VIBRATIONS AND INSTABILITIES

OF A DISK GALAXY

WITH MODIFIED GRAVITY

by

STANLEY ARVID ERICKSON, JR.

S.B., Massachusetts Institute of Technology
1964

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF
PHILOSOPHY
at the
MASSACHUSETTS INSTITUTE OF
TECHNOLOGY
June, 1974

Signature of Author . . . . . .
Department of Mathematics
May 3, 1974

Certified by . . . . . .
Thesis Supervisor

Accepted by . . . . . .
Chairman, Departmental Committee
on Graduate Students

VIBRATIONS AND INSTABILITIES OF A DISK GALAXY WITH MODIFIED GRAVITY

by

STANLEY ARVID ERICKSON, JR.

ABSTRACT

This thesis studies the effect of a modification of the gravitational potential upon the vibrations and instabilities of a thin, differentially rotating disk subject only to its own gravity. Instead of the usual potential $\phi = -GM/R$ at a distance R from a mass point M, it is assumed here that

$$\phi = GM/(R^2 + a^2)^{\frac{1}{2}} ,$$

meaning that gravitational force has been greatly reduced at distances smaller than the fixed length a. The disturbances examined encompass the entire disk, but are restricted to particle displacements that are both infinitesimal in amplitude and horizontal in direction. The approach taken is largely numerical, yet some physical interpretation is also provided.

After a brief historical introduction and some remarks in defense of the modified gravity in Chapter I, the dynamical equations for the perturbations of the disk are developed and the specific choice

of equilibrium model is discussed in Chapter II. In Chapter III a detailed description of the axisymmetric modes of the disk for both normal gravity and modified gravity is given, stressing the excellent agreement between the estimates of short wavelength theory and the computed point of stabilization of the disk, the existence and disappearance of discrete modes and the shape of the eigenfunctions. Finally, in Chapter IV a number of the nonaxisymmetric modes of the disk are discussed; included among them are open spirals as well as other unstable modes which do not become stable with the use of modified gravity.

June 1974
Massachusetts Institute of Technology
Department of Mathematics
Thesis Supervisor: Professor Alar Toomre

TABLE OF CONTENTS

ACKNOWLEDGEMENT

First and foremost, I must thank my thesis advisor, Professor Alar Toomre, for his patient help and guidance in every phase of this work. Without his forbearance and diligence the thesis would never have been completed.

The support I received at my place of employment, the Naval Underwater Systems Center, was also crucial to my being able to continue this work. Especially Dr. John Brady, Louis Bisci and A.M. Bottoms strongly and persistently encouraged my efforts. NUSC was also responsible for some of my tuition and salary support during the work on the thesis.

For their support and assistance I would also like to thank Professors C.C. Lin and C. Hunter; and A. Kalnajs for his comments on an early draft of the thesis.

I am also greatly indebted to the typists who prepared the drafts of my work, especially Mersina Christopher, and most of all, Fay Wark who typed the final manuscript under extreme deadline pressure. Also the assistance of the computer personnel at NUSC and M.I.T. is appreciated.

Lastly, I must thank my wife for tolerating my absence for long periods and for never failing to encourage me; also several friends whose moral support was invaluable at certain periods.

I. INTRODUCTION

During the past decade, the main preoccupation of theorists concerned with the dynamics of galaxies has doubtless been the problem of the spiral structures present in so many of the highly flattened systems. Owing largely to the pioneering work of Lindblad, and later of Lin and his associates, it is widely believed now that gravitation is the prime cause of the spirals. Gas and magnetic forces surely also affect these structures, but their role seems largely confined to such subordinate processes as interstellar cloud dynamics and the much-needed formation of young and highly visible stars.

(a) Background

Speculations as to the dynamical causes of the spirals are almost as old as the discovery by Lord Rosse in 1845 of the "whirlpool" nature of Messier 51. Among the more notable suggestions involving gravitational dynamics are those made already by Alexander (1852) of various clumpings and shearing of nebular matter, by Chamberlin (1901) of tidal forces from a neighboring galaxy, and by Jeans (1929 and earlier) of the shedding of material essentially from unstable Maclaurin spheroids.

These conjectures were carried much farther by Bertil Lindblad, whose extensive work in this subject (see especially Lindblad 1927, 1948, 1957, 1961, 1963) spanned nearly four decades. To be sure, there exists no unified Lindblad theory of spiral structure that one can point to in retrospect: from early ideas similar to those of Jeans, Lindblad proceeded through a long and varied series of attempts at understanding the spiral structures in terms of the motions and oscillations of many separate stars. It was probably such over-emphasis on individual orbits rather than on truly collective effects that prevented him from producing any satisfactory theory of the spirals. Yet it was certainly Lindblad who most inspired the modern theoretical analyses; indeed he also introduced many of the concepts in use today, and in his last papers he even discussed the idea of density waves.

Even so, the real flurry of theoretical analyses of spiral structure dates only from the 1960's. The most deservedly well-known work which began then is that of C.C. Lin and his associates. Lin and Shu (1964, 1966; Lin 1967, 1971; Shu 1968) greatly developed Lindblad's idea that at least the most striking spirals are wave phenomena. Their key contribution to the understanding of such presumed "quasi-stationary" waves involving entire galactic disks was a dispersion relation for stellar-dynamic density waves coupled by self-gravity. Although strictly based only on a short-wavelength approximation to the Poisson and collisionless Boltzmann equations, this dispersion relation and its implications were applied with con-

siderable success by Lin, Yuan and Shu (1969), by Shu, Stachnik and Yost (1971), and by Mathewson, van der Kruit and Brouw (1972) to interpreting the observed structures of the Milky Way and several other galaxies.

Certain difficulties with the theory remain, however. Many of the grand spirals are not tightly wrapped; this obviously strains the Lin-Shu assumption that all radial wavelengths are short compared to the scale of the disk. Remarkably, their theory does not even discriminate clearly between leading and trailing spirals. Furthermore, boundary conditions for the waves at the center and edge, and at any internal resonances, are hard to supply for such asymptotic analyses. And perhaps most worrisome is that even spiral waves have a group velocity and propagate in the radial direction (Toomre 1969) -- tending to empty a disk of such structures in a few galactic years unless some source of replenishment exists. Although several sources have been proposed (cf. Toomre 1969, Lin 1970, Lynden-Bell and Kalnajs 1972), this portion of the theory remains distressingly unresolved. Thus there is clearly a need for global, non-WKBJ analyses incorporating automatically (if perhaps not explicitly) the various transport and boundary conditions for the instabilities and vibrations of entire disks. In fact, a number of analyses of this sort have already been attempted, using widely varying techniques.

The first full-disk studies were performed by Hunter (1963, 1965). He began by constructing a family of cold disk models (no

random stellar motions) based on polynomial representations of the surface density. Then using techniques that were mostly analytical, he was able to determine both axisymmetric and non-axisymmetric modes of vibration for the first member of that family -- namely a Maclaurin disk rotating as a solid body -- as well as the axisymmetric frequencies for several other members. Many of the eigenfrequencies proved to be complex, representing a tendency of the disks to fragment into dense clumps of matter due to Jeans instability. Later using WKBJ techniques, Hunter (1969) also uncovered some strongly growing spiral waves of a "leading" sense in several of these models, but he did not find any of the trailing waves which seem much preferred by observations.

Some completely numerical studies of cold disks were also undertaken. Rehm (1965) and Miyamoto (1969), just as Toomre (1964) had done in an axisymmetric context, divided a disk into thin concentric rings to estimate its non-axisymmetric modes of oscillation. Both Rehm and Miyamoto again found numerous spiral instabilities of leading planform, as well as certain others. However, also like Hunter's analyses, these numerical studies were plagued by the severity of the deduced gravitational clumping. This made it difficult to sort out any subtler behavior, and also risky to extrapolate it to actual galaxies. Hence all the cold-disk studies served mainly to emphasize that rotation alone cannot stabilize thin model galaxies: a significant amount of random stellar motion is required as well (cf. Toomre 1964).

To date, Kalnajs (1965, 1970, 1971) has made the most extensive efforts to explore analytically the truly non-local behavior of

disks with the requisite random motions. His work, which began independently of Lin, has at least in principle allowed spiral waves of very open planform. Even so, the only complete solutions yet obtained by Kalnajs (1972) have been those for a class of uniformly rotating Maclaurin-Freeman disks: as it turned out, their most enduring instability is not spiral at all, but is bar-like, and its suppression requires a good deal more random motion than the demise of the last axisymmetric instabilities. Also according to Kalnajs, other model galaxies involving shear as well as random motions have so far proved intractable, except via numerical methods requiring severe approximations. Similar difficulties were encountered by Shu (1968, 1970).

To circumvent such complexity, a totally different approach to the dynamics of hot disks was followed by Prendergast and Miller (1968), Hockney and Hohl (1969), Prendergast, Miller and Quirk (1970), and Hohl (1971). Those authors simulated a hot disk galaxy directly as a kind of N-body problem. In other words, they followed a large number of mass points along trajectories confined to a single plane, allowing each particle to be affected only by the collective gravity. As expected from the studies already mentioned, these simulated disks were again found to be unstable when cold or cool, and to become smooth and stable only when the random motions had increased sufficiently. Again, nonaxisymmetric instabilities of large spatial extent proved hardest to stabilize, requiring several times the random velocities predicted by Toomre's axisymmetric analysis. In some of the experiments,

these instabilities took the form of trailing spirals which appeared and dissolved; in other instances they were more bar-like. However, owing to difficulties in interpreting all the outcomes -- not helped by the impermanence of the computed spirals -- no thorough understanding of spiral mechanisms emerged from those studies either.

(b) Rationale for the Modified Gravity

In this thesis we attempt to steer between the two hazards of the instabilities of cold disks and the complexities of hot ones. Our strategy will be to simulate the effects of stellar random motions through the artifice of a greatly reduced gravitational force between nearby stars. In essence, we seek thereby to avoid the great bother of having to deal explicitly with random motions, but we do indeed pretend that it is their diffusive or smearing action which renders the local gravity much less effective than it would otherwise have been.

This softening is accomplished here in about the simplest manner possible: the standard potential $-GM/R$ at a distance R from a point mass M is just replaced by

$$- GM / (R^2 + a^2)^{\frac{1}{2}} , \qquad (1)$$

where G remains the familiar gravitational constant but where the newly-introduced length a represents a finite if gradual close-

range cut-off distance for the effective gravity. Notice that this assumed force per unit mass,

$$F_R = - GM\, R/(R^2 + a^2)^{3/2} , \qquad (2)$$

thus attains its maximum magnitude when $R = a/\sqrt{2}$, and that it actually vanishes as $R \to 0$. Obviously the force (2) can also be regarded as the "horizontal" component of the full force between two particles differing in "height" by distance a and otherwise separated by a projected distance R. Hence we will often refer to a simply as the vertical offset.

As recognized also by Miller (1971), such a modification of the gravity already promises axisymmetric stability for a disk without random motions, provided the offset a is chosen sufficiently large.

This promise can best be understood by recalling (cf. Toomre 1964) that any radially sinusoidal wave of (positive) wavenumber k, in a cold thin disk of local surface density μ that is subject only to standard gravity, would vibrate with a frequency ω such that

$$\omega^2 = \kappa^2 - 2\pi\, G\, \mu\, k . \qquad (3)$$

As indicated by this approximate formula, the net effect of the local self-gravity represented by the second term is always to reduce the wave frequency ω below the single-particle or "epicyclic" frequency

κ contained in the first term. Obviously for wavenumbers

$$k > k_{crit} \equiv \frac{\kappa^2 (r)}{2\pi G \mu (r)} \tag{4}$$

the expected ω^2 is negative, and the wave is unstable.

Using the modified gravity, however, the dispersion relation becomes

$$\omega^2 = \kappa^2 - 2\pi G \mu k e^{-ak} . \tag{5}$$

Left implicit here is any (generally small) change in the central force field itself, and hence in κ. Much more significant is the exponential factor in equation (5): it stems from the fact that, as we have already implied, our new "gravitational" potential associated with any given sinusoidal density field of short wavelength is equivalent to the standard potential

$$\phi' (r,z) = A e^{-k|z|} \cos kr \tag{6}$$

evaluated at a fixed height $|z| = a$ above or below the plane of the disk.

Since the largest possible value of $x e^{-x}$ is $1/e$, the local force term in equation (5) cannot exceed $2\pi G \mu/a e$. Hence,

if at every radius r it happens to be true that the given

$$a > a_{crit}(r;a) \equiv \frac{2\pi}{e} \frac{G\,\mu\,(r)}{\kappa^2\,(r;a)} , \tag{7}$$

it is clear that that second term can nowhere overwhelm κ^2 -- and then, at least superficially, we seem assured of small-scale axisymmetric stability everywhere.

To be sure, such easy stabilization of all short axisymmetric disturbances is by no means unique to our device of the modified gravity. One simple alternative that has often been contemplated is to mimic the stellar random motions instead by an isotropic gas pressure: indeed if such a phoney "gas" disk is imagined somehow to remain very thin in the z-direction, its axisymmetric dispersion relation becomes

$$\omega^2 = \kappa^2 - 2\pi\,G\,\mu\,k + c^2k^2 \tag{8}$$

(cf. Hunter 1972), where c denotes the local speed of sound. Clearly

$$c > c_{crit} \equiv \frac{\pi\,G\,\mu(r)}{\kappa(r)} \tag{9}$$

then suffices to ensure the same sort of local stability as our $a > a_{crit}$.

There is, nevertheless, one strong a priori reason for preferring something like the reduced gravity to the gas pressure: no upper limit at all exists to the possible frequency ω implied by the gas equation (8) for shorter and shorter wavelengths, whereas our relation (5) clearly indicates that $\omega \to \kappa$ as $k \to \infty$. Although the former, $\omega \to \infty$ behavior is understandable as a kind of acoustic limit, it is hardly what one expects for short, radially propagating waves in an actual disk of stars. On the contrary, the well-known dispersion relation

$$\omega^2 = \kappa^2 - 2\pi G \mu k F_\nu(\chi) \qquad (10)$$

of Lin and Shu (1966), (but see also Kalnajs, (1965)), for such stellar waves does not admit any wave frequencies ω in excess of the epicyclic frequency κ -- as emphasized especially by the appropriate label "reduction factor" coined for the function $F_\nu(\chi)$ by Lin and Shu. More exactly, that reduction factor (which seems unnecessary to exhibit here) tends to zero faster than $1/k$ as $k \to \infty$, meaning that very short-wave stellar vibrations have the same limiting frequency $\omega \simeq \kappa$ as the waves governed by our modified gravity. Hence, unlike any gas disks, our models with the faked gravity should at least mimic the Lindblad resonances expected in disks of stars.

These differences and similarities may also be judged from Figure 1, which shows the relative frequence ω/κ plotted as a

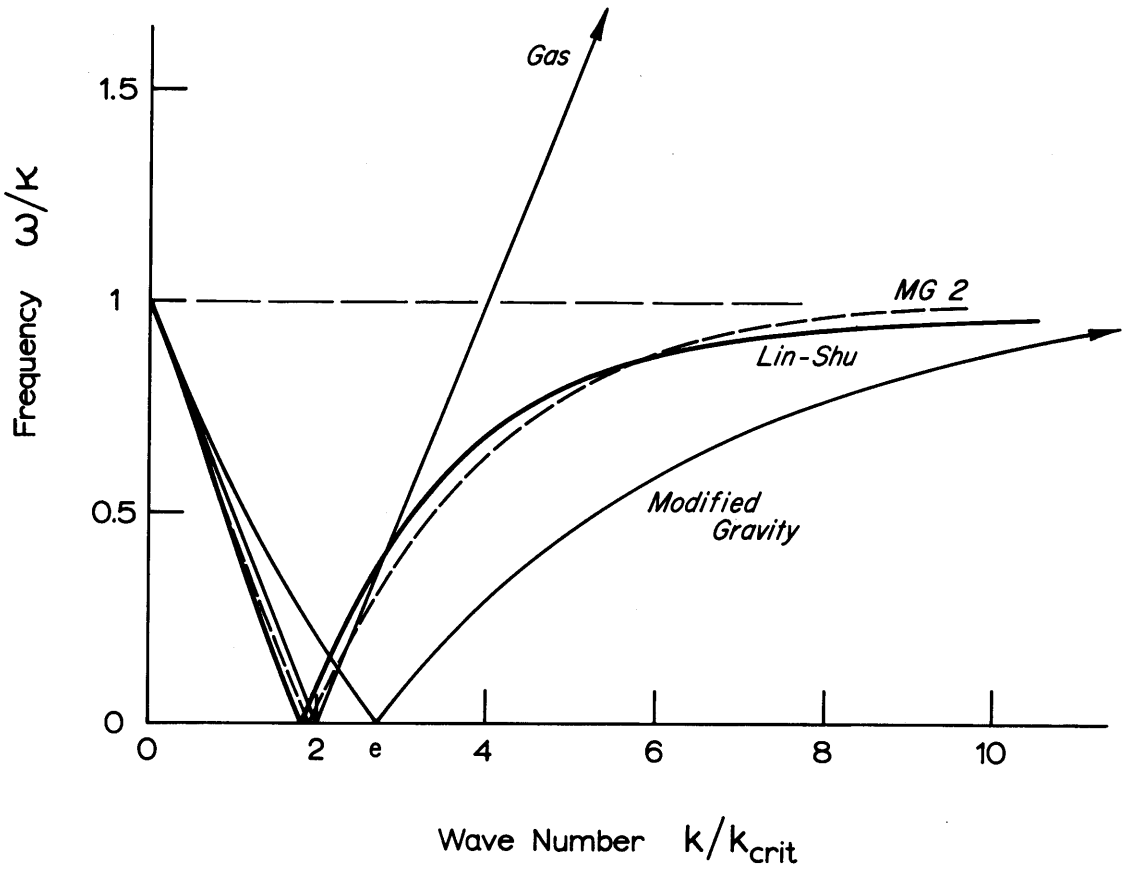

Figure 1. This diagram compares the frequency ω predicted for various wavenumbers k by (i) the Lin-Shu formula for a disk of stars, (ii) a thin gaseous disk, and (iii) our modified gravity, all in circumstances of marginal stability. Also shown by broken curve "MG2" is a second modified gravity relation which fits the Lin-Shu curve more closely; this curve was generated using a potential $\{\phi(r,a) - a\partial\phi(r,a)/\partial a\}$ where $\phi(r,a)$ is the usual gravity modification.

function of the dimensionless wave number k/k_{crit}, as predicted from each of the three dispersion relations (5), (8) and (10) in circumstances that are barely stable. Although these curves for the reduced gravity and the Lin-Shu relation are by no means coincident, they seem close cousins indeed when compared with the gas example for the larger values of k. Of course, it should be possible to modify the "gravitational" potential further, so as to create a dispersion relation which matches the Lin-Shu curve really closely (see Figure 1) -- but understandably for this initial exploration we have simply opted to confine all our efforts to disks subject only to the elementary potential (1).

II. DYNAMICAL EQUATIONS

Throughout this thesis we will be dealing with infinitesimal horizontal displacements of particles from their circular equilibrium orbits in an originally axisymmetric thin disk. The details of our equilibrium model will come later. What matters at present is simply that any such disk can be completely specified by its surface density $\mu(r)$ given as a function of the radius r -- provided one assumes, as we do here, that all its mass is strictly confined to the plane $z = 0$. Even with our modified gravity, this density then obviously defines a unique gravitational potential $\phi(r, z; a)$ and that in turn dictates that the angular speed $\Omega(r; a)$ must satisfy

$$r\,\Omega^2\,(r, a) = \partial\phi/\partial r\,|_{z=0} \qquad (11)$$

for centrifugal balance; for all values of the vertical offset a, this function $\Omega^2\,(r, a)$ is non-negative and thus sensible for any $\mu(r)$ that nowhere increases outwards.

(a) Motion Of A Single Test Particle

Before we begin the study of the perturbation of the whole disk, let us review the characteristic behavior of a massless test

particle when slightly perturbed from a circular equilibrium orbit. If we take r and θ to be the polar coordinates of the particle relative to the disk center, any trajectory of the particle through the disk must be governed by

$$\begin{aligned} \ddot{r} - r\,\dot{\theta}^2 &= -r\,\Omega^2(r) + f_r(r,\theta,t) \\ r\,\ddot{\theta} + 2\dot{r}\,\dot{\theta} &= f_\theta(r,\theta,t) \end{aligned} \tag{12}$$

where dots denote differentiation with respect to time t, and f_r and f_θ denote the radial and tangential components of the acceleration caused by an additional applied force.

In this thesis, we will only be dealing with infinitesimal deviations from exactly circular trajectories, so we immediately linearize equation (12) in terms of two new variables

$$\begin{aligned} \xi &= r - r_0 \\ \eta &= (\theta - \theta_0 - \Omega t)\, r_0 \,, \end{aligned} \tag{13}$$

the radial and tangential displacements of the particle from its undisplaced position r_0, $\theta_0 + \Omega t$. To first order in ξ and $\dot{\eta}$, equation (12) becomes

$$\begin{aligned} \ddot{\xi} - 2\Omega\dot{\eta} &= -(\kappa^2 - 4\Omega^2)\,\xi + f_r \\ \ddot{\eta} + 2\Omega\dot{\xi} &= f_\theta \end{aligned} \tag{14}$$

For good reason we have introduced here $\kappa(r)$, the epicyclic frequency, which is defined by:

$$\kappa^2(r) = 4\Omega^2(r) + r \frac{d}{dr} \Omega^2(r). \tag{15}$$

This particular frequency is of special importance because it is the only frequency at which test particles oscillate around their equilibrium orbits when perturbed.

As is quickly shown, two of the four characteristic or modal solutions of the unforced version of equation (14) have zero frequency; the other two have $\pm\kappa$. Just as with a simple harmonic oscillator, these latter two solutions differ only in phase and can be written

$$\begin{aligned} \xi &= c \cos(\kappa t + \alpha) \\ \eta &= -c' \sin(\kappa t + \alpha) \end{aligned} \tag{16}$$

where c and α are constants of integration and $c' = 2\Omega c/\kappa$.

Equation (16) describes the well-known epicyclic motion of a test particle. If we watch the particle from a coordinate system rotating with angular speed Ω, we will see it trace out an ellipse with its center (the "epicenter") at the unperturbed position of the particle. In the usual case of Ω decreasing with radius, the minor axis of the ellipse points directly outwards from the center of the disk.

However, viewed in fixed, non-rotating coordinates, the motion of the particle is an infinitesimally wavering departure from the equilibrium orbit. The epicenter does move in the original orbit with constant speed Ω, but the test particle itself revolves around the epicenter, repeatedly crossing the equilibrium trajectory.

Currently, the word "epicycle" is used to describe such motion, as it is a mild generalization of the ancient Ptolemaic circles upon circles. Similar motions occur in plasma theory, as a charged particle in a magnetic field revolves around a "guiding center", which itself drifts along the magnetic field lines (see, for example, Spitzer (1962)).

To finish off the discussion of the motion of our test particle, let us note that the other two non-epicyclic solutions of equation (16) are physically simple. One merely corresponds to a constant change in longitude:

$$\begin{aligned} \xi &= 0 \\ \eta &= c \end{aligned} \tag{17}$$

and the other to a constant rate of change in longitude, accompanied by a slight change in radius:

$$\begin{aligned} \xi &= c \\ \eta &= c't \end{aligned} \tag{18}$$

where c is an infinitesimal constant, and

$$c' = (\kappa^2 - 4\Omega^2)\, c/2\Omega .$$

In the latter mode, note that η can also be written as

$$\eta = r_0 \frac{d\Omega}{dr} ct$$

which implies that the angular speed of the particle is

$$\dot{\theta} = \Omega(r_0) + \frac{d\Omega}{dr} c ,$$

or just the equilibrium angular speed at the new orbital radius, $r_0 + c$.

(b) Motion Of A Ring Of Test Particles

As an intermediate step in the development of the perturbation equations for the entire disk, let us examine the perturbation of a ring of many particles, equally spaced along a single equilibrium orbit. This will elucidate the separate and combined roles of Ω and κ and also serve to introduce the notation we shall use. Again let us discuss only massless particles, which move in the field of the equilibrium potential without interacting with each other.

Rather than discuss the complicated general motions of the particles, let us focus our attention on motions periodic in azimuth,

specifically, varying sinusoidally as $m\theta$, where m is an integer. If we desired we could create kinematic waves of any azimuthal periodicity by cleverly arranging the phases of the epicyclic motions of our ring of particles. Such a contrivance is shown in Figure 2, where the phases were chosen at $t = 0$ proportional to $\cos 2\theta$. Here each epicycle has identical amplitude, and the epicenters lie on the same equilibrium orbit. At a time $t = (\pi/2)\kappa^{-1}$, the particles have moved a quarter turn around their epicycles, and the pattern has rotated $\pi/2m$ around the equilibrium orbit. Evidently, this wave has a pattern speed of κ/m in rotating coordinates.

However, since we will soon be dealing with particles of different angular speeds, it seems more in order to examine the motion in a fixed coordinate system, rather than one rotating with some particular Ω. Therefore, we need to write equation (16) in Eulerian variables. Let $\xi'(r, \theta, t)$ and $\eta'(r, \theta, t)$ be the infinitesimal displacements from (r, θ) of that particle now at (r_1, θ_1) which we would have found at (r, θ) had it continued in its equilibrium trajectory:

$$\begin{aligned} \xi' &= r_1 - r \\ \eta' &= (\theta_1 - \theta)\, r \,. \end{aligned} \tag{19}$$

Traditionally, the transformation from Lagrangian to Eulerian variables is handled very compactly by replacing the Lagrangian deriva-

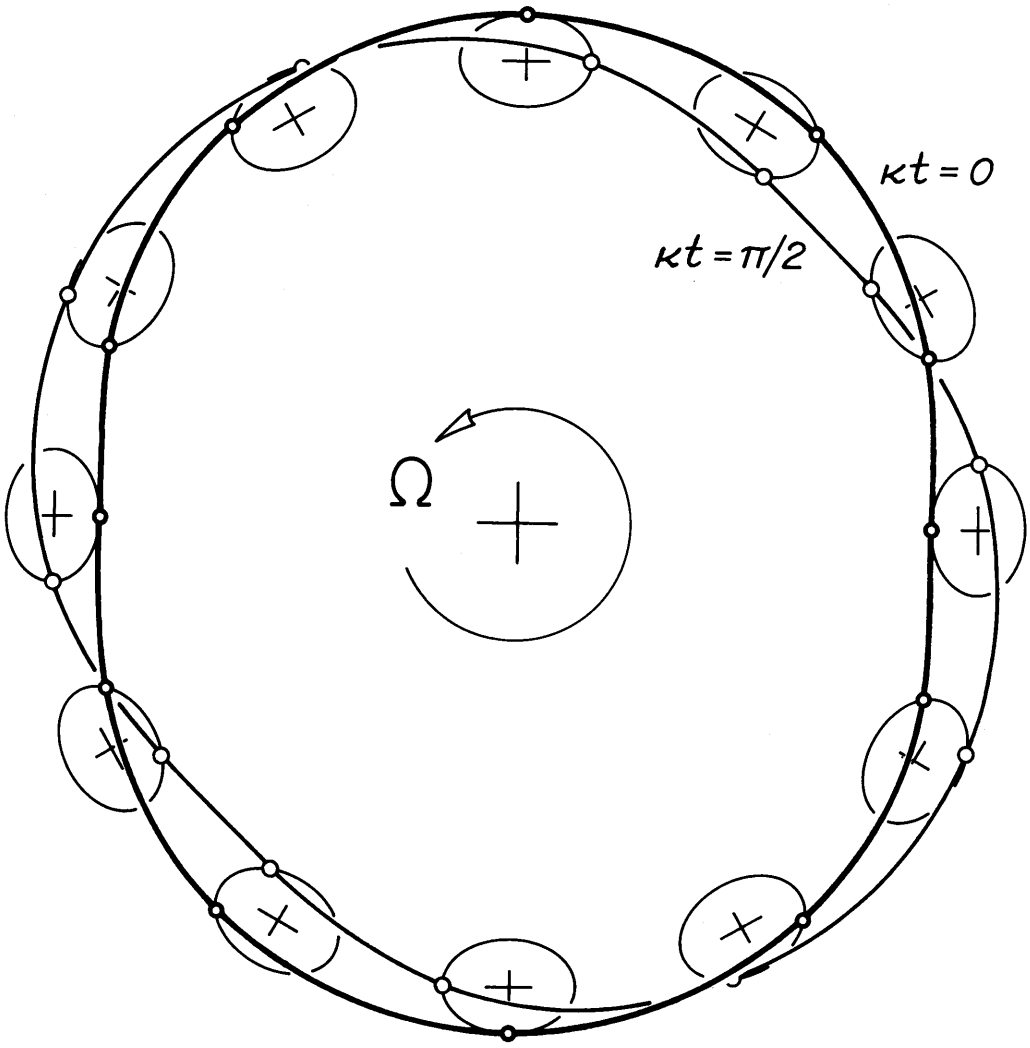

Figure 2. Fast kinematic wave with periodicity m = 2 on a ring of massless particles. The angular speed is $\Omega + \kappa/m$ in fixed coordinates.

tive d/dt, as applied to a single particle, by its equivalent Eulerian derivative, D_t, which is applied with respect to fixed coordinates. In general,

$$D_t = \partial/\partial t + (\vec{v} \cdot \vec{\nabla})$$

but our linearization simplifies it to

$$D_t = \partial/\partial t + \Omega \, \partial/\partial\theta$$

in our polar coordinates. Hence, equation (16) becomes

$$\begin{aligned} D_t^2 \xi' - 2\Omega D_t \eta' &= -(\kappa^2 - 4\Omega^2)\xi' + f_r' \\ D_t^2 \eta' + 2\Omega D_t \xi' &= f_\theta' \end{aligned} \tag{20}$$

where f_r' and f_θ' are now expressed in terms of fixed coordinates.

Fortunately, this equation, and indeed all the dynamical equations of this thesis, are easier to solve using complex variables as a shorthand notation. The simple reason for this is that the m^{th} Fourier components of equation (20), expressed in terms of $e^{im\theta}$ and $e^{-im\theta}$ do not mix with each other in any situation, which cannot be said for components written using $\sin m\theta$ and $\cos m\theta$. Let us define the Fourier coefficients of ξ, η by

$$\xi' = \mathrm{Re}\,\{\xi_m(t)\, e^{i m\theta}\}$$
$$\eta' = \mathrm{Re}\,\{\eta_m(t)\, e^{i m\theta}\}\ , \tag{21}$$

and similarly decompose the forces

$$f_r' = \mathrm{Re}\,\{f_{r,m}\, e^{i m\theta}\}$$
$$f_\theta' = \mathrm{Re}\,\{f_{\theta,m}\, e^{i m\theta}\}\ . \tag{22}$$

In these new variables, equation (20) becomes a set of equations for the Fourier coefficients:

$$D_t^2\,\xi_m - 2\Omega\, D_t\,\eta_m = -(\kappa^2 - 4\Omega^2)\,\xi_m + f_{r,m}$$
$$D_t^2\,\eta_m + 2\Omega\, D_t\,\xi_m = f_{\theta,m}\ , \tag{23}$$

where D_t is now

$$D_t = \partial/\partial t + i\, m\, \Omega\ .$$

Since the dynamical equations are linear we can resolve them in terms of single frequency modes, or eigenmodes, varying with time as $e^{-i\omega t}$. The displacements and forces are expressed as

$$
\begin{aligned}
\xi' &= \mathrm{Re}\left(X\ e^{im\theta}\ e^{-i\omega t} \right) \\
\eta' &= \mathrm{Re}\left(i\ Y\ e^{im\theta}\ e^{-i\omega t} \right) \\
f_r' &= \mathrm{Re}\left(F_r\ e^{im\theta}\ e^{-i\omega t} \right) \\
f_\theta' &= \mathrm{Re}\left(i\ F_\theta\ e^{im\theta}\ e^{-i\omega t} \right)
\end{aligned}
\tag{24}
$$

and we will conveniently leave implicit the dependence of the four coefficients X, Y, F_r, and F_θ on m and ω. The imaginary number i has been included in the expressions for η' and f_θ' so that Y and F_θ will be real if ω is real, as we will see later on. In terms of these new variables, equation (23) turns into

$$
\begin{aligned}
D_t^2\ X - i\ 2\Omega\ D_t\ Y &= -(\kappa^2 - 4\Omega^2)\ X + F_r \\
D_t^2\ Y - i\ 2\Omega\ D_t\ X &= F_\theta
\end{aligned}
\tag{25}
$$

and D_t is now reduced to

$$
D_t = -i\omega + i\ m\ \Omega\ .
$$

Finally, we are ready to deduce the unforced motion of our ring of test particles. If we set

$$
F_r = F_\theta = 0
$$

in equation (25), we quickly arrive at a polynomial for the characteristic frequencies of the ring:

$$(\omega - m\Omega)^2 \left((\omega - m\Omega)^2 + \kappa^2\right) = 0 \,. \tag{26}$$

As we would expect, the solutions are the single particle solutions with a convective term added:

$$\begin{aligned} \omega &= m\,\Omega \pm \kappa \\ \omega &= m\,\Omega \quad \text{(twice)}. \end{aligned} \tag{27}$$

Now the uniqueness of the four eigenmodes becomes more apparent than in the single particle discussion. If we look at the first eigenmode:

$$\begin{aligned} \xi' &= \mathrm{Re}\left(X \exp i \{m\,\theta - (m\,\Omega + \kappa)t\}\right) \\ \eta' &= \mathrm{Re}\left(iY \exp i \{m\,\Omega - (m\,\Omega + \kappa)t\}\right) \end{aligned} \tag{28}$$

with X arbitrary and $Y = 2\Omega X/\kappa$, we find the formula for the well known rotating pattern of particles. As shown in Figure 3a, for $m = 2$, a ring of equispaced particles is perturbed into an ellipse, (of infinitesimal eccentricity), on which particles are bunched on the inner portions and separated on the outer portions. This ellipse rotates without change of shape with pattern speed

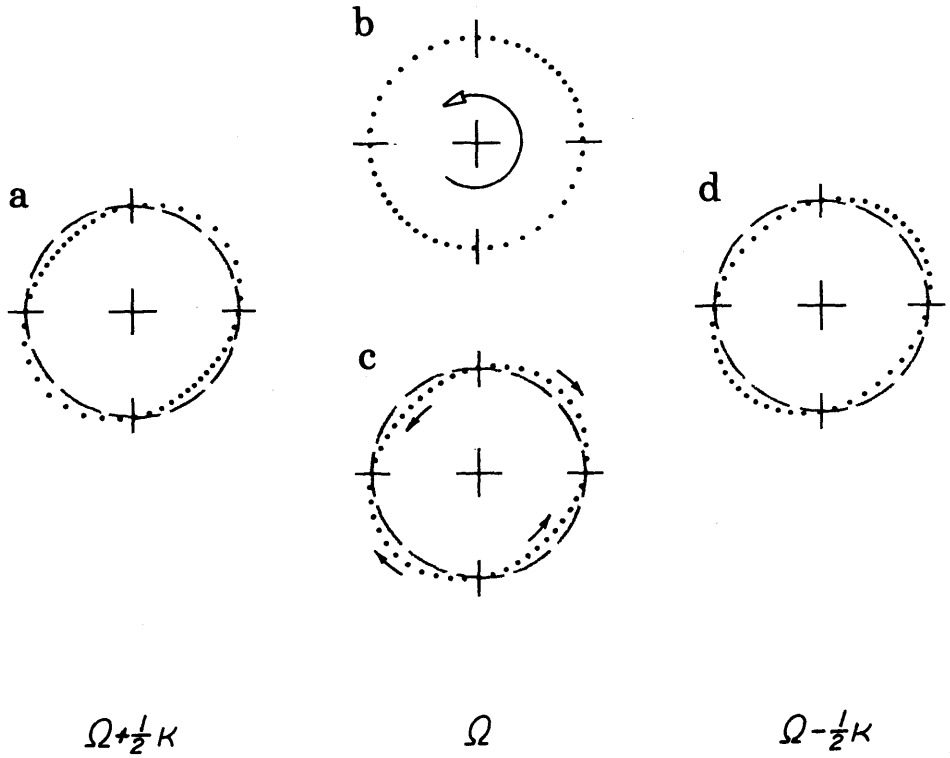

Figure 3. The four elementary modes of a ring of massless particles displaced with periodicity m = 2. Pattern speeds ω_p are shown below the diagrams.

$$\omega/m \quad = \quad \Omega + \kappa/m \ ,$$

and any observer fixed in space just inside the ring will see it oscillate with the frequency ω. If we had endowed the particles with a slight mass, this observer would measure a gravitational force with frequency ω exactly in phase with ξ'. This force arises from two causes, the bunching of particles, and the reduction in distance between the observer and the ring where it bulges outward.

Similarly, the second eigenmode, given by

$$\begin{aligned} \xi' &= \text{Re} \left(X \exp i \{m\theta - (m\Omega - \kappa)t\} \right) \\ \eta' &= \text{Re} \left(iY \exp i \{m\theta - (m\Omega - \kappa)t\} \right) \end{aligned} \tag{29}$$

with again X arbitrary and $Y = 2\Omega X/\kappa$, can be represented by a rotating ellipse, (for $m = 2$), but several differences exist between it and the first eigenmode. It rotates with a pattern speed

$$\omega/m \quad = \quad \Omega - \kappa/m$$

which may be negative if κ exceeds $m\Omega$ at the ring radius. If the pattern speed is negative, the ellipse rotates in an inverse or retrograde direction compared to the rotation of the disk. Also, the bunching we mentioned before occurs on the portion of the ellipse outside the equilibrium orbit, as is quite visible in Figure 3.d. In this

eigenmode our observer would note a lesser gravitational effect than in the first mode, as the force caused by the bulging of the ring counteracts that caused by bunching of the particles. If we moved our observer to just outside the ring, the inverse would occur, and the second eigenmode would produce more gravitational force at his new location.

Now let us turn to the ring motion derived from the third particle eigenmode, in which

$$\begin{aligned} \xi' &= 0 \\ \eta' &= \mathrm{Re}\,(i\, Y \exp i\,\{m\,\theta - m\,\Omega\, t\}) \end{aligned} \tag{30}$$

with Y arbitrary. This consists simply of sinusoidal bunchings and separations along the equilibrium orbit, as depicted in Figure 3b. The pattern rotates exactly at orbital speed, so the pattern speed is merely

$$\omega/m = \Omega\ .$$

Finally, we note that the fourth mode, also of frequency $m\Omega$, has a secular rather than purely sinusoidal nature:

$$\begin{aligned} \xi' &= \mathrm{Re}\,(X \exp i\,\{m\,\theta - m\,\Omega\, t\}) \\ \eta' &= \mathrm{Re}\,(\dot{Y}\, t \exp i\,\{m\,\theta - m\,\Omega\, t\}) \end{aligned} \tag{31}$$

where X is arbitrary and $\dot{Y}$ is a speed given by

$$\dot{Y} = \frac{\kappa^2 - 4\Omega^2}{2\Omega} X .$$

The particles here oscillate with constantly increasing amplitude in the tangential direction. The picture of this eigenmode at $t = 0$ is shown in Figure 3c for $m = 2$, and is again an ellipse, but lacks the bunching of the earlier modes. This ellipse immediately begins to distort as it rotates. If the motion had finite amplitude instead of infinitesimal, the particles would eventually pass one another in longitude and shear the ellipse into a quite complicated figure.

Before leaving this discussion of the motion of a ring of particles, we want to remark that complex notation is a convenient way of manipulating the displacements X and Y, as well as F_r and F_θ. To see this we expand, as an example, the displacements of the first eigenmode, as given in equation (28):

$$\begin{aligned} \xi' &= \text{Re } X \cos\{m\theta - (m\Omega + \kappa)t\} - \text{Im } X \sin\{m\theta - (m\Omega + \kappa)t\} \\ \eta' &= \text{Im } Y \cos\{m\theta - (m\Omega + \kappa)t\} - \text{Re } Y \sin\{m\theta - (m\Omega + \kappa)t\} . \end{aligned} \tag{32}$$

It is clear from this expansion that the real part of X specifies the $t = 0$ radial displacement at $\theta = 0$, and the imaginary part, that at $\theta = \pi/2m$. These two displacements completely characterize ξ' in the first eigenmode. Similarly the real part of Y specifies

the tangential displacement at $\theta = \pi/2m$ and the imaginary, that at $\theta = 0$. The force terms F_r and F_θ as well follow this pattern.

Note also that the stable, solely vibratory, eigenmodes, the first three given here, can be expressed by a choice of phase as

$$\xi' = c \cos (m\theta - \omega t)$$

$$\eta' = c' \sin (m\theta - \omega t)$$

with radial and tangential displacements $\pi/2$ out of phase. If only those stable eigenmodes existed, we would have no need for complex coefficients, at least in the case of a single ring. However, as we noted before, an unstable eigenmode, such as the fourth, breaks this pattern and forces us to include one or both of the other two terms:

$$\xi' = c'' \sin(m\theta - \omega t)$$

$$\eta' = c''' \cos(m\theta - \omega t)$$

in our calculations. As we will see, this phenomenon extends to the more complicated case of perturbations involving the whole disk.

(c) Disturbance Densities And Forces

So far we have been examining the motions of massless particles, affected only by the gravitational field of the unperturbed disk. How-

ever, we really need to discuss the perturbed motions of all the particles constituting the disk. These perturbations will almost always give rise to a surface density which in turn will cause a gravitational force. And in turn, that force -- left implicit as the f_r' and f_θ' terms in the dynamical equation (20) -- will affect the motion of the particles. In essence, we need to complete equation (20) by making those forcing terms self-consistent, that is, dependent on the displacements ξ' and η' themselves.

Of course, the perturbation density μ' can be deduced immediately from the pertinent continuity equation:

$$\mu' = -\frac{1}{r}\frac{\partial}{\partial r}(r\,\mu\,x') - \frac{1}{r}\,\mu\,\frac{\partial}{\partial\theta}y' \,. \tag{33}$$

As with the displacement variables, we decompose the density into Fourier components:

$$\mu' = \mathrm{Re}\,\{s_m(r,t)\,e^{im\theta}\}\,, \tag{34}$$

and further into eigenmodes:

$$\mu' = \mathrm{Re}\,\{S(r)\,e^{im\theta}\,e^{-i\omega t}\}\,. \tag{35}$$

In terms of the latter coefficient, equation (33) becomes

$$S(r) = -\frac{1}{r}\frac{d}{dr}(r\,\mu\,X) + \frac{m}{r}\,\mu\,Y\,, \tag{36}$$

and we are finished with continuity.

For any disk with surface density μ', the potential ϕ' at $\vec{r}$ is found by merely superimposing the contribution from each infinitesimal mass dm,

$$- \frac{G \quad dm}{\sqrt{\Delta^2 + a^2}} \quad ,$$

where Δ is the distance between the mass dm and $\vec{r}$, and a is our now-familiar offset. This leads to

$$\phi'(r,\theta,t) = -G \int_0^{\infty} \int_0^{2\pi} \frac{\mu'(\rho,\beta,t)\ \rho\ d\rho\ d\beta}{(r^2 + \rho^2 + a^2 - 2r\rho\ \cos(\theta - \beta))^{\frac{1}{2}}} \ . \tag{37}$$

Clearly, if μ' consists of only one Fourier coefficient, as in equation (34), ϕ' will also, and we may write it as

$$\phi' = \text{Re}\ \{p_m\ (r,t)\ e^{im\theta}\} , \tag{38}$$

where

$$p_m\ (r,t) = \int_0^{\infty} K_m\ (r,\rho)\ s_m\ (\rho,t)\ d\rho$$

in terms of the kernel

$$K_m\ (r,\rho) = -G\ \rho \int_0^{2\pi} \frac{\cos\ m\beta\ d\beta}{(r^2 + \rho^2 + a^2 - 2r\rho\ \cos\ \beta)^{\frac{1}{2}}} \ . \tag{39}$$

Our kernel is also expressible in a variety of other ways: using Bessel functions

$$K_m(r, \rho) = -2\pi G\rho \int_0^\infty J_m(kr)\, J_m(k)\, e^{-ak}\, dk , \quad (40)$$

hypergeometric functions

$$K_m(r,\rho) = -G \left(\frac{\pi \rho}{2^{2m-1} r}\right) \frac{\Gamma(m + \frac{1}{2})}{\Gamma(m + 1)} \left(\frac{2 r}{a^2 + r^2 + \rho^2}\right)^{m + \frac{1}{2}} \cdot \quad (41)$$

$$F\left\{\frac{m}{2} + \frac{3}{4}, \frac{m}{2} + \frac{1}{4}, m + 1; \left(\frac{2 r \rho}{a^2 + r^2 + \rho^2}\right)^2\right\} ,$$

associated Legendre functions

$$K_m(r, \rho) = -2 G \sqrt{\frac{\rho}{r}}\, Q_{m-\frac{1}{2}}\left(\frac{a^2 + r^2 + \rho^2}{2 r \rho}\right) , \quad (42)$$

and doubtless others. As discussed in Appendix V, we found the last-named formula convenient for computation. This kernel is of course used without change in the decomposition of the potential into eigenmodes:

$$\theta' = \mathrm{Re}\{P e^{im\theta} e^{-i\omega t}\} , \quad (43)$$

where

$$P(r) = \int_0^\infty K_m(r,\rho)\, S(\rho)\, d\rho \,. \tag{44}$$

To complete this section, a decomposition of the self-consistent forces:

$$\begin{aligned} f'_r &= -\frac{\partial}{\partial r}\phi' \\ f'_\theta &= -\frac{1}{r}\frac{\partial}{\partial \theta}\phi' \end{aligned} \tag{45}$$

is in order. Azimuthal components are simply

$$\begin{aligned} f_{r,m} &= \frac{\partial}{\partial r}\phi_m \\ f_{\theta,m} &= \frac{im}{r}\phi_m \,, \end{aligned} \tag{46}$$

and the eigenmodal components are

$$\begin{aligned} F_r &= \frac{d}{dr}P \\ F_\theta &= \frac{m}{r}P \,. \end{aligned} \tag{47}$$

Note that F_θ has previously been defined with a factor of i multiplying it, as was Y, so that the complete dynamical equations are entirely real:

$$D^2 X - 2\Omega D Y = (\kappa^2 - 4\Omega^2) X + \frac{d}{dr} P$$

$$D^2 Y - 2\Omega D X = \frac{m}{r} P$$

$$D = -\omega + m \Omega \tag{48}$$

$$P = \int_0^\infty K_m (r,\rho) \left(-\frac{1}{\rho} \frac{d}{d\rho} (r \mu X) + \frac{m}{\rho} \mu Y \right) d\rho \ .$$

These equations are very suggestive of a characteristic value problem for a pair of coupled singular Fredholm integral equations. The similarity can be made even stronger by some simple manipulations, such as integrating the X-related force terms by parts to eliminate the derivative of X in the integrands, and by introducing two new variables, DX and DY. However, we have no need here for the formal modifications, as we only wish to discuss the choice of method of solution.

(d) Spline Approximation and Force Integrals

The literature on Fredholm integral equations is quite rich, containing several methods which might be applied to our equation (48). The formal theory developed by Fredholm (see, for example, Mikhlin (1957)), involves the calculation of a series of iterated kernels, each of which is the integral of a previous pair of kernels. Clearly, this method

would require a prodigious amount of calculation, and seems better suited to the proof of existence of solutions. We are more concerned with obtaining solutions here, and so we need to consider now the approximate methods available.

If we intended to find only a few eigenfunctions, our best choice would probably be an iterative method of solution similar to that of Kellogg (see Mikhlin (1957)) in which we would make an initial guess for X and Y, evaluate the force integrals, use equation (48) to find a second guess for X and Y, and then continue the process until convergence hopefully occurs. However, the results of Hunter and Toomre (1969) on the bending of a disk, which involved similar equations, raise the suspicion that the whole spectrum would itself be extremely interesting to see, whereas the first few eigenfunctions would not necessarily be so. Hence, we need a method more suitable for obtaining the whole spectrum, and, by elimination, we are left with the choice of some method of approximating equation (48) by a set of algebraic equations.

Broadly speaking, a method of approximating integral equations by a set of algebraic equations is specified by only two things: the set of functions used to expand the dependent variables, and the criterion used to minimize error. We discuss the first point here, and the second in the next section.

Initially we expanded the displacements as sums of the first n Chebyshev polynomials (Fröberg (1965)), written as functions of $\zeta = 1 - \exp(-r^2/2)$ which has the proper finite range for an edgeless

disk. We chose Chebyshev polynomials rather than any other set of polynomials principally because decomposition of an arbitrary function into a sum of the first n Chebyshev polynomials can be accomplished by merely evaluating the function at n points and then summing these values with n sets of weights. No extra integration is required.

This technique worked tolerably well in the axisymmetric case ($m = 0$), but we realized that as n increased, we were introducing very rapidly varying functions of radius into the displacements. Intuitively, we should better have chosen a set of smooth functions whose main feature was that they "localized" disturbances well, so that concentrated perturbations do not cause spurious forces large distances away from the perturbation. In other words, if our set of functions was used to synthesize a trial function which was zero everywhere except in a small region, the reproduction of the trial function should be nearly zero everywhere outside that region. The Chebyshev polynomials are not especially good at this, as they tend to distribute the error made in fitting trial functions over their whole domain, whereas we would like them to concentrate any error near the disturbance region, and leave the rest zero. No set of polynomials seems to do that well, so we turned to splines (see, for example, Ahlberg, Nilson and Walsh (1969)).

The localizing ability of splines is a result of their piecewise character. They are merely cubic polynomials fitted between adjacent points r_i and r_{i+1} of a set of well-chosen sampling

points $\{r_i, i = 1, n\}$. The polynomial coefficients for each interval are picked so that the resultant function and its first two derivatives are continuous, as well as satisfying the obvious requirement of producing the exact result at all the r_i. The details of the splines used, and of a companion method we used for testing purposes, are discussed in Appendix V.

We should also note here that X and Y were not themselves the subject of our "splining", but rather we used

$$X(r)/r^{|m-1|}, \; Y(r)/r^{|m-1|}, \tag{49}$$

for, as described in Appendix I, the leading term of the expansion of X and Y in powers of radius is $cr^{|m-1|}$, where m is the azimuthal modal number used frequently above, and c is some constant. Furthermore, we symmetrized the two functions (49) as they depend only on r^2 rather than on r, as also discussed in Appendix I.

One last point to be made concerning the splines is the obvious fact that they are linear functions of the displacements $X(r_i)$ and $Y(r_i)$. This allows us to write the interpolated displacements as sums:

$$\begin{aligned} X(r) &= \sum_{i=1}^{n} S_n^i(r) \, X(r_i) \\ Y(r) &= \sum_{i=1}^{n} S_n^i(r) \, Y(r_i) \end{aligned} \tag{50}$$

where we use the symbol $S_n^i(r)$ to indicate the interpolation curve obtained when we have started with a unit displacement of r_i and zero displacement at all the other n-1 radii, r_j, $j \neq i$. These expansions let us write the integrals above also as sums:

$$\begin{aligned} P(r) &= \sum_{i=1}^{n} \left[A_1(r,r_i)\, X(r_i) + A_2(r,r_i)\, Y(r_i) \right] \\ \frac{dP}{dr}(r) &= \sum_{i=1}^{n} \left[A_3(r,r_i)\, X(r_i) + A_4(r,r_i)\, Y(r_i) \right] \end{aligned} \tag{51}$$

where

$$A_1(r,r_i) = \int_0^\infty K_m(r,\rho) \left\{ -\frac{1}{\rho} \frac{d}{d\rho} \left(\rho\mu(\rho)\, S_n^i(\rho) \right) \right\} d\rho$$

$$A_2(r,r_i) = \int_0^\infty K_m(r,\rho) \left\{ \frac{m}{\rho} \mu(\rho)\, S_n^i(\rho) \right\} d\rho$$

$$A_3(r,r_i) = \int_0^\infty \frac{\partial}{\partial r} K_m(r,\rho) \left\{ -\frac{1}{\rho} \frac{\partial}{\partial \rho} \left(\rho\mu(\rho)\, S_n^i(\rho) \right) \right\} d\rho$$

$$A_4(r,r_i) = \int_0^\infty \frac{\partial}{\partial r} K_m(r,\rho) \left\{ \frac{m}{\rho} \mu(\rho)\, S_n^i(\rho) \right\} d\rho \ .$$

In a similar fashion, F_r and F_θ become

$$F_r(r) = -\sum_{i=1}^{n} \left(A_3(r,r_i)\, X(r_i) + A_4(r,r_i)\, Y(r_i) \right)$$

$$F_\theta(r) = \frac{m}{r} \sum_{i=1}^{n} \left(A_1(r,r_i)\, X(r_i) + A_2(r,r_i)\, Y(r_i) \right) \tag{52}$$

(e) Matrix Equations

Last to be considered in our development of the dynamical equations is the method used to minimize the error of approximation. The orthogonality properties of the Chebyshev polynomials led us to use a natural criterion for them: making the residual of equation (48), that is, the difference between the left-hand and right-hand sides, orthogonal to the first n Chebyshev polynomials. This can be accomplished by simple summation.

No similar property obtains for the splines, but the method of collocation (Hildebrand (1952)), or the setting of the residual equal to zero at n selected points, fits very naturally with the use of splines, provided, of course, that the set of collocation points is identical with the set of sampling points. If this proviso holds, we will be basing our spline curves directly on the values we compute at the collocation points.

There are surely many other choices of error criterion for spline-based approximations. Two other frequently-used criteria, minimax error and least-squares fitting, could be used, but they require

maxima hunting or additional integrations and so appear less desirable. However, no actual comparison of the relative accuracy of the different criteria was made, as collocation proved quite satisfactory.

To write the dynamical equations conveniently in our new form, let us first define several quantities in terms of the sampling points $\{r_i\}$:

$$\begin{aligned} X_i &= X(r_i) & Y_i &= Y(r_i) \\ \Omega_i &= \Omega(r_i) & \kappa_i &= \kappa(r_i) \\ A_{nij} &= A_n(r_i, r_j) & D_i &= D(r_i). \end{aligned} \tag{53}$$

Furthermore, two velocity variables, $\dot{X}$ and $\dot{Y}$, will be needed, which we define by

$$\begin{aligned} D_i X_i &= \dot{X}_i \\ D_i Y_i &= \dot{Y}_i . \end{aligned} \tag{54}$$

Then the dynamical equations in collocation form are expressed by:

$$\begin{aligned} D_i \dot{X}_i &= 2\Omega_i \dot{Y}_i + (\kappa_i^2 - 4\Omega_i^2) X_i + \sum_{j=1}^{n} (A_{3ij} X_j + A_{4ij} Y_j) \\ D_i \dot{Y}_i &= 2\Omega_i \dot{X}_i - \frac{m}{r_i} \sum_{j=1}^{n} (A_{1ij} X_j + A_{2ij} Y_j) . \end{aligned} \tag{55}$$

Finally, these last four equations can be written compactly in terms of a single vector, $\vec{Z}$, where

$$\vec{Z}^T = (X_1, X_2, .., X_n, \dot{X}_1, ... \dot{X}_n, Y_1, ..., Y_n, \dot{Y}_1, ... \dot{Y}_n).$$

The resulting equation is then

$$\omega \vec{Z} = M \vec{Z} \tag{56}$$

where M is a $4n \times 4n$ matrix composed of 16 nxn blocks:

$$M = \begin{pmatrix} m\Omega & -I & 0 & 0 \\ B_1 & m\Omega & B_2 & -2\Omega \\ 0 & 0 & m\Omega & -I \\ B_3 & -2\Omega & B_4 & m\Omega \end{pmatrix}.$$

In this matrix 0 denotes an nxn zero matrix, I is an identity matrix, and finally Ω is a matrix whose diagonal terms are Ω_i and the off-diagonal terms zero. Also

$$\begin{aligned} (B_1)_{ij} &= -A_{3ij} - (\kappa_i^2 - 4\Omega_i^2)\, \delta_{ij} \\ (B_2)_{ij} &= -A_{4ij} \\ (B_3)_{ij} &= \frac{m}{r_i} A_{1ij} \\ \text{and} \quad (B_4)_{ij} &= \frac{m}{r_i} A_{2ij} \end{aligned} \tag{57}$$

are the blocks holding the force terms.

Note that our dynamical matrix M is entirely real, which is the result of our particular choices of phase for the variables. This allows us to use considerably simpler programs to compute the eigenvalues of M than if the matrix were complex.

A greater simplification can be obtained in the axisymmetric case, where $m = 0$. Here we have $F_\theta = 0$, and so we can eliminate Y from Equation (25) to obtain

$$\omega^2 X = -\kappa^2 X + F_r , \tag{58}$$

and

$$F_r(r) = -\sum_{i=1}^{n} A_3(r,r_i)\, X(r_i)$$

as A_4 is also zero. If we write X as a vector:

$$\vec{X}^T = (X(r_1),\ldots X(r_n)) .$$

Thus we obtain a matrix equation for $\vec{X}$ immediately as

$$\omega^2 \vec{X} = N \vec{X} \tag{59}$$

where the nxn matrix N is defined by

$$(N)_{ij} = -A_{3ij} - \kappa_i^2 \delta_{ij}$$

This equation produces n eigenvalues for ω^2, or a total of $2n$ values for ω. Naturally, this corresponds to the loss of one of the two degrees of freedom in the dynamical equations, as Y can no longer be specified independently of X, but instead must obey

$$Y = 2\,\Omega\,X/\omega\ . \qquad (60)$$

(f) Equilibrium Model and Scaling

The only thing left unspecified thus far is the choice of the density, μ, for the equilibrium model, which in turn specifies Ω and κ. Several classes of models could be considered as candidates. There are a number of finite-radius models, where μ, Ω, and κ are algebraic functions, for example the models in Hunter (1965). These would be easy to use, but the presence of a sharp edge can only be expected to introduce new complications into the results. Since galaxies have no sharp edge, there is a risk that these edge effects might obscure something more fundamental. One would expect that Toomre's modification of the models (Hunter and Toomre (1969)) in which the surface density goes to zero near the edge like $(R-r)^{-n}$, where R here is the disk edge, might have removed some of the edge effects. However, there is no need to have any edge at all in our models.

Probably the most obvious choice would be an exponential density as the surface brightness of the outer portions of many galaxies is roughly exponential with radius, and this suggests that the density

might also be exponential:

$$\mu = \mu_c \exp(-r/R) . \qquad (61)$$

Here μ_c is the central density and R is now a scale radius. There is no major objection to this model, and we might well have used it; nevertheless, one could object that the mass in the model is rather widely dispersed. If we arbitrarily define the central portion of the model as that which contains the inner half of the mass, we find that it extends to 1.68R. The radius encircling 99% of the mass is quite far out, at 6.64R or 3.96 times the central disk radius. This is excessive compared with almost all other models. For example, the Maclaurin disk, which is the first member of Hunter's series of models, has 1.60 for the same ratio.

A family of more centrally concentrated models is that given by Toomre (1963), which have the added convenience that the density and rotation curves are simple algebraic functions. Yet the surface density decays only algebraically for large r, implying that too much mass stretches to large radius, as compared with the exponential density (61). However, the limiting member of their family, the gaussian disk does not suffer from this diffuseness. If we let

$$\mu = \mu_c \exp(-r^2/2R^2) \qquad (62)$$

we find the 99% radius at 2.58 the central disk radius, and furthermore,

it decays faster than the exponential model for large radius. In addition, it is also quite simple to use computationally, and so we chose to use it for all our calculations.

In order to make any computations at all, it is necessary to define in what units our variables are measured. The most convenient set of units for our problem is one based on the natural units inherent in the model itself. Besides a, the equilibrium model has three constants in it, R, μ_c, and G. In scaling we have the choice of three units, typically mass, length and time, and these three are exactly specified by choosing R, μ_c, and G. We will set $R = 1$, $\mu_c = 1$, and $G = 1/2\pi$. The latter value is just the reciprocal of the total mass of the disk, $M = 2\pi$, and was chosen so that the angular velocity would have the simple form $r^{-3/2}$ at large distances from the center of the disk. This choice of G also conveniently cancels the factor of 2π in the Bessel function integral representation of the kernel $K_m(r, \rho)$ of equation (40). All of the results and discussions which follow refer to this choice of units.

For reference, the rotation curve

$$V(r; a) = r\,\Omega(r; a) \tag{63}$$

and the surface density $\mu(r)$ are shown in Figure 4. As indicated by the notation, the velocity curve is also dependent on a, but as the figure demonstrates, that dependence is weak. We chose to let the

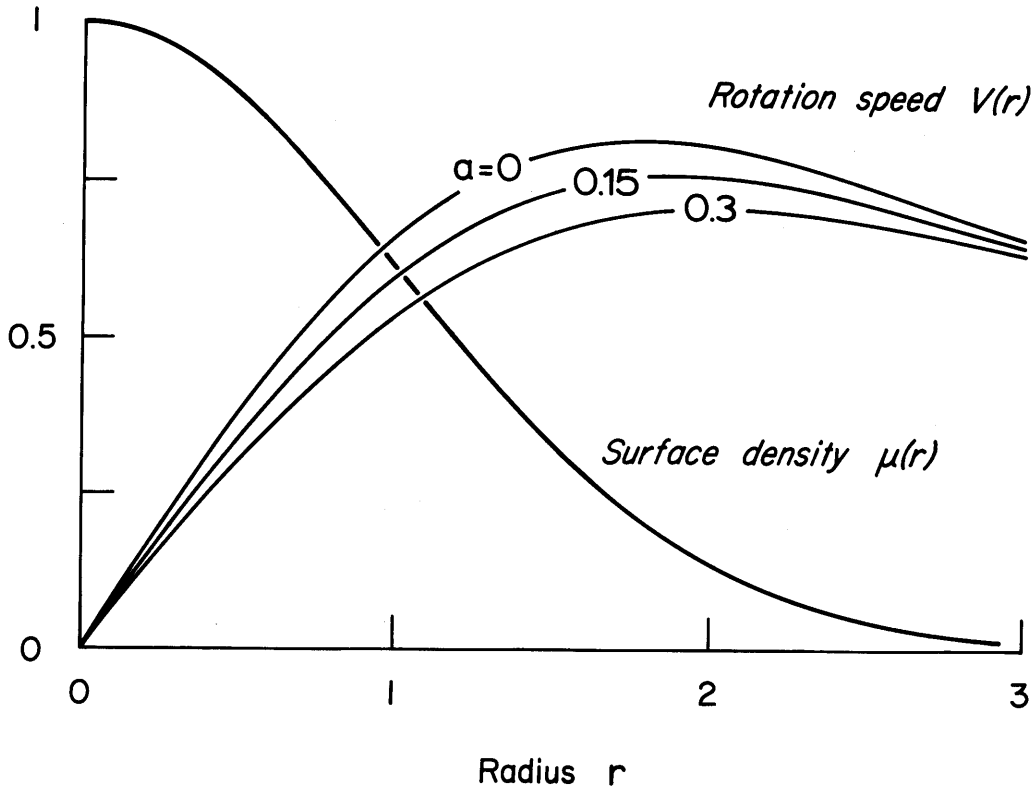

Figure 4. The linear speed of rotation $V = r\,\Omega$ and the surface density μ for the Gaussian disk, shown as functions of the radius r. The latter is independent of the offset a, but the rotation curves vary with a as shown.

equilibrium model depend on a just for self-consistency; technically, it could just as easily have been computed with $a = 0$ for all our work.

III. RESULTS OF AXISYMMETRIC CALCULATIONS

The simplest class of eigenmodes is obviously the one where all perturbations are independent of longitude. We concentrate here on these $m = 0$ disturbances, both to demonstrate the performance of our approximation schemes, and to show that axisymmetric stability in the large is indeed achieved upon sufficient modification of the gravity.

For $m = 0$, two elementary facts greatly simplify the presentations of spectra. For one thing, as noted already in Chapter II, axial symmetry implies that only $2\underline{n}$ nontrivial characteristic frequencies ω exist in modal calculations performed using n sampling radii, rather than the $4n$ frequencies found more generally. Moreover, these $2n$ frequencies occur in pairs differing only in sign, since they are merely the positive and negative square roots of the eigenvalues ω^2 of equation (57). In this chapter we can therefore describe the spectra simply in terms of ω^2.

The second simplification is closely related to the $m = 0$ version of the dynamical equation (48):

$$\omega^2 X(r) = \kappa^2(r) X(r) + \int_0^\infty \frac{d}{dr} K_0(r,\rho) \cdot \left[-\frac{1}{\rho} \frac{d}{d\rho} (\rho \mu X) \right] d\rho \qquad (63)$$

Upon introducing here the new variable $X'(r) = (r\mu(r))^{\frac{1}{2}} X(r)$ and integrating by parts, the resulting integral would have the symmetric kernel

$$(r\mu(r))^{\frac{1}{2}} (\rho\mu(\rho))^{\frac{1}{2}} \frac{\partial^2}{\partial r \partial \rho} (\rho K_0(r,\rho)).$$

Hence by arguments of Hilbert-Schmidt theory, all exact values of ω^2 must be real. Our approximations do not strictly preserve this symmetry implicit in equation (63), except in the limit as $n \to \infty$. Using finite n, it is conceivable that complex values of ω^2 could be obtained using our matrix scheme. However, in practice we have not discovered any instances where this nuisance occurred, even with values of n as low as two, provided the integrations were carried out with sufficient care. Thus the entire spectrum of n important eigenvalues for any axisymmetric case studied may conveniently be displayed on a single line.

(a) Results for Normal Gravity ($a = 0$)

We first show that the approximation machinery used here produces quite reasonable results for the traditional potential, meaning $a = 0$. Intuitively, if any axisymmetric case is likely to be poorly approximated it is this one, as the kernels of the force integrals there are singular. However, even here this is not a very worrisome problem, since the integrals themselves are well-defined -- it means only that

more than usual care is needed in calculating them, as we discuss in Appendix II. Indeed, the interpretation of the results is more of a problem.

i) Predictions

Before examining the computed data, let us first anticipate what results could be expected, both from previous work and from short wavelength theory. As noted before, a very similar study of the axi-symmetric perturbations of a cold disk was carried out by Toomre (1964). He used an algebraic disk model having no outer boundary, and therefore no edge effects, just as with our gaussian disk. Stated very briefly, his technique was to divide the mass of this disk into N discrete rings, which were allowed to move in accordance with the total self-generated gravitational force. As in our case, this procedure led to a matrix eigenvalue problem, with the solutions corresponding to the vibratory modes of the disk.

For his model, Toomre found modes both with well-defined frequencies, that is, virtually independent of the number of rings, and also with ill-defined frequencies. From this result and other work, he conjectured that the cold disk had a bipartite spectrum, containing both discrete modes with particular frequencies for $\omega^2 > 0$, and continuum modes, with $\omega^2 < 0$. If this were so, one might expect that in any computer experiment, an arbitrary selection of continuum frequencies would appear, the choice being dictated not by the model, but by the particulars of the numerical approximation used.

Later, an analytical proof of the existence of this type of bipartite spectrum was obtained by Hunter (1969) for a broad family of cold disk models of finite extent. His method was to express the perturbations as sums of Legendre functions, and to then obtain difference equations relating the different orders of these functions. By approximating the difference equations he was able to show that, provided the equilibrium density declines sufficiently rapidly with radius, a semi-infinite continuum of axisymmetric vibrations will exist, with upper bound

$$\omega^2 = \kappa^2(r_e).$$

Here r_e is the edge radius of the disk. By numerically solving the difference equations, Hunter further showed the existence of discrete modes with positive ω^2 for several models of the family.

Since our model does not have a finite edge, Hunter's results cannot be unhesitatingly used as predictions of our numerical calculations. However, the implication from his work is the same as that of Toomre's, that we might expect a continuous spectrum for negative values of ω^2, and a discrete spectrum for positive.

Short wavelength theory (SWLT) also implies the same prediction. Let us solve equation (3) for the wavenumber k :

$$|k| = (\kappa^2 - \omega^2)/2\pi G\mu \ ; \qquad (64)$$

this function is graphed for a selection of values of ω^2 in Figure 5. Since k may take equal positive and negative values, we show only the positive. Note that the curves in this figure are of two types. For non-positive squared frequencies, k grows positively infinite, while for positive squared frequencies, a radius r_L exists where $\kappa^2(r_L) = \omega^2$ and $k = 0$. At radii beyond r_L no short wavelength solution of equation (3) exists.

As implied by the figure, no boundary of any kind appears for non-positive squared frequencies, thus there is nothing to prevent an eigenmode from existing at any of these frequencies, and SWLT predicts a continuum. On the other hand, r_L is definitely a boundary of some sort for eigenmodes of positive frequency.

To determine the conditions at this boundary, we found it necessary to supplement SWLT, which is hardly valid at a point where wavenumber is zero, with the phase condition at r_L determined by a highly simplified model equation. As described in Appendix IV, a total phase change of $n\pi$ is required from center to resonance:

$$\int_0^{r_L} k \, dr = n\pi \qquad n = 1, 2, 3, \ldots \quad , (65)$$

at least in the sense of the SWLT approximation of completely separate short and long-wavelength oscillations.

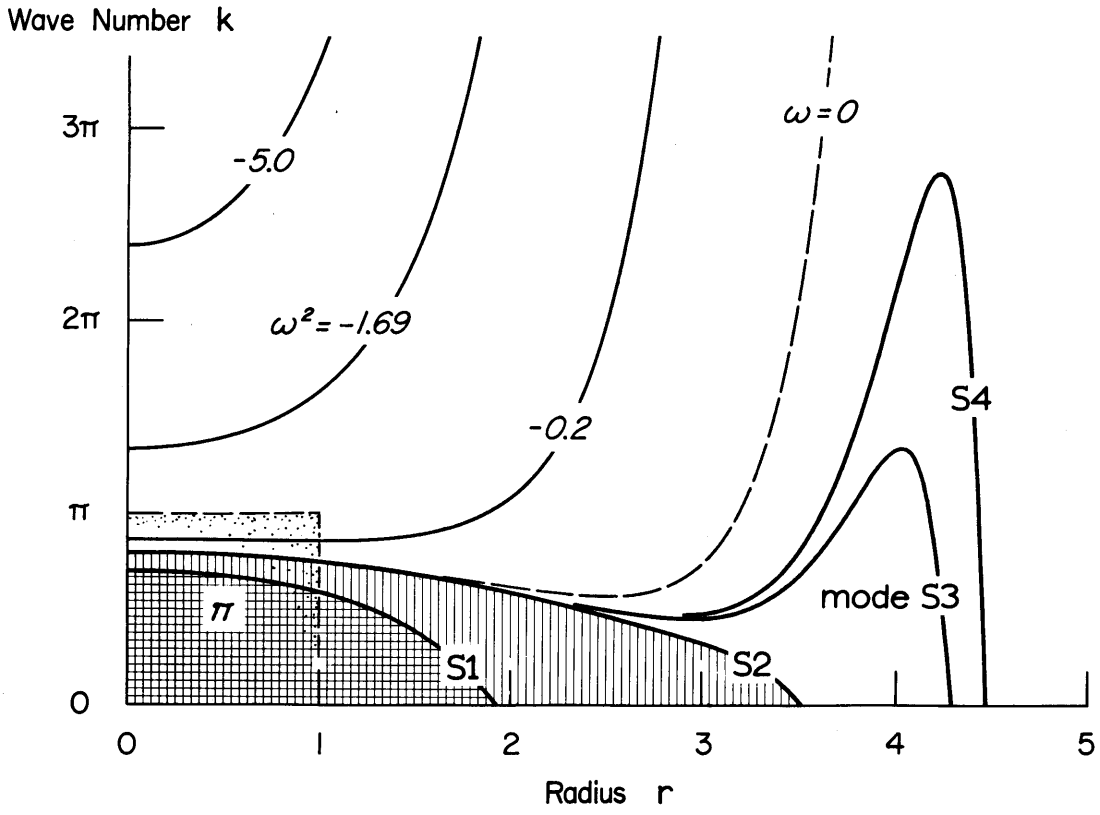

Figure 5. Wavenumbers $k(r;\omega^2)$ as functions of the radius r predicted from the axisymmetric dispersion relation with normal gravity ($a = 0$). The curves labeled S1, S2, ... correspond to the squared frequencies ω^2 predicted for the discrete modes; they bound an area measuring an integer multiple of π.

ii) Calculations

Ideally we would like to confirm or contradict the above predictions by calculating the complete spectrum of the exact equation (48). It would be hoped that if N increased toward infinity, our set of N approximate eigenvalues would converge towards those of the exact equation. Even granting this hope, we are faced with a problem dictated by practical limitations on our computing: We must extract the essential characteristics of the exact spectrum from sparse representations of it. Of course, we get around this obstacle in the only feasible manner, by intercomparing different approximate versions of the same spectrum and thereby gaining a familiarity with typical patterns and arrangements. These versions are generated by computing the same case (here $m = 0$, $a = 0$) with different numbers of samples, subject to a practical maximum, and by varying their location.

Our first choice of sampling points can be regarded as the most natural, that which divides the mass equally over the samples. To be precise we put the ith location at the radius encompassing $(i - 1/2)/N$ of the total mass, e.g. for 8 samples, the rings through the sampling points surround 1/16, 3/16, 5/16, ..., 13/16 and 15/16 of the total mass respectively, as is graphically shown in Figure 6.

The results for this arrangement of "rings" appear in Figure 7a, where the spectra for 2 through 15 samples are shown on the left side of the figure. We can immediately tell that our predictions are justified:

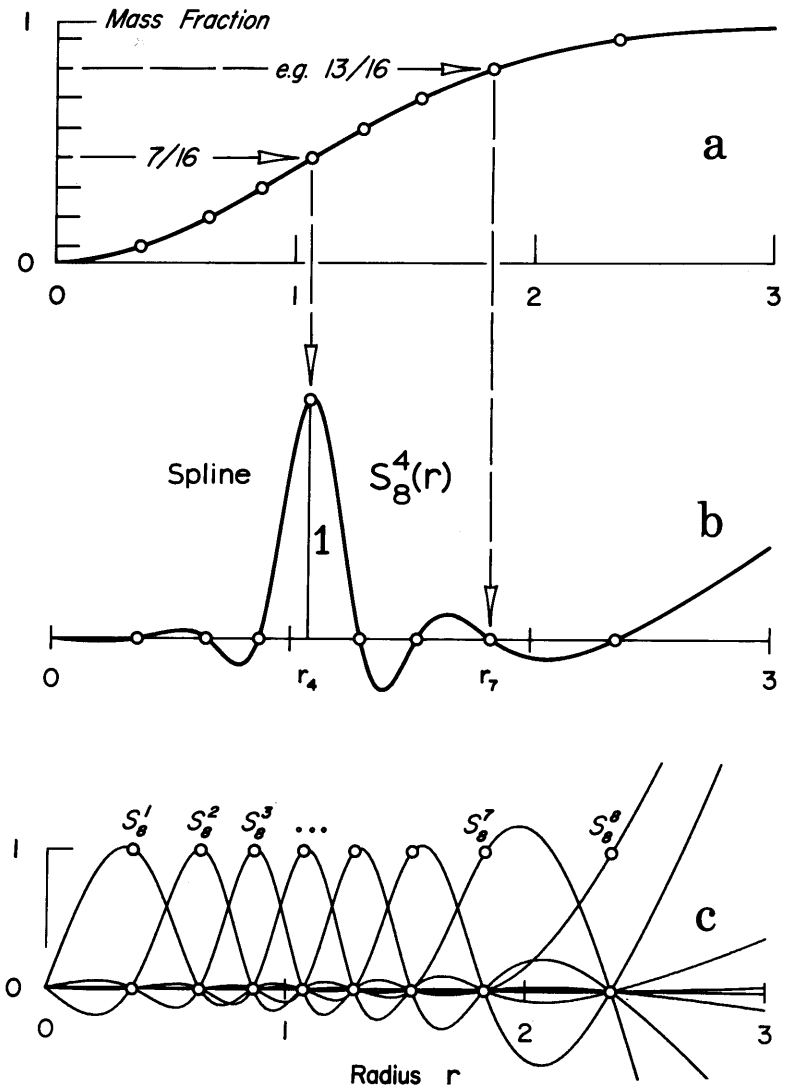

Figure 6. Demonstration of mass-spaced splines. In (a), radii which divide the disk into annuli with equal mass are shown as open circles; in (b) the curve for the fourth 8-point spline appears; in (c) we display the complete set of 8-point splines, normalized to unit height at the displaced radius.

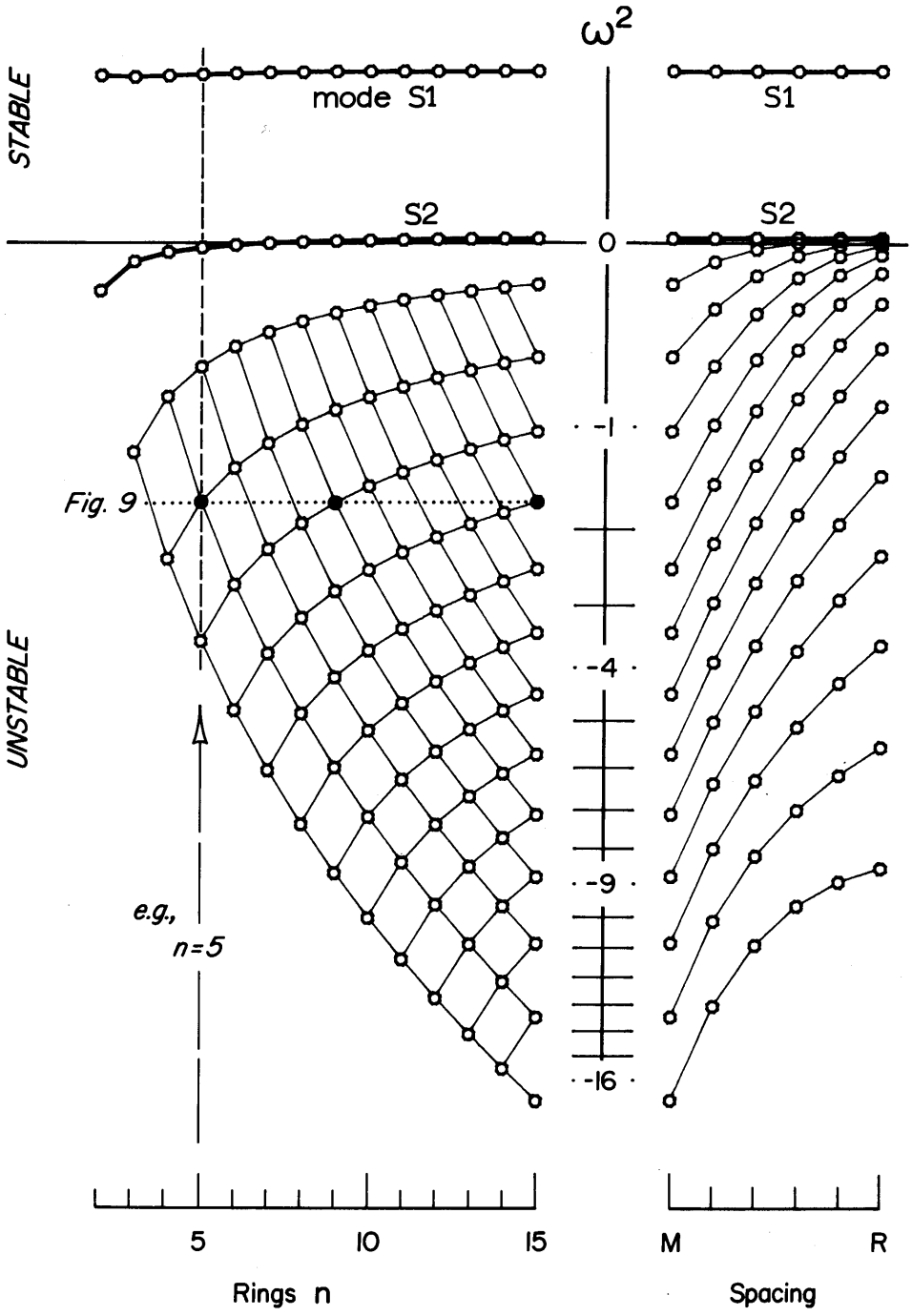

Figure 7. On the left side the eigenvalues ω^2 for mass-spaced sampling rings are shown. The number of rings used varied from 2 to 15. On the right side the eigenvalues for 15 mass-spaced rings are repeated in the left column, in the rightmost column the eigenvalues for rings equally spaced with overall extent 4 are given. The columns between these correspond to five equal shifts of the ring locations from mass-spaced to radially-spaced.

Two classes of modes are clearly identifiable for larger N. As expected, there are both positive eigenvalues which are nearly independent of N, and also a band of negative eigenvalues that stretches from near zero down toward negative infinity.

Note that the scale had to be highly non-uniform to clearly accommodate both the large negative values and the small values near zero. We plotted the vertical ω^2 scale by adjusting the heights by the rule $(0.5 - \omega^2)^{\frac{1}{2}}$.

The squared frequency of the first isolated mode is actually quite precise even for small N, as seen in Table 1.

TABLE 1

N	ω_1^2
2	.29909
3	.29797
4	.29895
5	.29978
10	.30141
15	.30182
30	.30211

Table 1. Eigenvalues of the first discrete mode for $a = 0.0$, computed with varying numbers of sampling points.

This is to be compared with a value of .3132 predicted by SWLT. As noted below, the second eigenvalue is much less precisely determined by this chain of experiments.

These results were corroborated by a second series of experiments in which the sampling locations were gradually shifted from experiment to experiment. Rather than have the sampling points equally spaced in mass, we moved them in five equal steps to points equally spaced in radius, located at

$$r_i = 4(i - \tfrac{1}{2})/N \ . \qquad (66)$$

In Figure 7b we show the resulting eigenspectrum at each of these steps. The leftmost column of eigenvalues is the same as the rightmost column of Figure 7a: the 15 sampling points are equally spaced in mass. The second column of eigenvalues is the spectrum obtained by moving each ring 20% of the distance from its original location to its final location as given by equation (66). The third column refers to a 40% shift and so on.

Again the discrete frequencies stay fixed and the continuum frequencies flow to new positions in the spectrum. Each step causes all of the continuum eigenvalues to move to smaller frequencies. The reason for this can be deduced from our Figure 5. As we move the sampling locations outwards and simultaneously increase the spacing between adjacent locations, we reduce the maximum value of k or equivalently increase the minimum wavelength that our sampling can mimic at any radius. And Figure 5 shows that lower values of k at any radius correspond to smaller values of ω^2; thus increasing the spacing moves us to lower ω^2.

Furthermore, the same value of k at a larger radius will in general relate to a smaller value of ω^2, so, moving the sampling rings outwards also will move the spectrum to lower ω^2. Both of these effects combine to produce the large frequency shifts shown in Figure 7b.

Although difficult to distinguish in Figure 7b, sliding the rings outward has also caused more stable discrete eigenvalues to appear. In Figure 5 we deliberately chose to show the values of k corresponding to the positive discrete mode frequencies for the first four modes of this type predicted by theory. Since our mass-spaced set of samples did not extend even to $r = 3$, and these modes extend beyond $r = 4$, it is little wonder that the third and fourth discrete eigenmodes were entirely missed by our first series of experiments. Extending the set of samples to 4.0 did cause them to appear but to obtain an accurate description of their shape it was necessary to calculate with samples extended to $r = 5.5$. For this same reason, even the second stable frequency was poorly determined by the mass-spaced samples.

The eigenfunctions of the first three discrete eigenmodes are displayed in Figure 8. As we would expect, they are very similar to Toomre's eigenfunctions in several respects. Just as with a great number of physical systems, such as strings, beams, plates, etc., they have respectively zero, one and two modes. Furthermore, the highest peak is the outermost one, and the boundary radius, r_L, is located

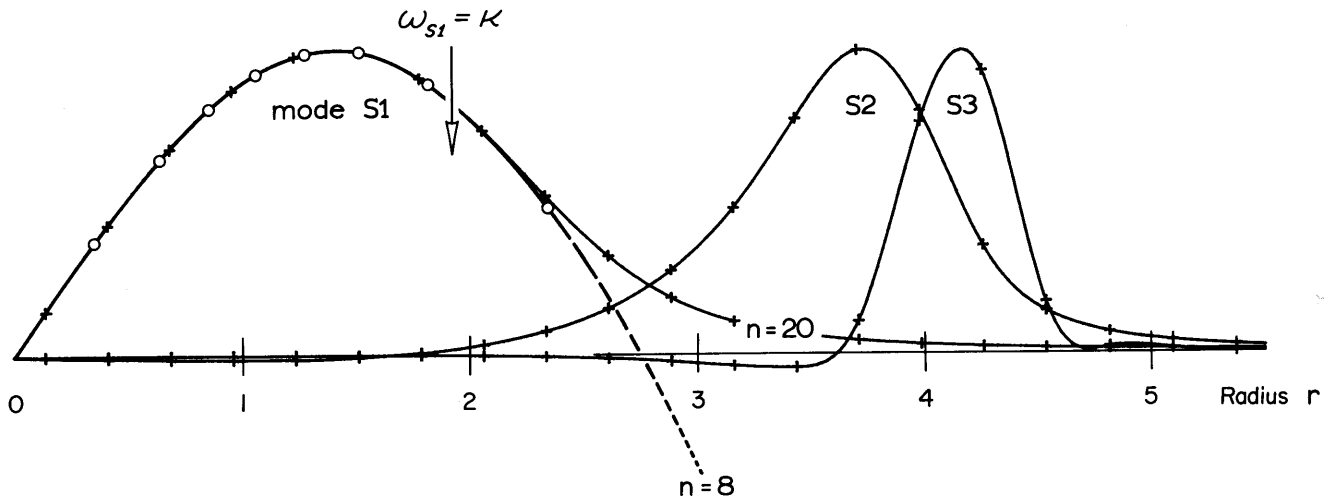

Figure 8. The first three discrete stable eigenfunctions for normal gravity (a = 0). All were computed with 20 sampling points, equally spaced out to 5.5. For comparison, an 8-ring mass-spaced eigenfunction for S1 is also shown.

on the downslope of this latter peak. This of course agrees with the model equation prediction of a $\pi/4$ phase at resonance.

The last thing we want to mention in the normal gravity case is the form of the unstable, continuum eigenfunctions. As can be seen in the pyramid of eigenvalues appearing in Figure 7a, squared frequencies very close to $\omega^2 = -1.69$ appear coincidentally at $N = 5$, 9, and 15 sampling locations, as well as for $N = 23$, which is not shown in Figure 7a. In Figure 9 we display the computed eigenfunctions corresponding to these eigenvalues.

The upper eigenfunction is a rather simple curve, as it must be with only five sampling points. Progressing downwards, the lower curves have successively more nodes and there is certainly little indication from the presented data that the last curve shown is a good representation of the true eigenfunction. On the contrary the superposition of the last two functions shows that the sequence has far from converged for radii beyond 2.5. Since SWLT predicts an unbounded number of nodes, stretching out to infinite radius, our computed results might optimistically be considered a mild corroboration of the theory. However, we have a much better indication of the accuracy of SWLT from a comparison of the computed nodal positions with the theoretically derived positions. As noted in Appendix III, a first approximation to the theoretical nodal positions can be found from the zeroes of the Bessel function $J_1(\alpha(r))$ where $\alpha(r) = \int_0^r k(s)ds$ and the wavenumber $k(r)$ is obtained

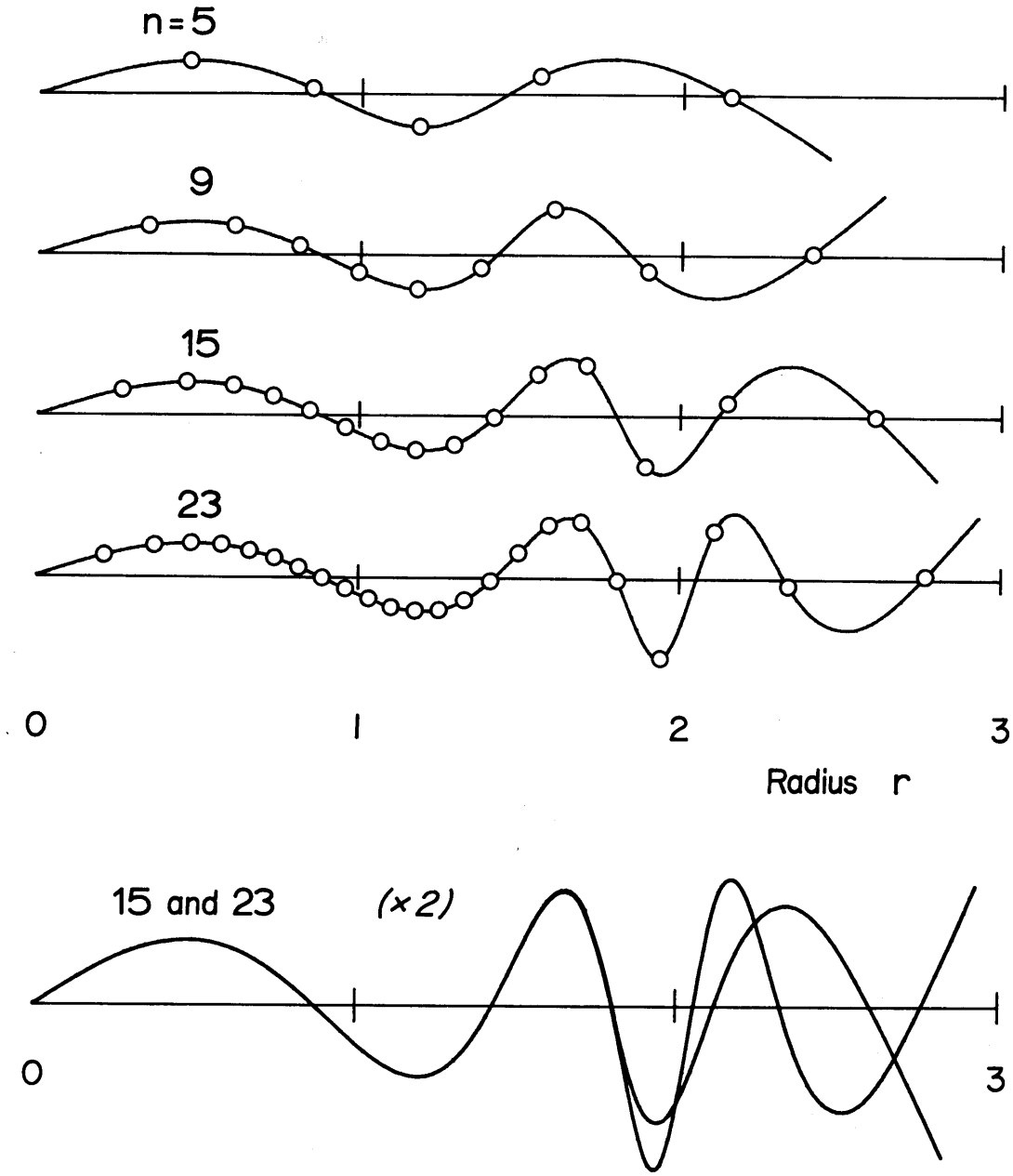

Figure 9. Representations of an unstable continuum eigenfrequency with squared frequency $\omega^2 = -1.69$. In the upper curves, 5, 9, 15, and 23 mass-spaced sampling points were used, the latter two are overlaid and expanded in scale in the lower portion of the figure to facilitate comparison.

from the dispersion relation. Table 2 shows that at least the first 4 nodes are well-defined by our 23-sample computation.

TABLE 2

Node	Theory	Computation
1	.8713	.8674
2	1.4320	1.4271
3	1.8038	1.8000
4	2.0482	2.0535

Table 2. Nodal radii of the continuum mode at $\omega^2 = -1.69$, as predicted by SWLT and as computed with 23 mass-spaced samples.

To find the computed nodes for this table, we first found the nodes at the exact frequency determined in the computation and then adjusted the radii according to SWLT -- amounting at most to a slide of .001. The location of even more nodes could be obtained from calculations with large numbers of samples; however, the point is already well made by the data at hand -- that the wavefunction seems to be fulfilling the predictions of SWLT and oscillates endlessly to infinity.

(b) Results for Modified Gravity ($\alpha \neq 0$)

As noted in the introduction, the idea of using the modified gravity in a cold disk model stems from the hope that at least the axi-

symmetric waves could all be stabilized for large enough a. Before checking this surmise, we will examine in detail the modifications to the eigenmodes that are caused by making a nonzero, both in SWLT prediction and in computational fact.

i) Predictions for a Sample Case ($a = .15$)

Let us start our examination of the effects of a with a case still possessing instabilities, $a = .15$. As in the introduction, our SWLT predictions will be wholly on the basis of equation (5), the dispersion equation for nonzero a:

$$\omega^2 = \kappa^2 - 2\pi G \mu k \exp(-ak) .$$

First we wish to discuss modes where the wavenumber k is real. We can immediately see that this requirement implies that for any radius r, there are both maximum and minimum values that ω^2 can take, namely

$$\omega^2_{max} = \kappa^2(r) \qquad \omega^2_{min} = \kappa^2(r) - 2\pi G\mu(r)/ae. \qquad (67)$$

Here we have used the fact that the maximum of $k \exp(-ak)$ is $1/ae$, as illustrated in Figure 10.

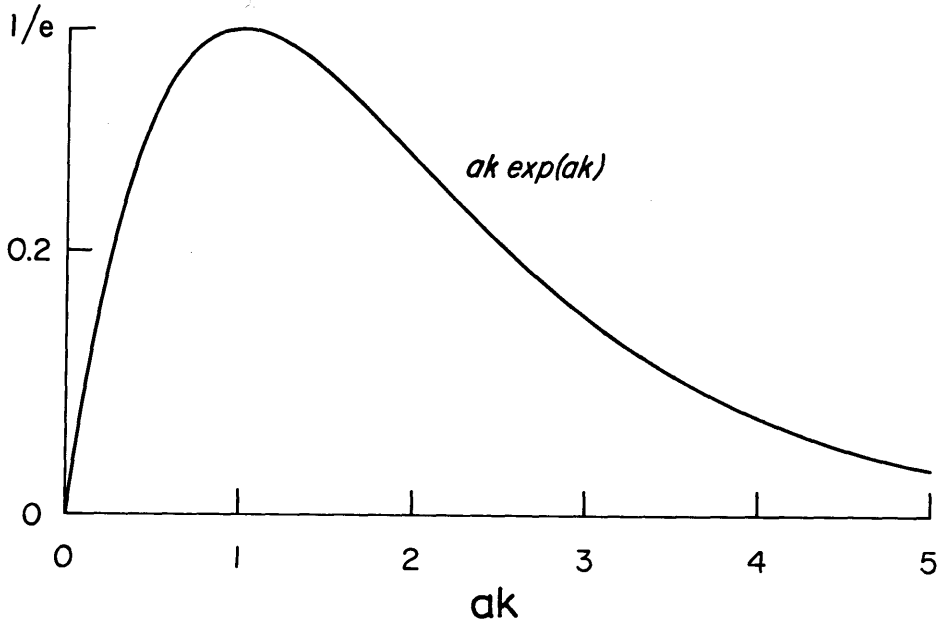

Figure 10. The shape of the curve ak exp(-ak) which appears in the local dispersion relation. The value $k = a^{-1}$ divides the short-wave from the long-wave solutions.

In this thesis, we will refer to these two $\omega^2(r)$ curves as resonance and cutoff respectively. This nomenclature differs distinctly from that used in plasma physics (cf. Stix (1962)). In plasma physics, resonance refers to a point where k becomes zero, and cutoff, where k becomes infinite. Both of these possibilities occur at the same radius in our galactic dynamics, and so we can make do with the single term resonance for this point. The superfluous word cutoff we borrow for a radius at which k is maximum. Incidentally, plasma dispersion relations for omnidirectional propagation do not have a maximum, but instead are monotonic, much like our $a = 0.0$ case.

These two special curves (67) have been plotted in the upper diagrams of Figure 11; the leftmost one corresponds to $a = 0.15$. As just noted, real values of k can exist only in the area between these two curves. But we can interpret this diagram in another way. In looking for eigenmodes of the disk, we are concentrating on solutions with one fixed frequency. Figure 11a shows us that, for any value of ω^2 in the allowed region, real solutions of the dispersion relation (5) can only extend from the center, $r = 0$, out to some outer radius.

Evidently there are three kinds of solutions of this general type, one for positive squared frequencies less than $\kappa^2(0)$ and greater than the maximum value having a cutoff, ω_x^2, which is here .0124; these solutions extend from center to cutoff. A second kind exists for negative squared frequencies greater than ω_x^2; these stretch out to resonance. A third kind exists for the remaining values of ω^2 :

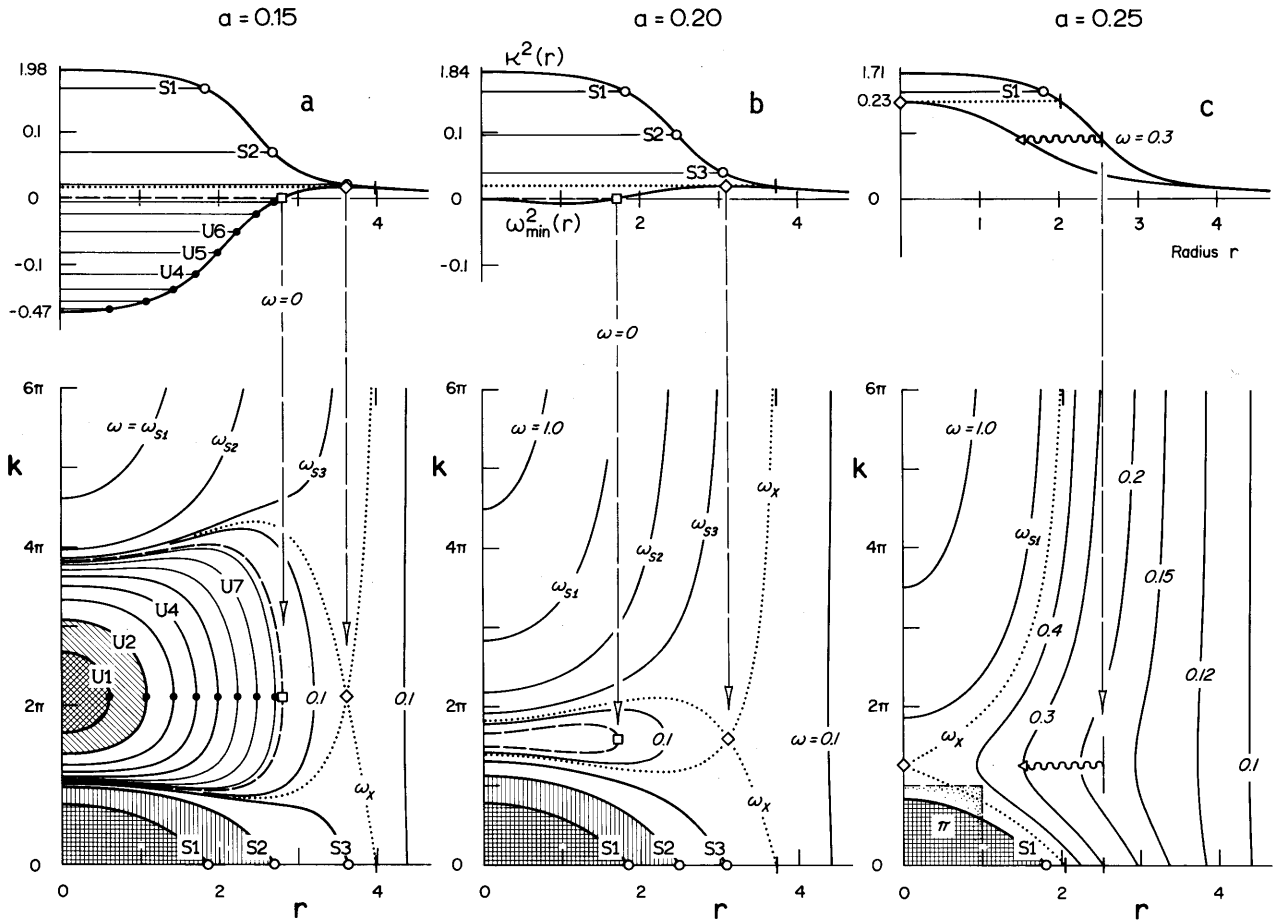

Figure 11. Graphical presentation of the dispersion relation for values of a near the critical value. The upper curves display $\omega^2_{max} = \kappa^2$ and ω^2_{min}, and the lower curves show the wavenumber $k(r;\omega^2,a)$ as predicted by the dispersion relation for selected values of ω^2 and for (a), left, a = 0.15; (b), center, a = 0.20; (c), right, a = 0.25. Curves labeled S1,..., and U1,... correspond to predicted discrete modes. The dotted curve labeled ω_x corresponds to the maximum value of ω_{min}. In (c), the extent of the allowed region is shown for $\omega = 0.3$ by a wavy line.

$0 < \omega^2 < \omega_x^2$. For eigenvalues ω^2 in the third band, real values of k can be found in two separated regions, one from the center out to cutoff, and another from a second cutoff radius out to resonance.

A better viewpoint for these three categories is provided by the kr plane. For each point inside the allowed region in Figure 11a, there are two values of k satisfying the dispersion relation, one with k greater than a^{-1}, one with k less than a^{-1}, as was already clear in the plot of ke^{-ak} in Figure 10. Let us draw the curves of k versus r corresponding to a fixed frequency in the lower part of Figure 11. There the value $k = a^{-1}$ divides the two solutions; above it appear the short wavelength solutions, called for convenience s-waves, and below it the long wavelength solutions, the l-waves. Note that for squared frequencies less than ω_x^2, the s-waves join the l-waves at this value of k, which is of course the translation into k and r of the lower curve of Figure 11a, the cutoff.

In contrast, eigenvalues above ω_x^2 have separated solutions which never join, but instead go off to $k = \infty$ (s-waves) or to $k = 0$ (l-waves). These two limiting values of k are the two translations of the single resonance curve of Figure 11a. Taken as a whole, this kr diagram is the analogue of Figure 5 for non-zero a.

With this detailed plot before us, the most prominent question is that of the existence of discrete and continuum modes, in analogy

with the discussion centering around Figure 5 for $a = 0$. Each type of curve in our new kr diagram can be characterized as a separate category of eigenmode. We will treat them sequentially.

1. S-modes

The lowest set of curves, comprised of 1-waves with squared frequency greater than ω_x^2, is structurally identical with the k-curves for positive ω^2 for the case of $a = 0$. The boundary conditions for 1-waves at the center and the resonance radius r_L again produce the requirement that

$$\alpha_1(r_L) = \int_0^{r_L} k_1(s)\, ds = n\pi ,$$

and therefore we should expect discrete frequencies to appear. Here we have used the subscript 1 to refer to the 1-wave solution of the dispersion relation.

Unlike the case of $a = 0$, only a finite number of S-modes can exist when a is non-zero. In the former case, an arbitrarily large number of discrete modes could be found by lowering the squared frequency toward zero. Here the S-mode spectrum is bounded below by ω_x^2, and the largest value of $\alpha_1(r_L)$ that can occur is that obtained for $\omega^2 = \omega_x^2$.

2. C-modes

The upper set of curves is the s-wave analogue of the curves representing the S-modes. In each of this upper set k rises to infinity at $r = r_L$. As we discuss in Appendix IV, there is no phase condition on α_s at r_L, nor is there any requirement for a node. Thus theory predicts that, instead of discrete modes, a continuum of s-wave eigenmodes, extending from ω_x^2 to $\kappa^2(0)$, will overlay the discrete spectrum of s-modes.

3. U-modes

Let us now study the modes corresponding to the U-shaped curves which occur for squared frequencies below ω_x^2. To the right of the crossover point in Figure 11 there are curves for ω^2 between 0 and ω_x^2, each of which can be associated with a curve of the same squared frequency among the U-shaped curves. These pairs of separated curves directly correspond to values of ω^2 which have no associated curve anywhere in the kr-plane. We will refer to only these modes as U-modes, as they alone are unstable. From an asymptotic analysis in Appendix IV, we found that the boundary condition for U-modes involved the difference of the phase of the 1-wave and s-wave at cutoff, r_c :

$$\alpha_s(r_c) - \alpha_1(r_c) = (n + \tfrac{1}{2})\pi \qquad n = 0, 1, 2, \ldots \quad (68)$$

where α_s and α_l are the phases of the s- and l-waves respectively. Strictly speaking the term wave is inapplicable here, as we are dealing with disturbances with purely imaginary frequency either growing or decaying in time.

The boundary condition (68) dictates that all unstable modes will be discrete. In graphical terms equation (68) says that the only unstable modes predicted to exist by SWLT are those which enclose an area of $(n - \frac{1}{2})\pi$ in the kr plane, such as those pictured in Figure 11.

This, of course, is very much like the requirement for the S-modes, which is that the corresponding curves in the kr plane enclose $n\pi$. The similarity between the two graphical interpretations should not obscure the actual differences between the role of the cut-off radius and that of resonance. The former serves to couple the long and short waves, rather than providing a boundary condition on either one of them alone. Thus we can also predict that the eigenfunctions for U-modes will be very different from those of S-modes.

4. T-modes

Matters are a bit more complicated for modes with ω^2 in the range from 0 to ω_X^2. Here two cutoff radii exist, r_{c1} and r_{c2}, as well as the resonance radius, r_L. Between the two cut-offs no real k can satisfy the dispersion relation, and we will therefore have a mixture of spatially growing and decaying waves there,

and in the two "allowed" regions, $0 < r < r_{c1}$ and $r_{c2} < r < r_L$, a mixture of s-waves and 1-waves will occur. This type of behavior cannot be studied with the model equation of Appendix IV, which has only one cutoff owing to the use of a linear $\kappa^2(r)$. Furthermore, except for a small band of frequencies just below ω_X^2, SWLT is inapplicable in the inter-cutoff annulus as no solution of the dispersion equation exists, even with complex k. For these reasons, and with the justification that very few modes occur in this region, we neglect quantitative prediction of the phase relations for T-modes. However, there should be little qualitative difference in the wavefunctions between C-modes and T-modes, inasmuch as the inter-cutoff region is generally quite narrow.

ii) Calculations for the Sample Case (a = .15)

To check our SWLT predictions for discrete and continuum modes, we performed an extensive series of experiments very similar to those done in the normal gravity case. Again, theory and experiment complement each other. On the right of Figure 12a, the spectrum of eigenvalues for 2 to 20 mass-spaced sampling points is presented. Both U-modes and S-modes are instantly visible against the background of continuum modes. The same is true for the right columns of eigenvalues in Figure 12a, in which we show the results of a complementary set of experiments, where we gradually slide the 20 sampling points from their mass-spaced positions into radially-spaced positions extending out to $r = 4.0$,

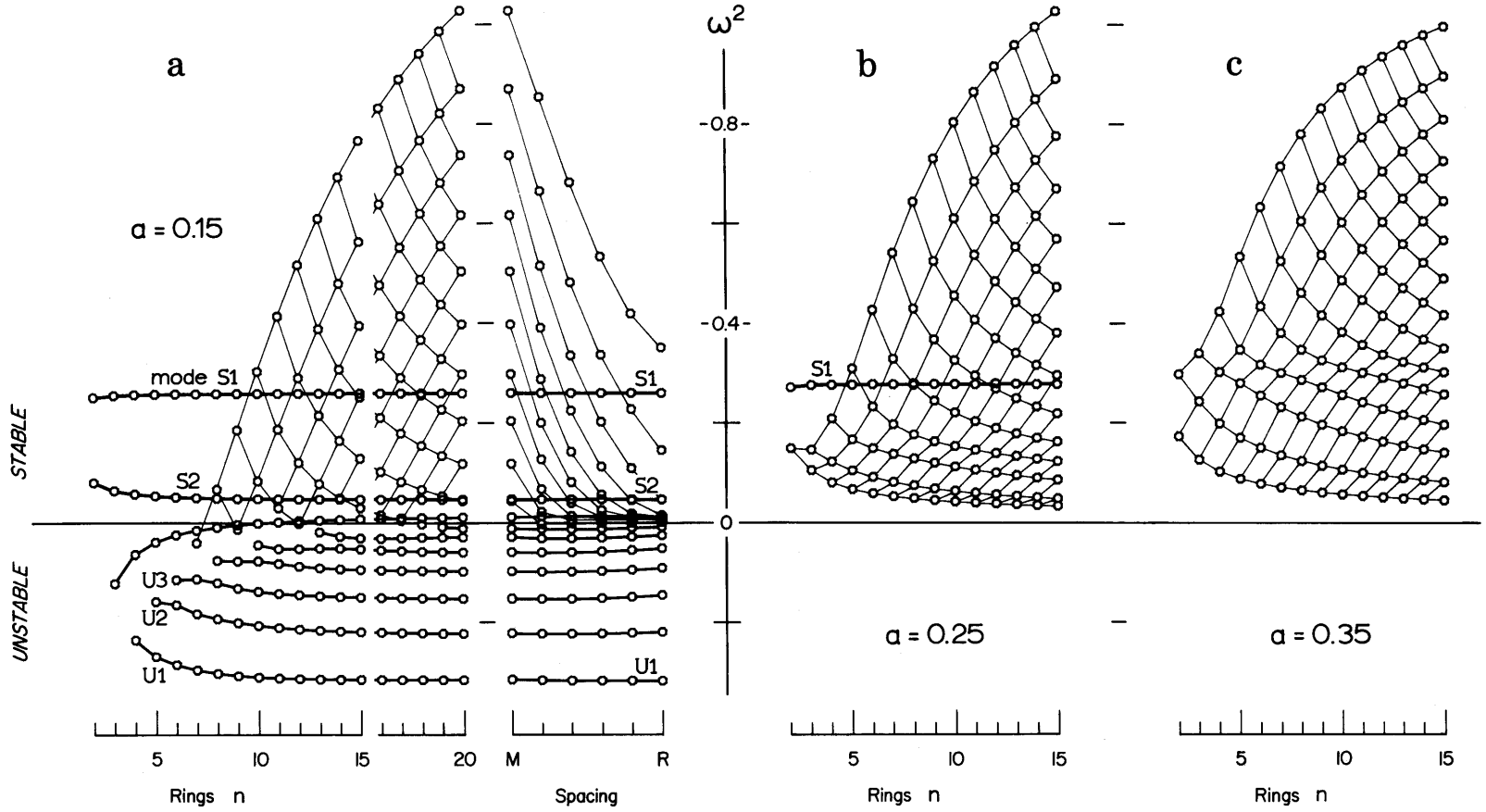

Figure 12. Calculated eigenvalues for three values of offset a. On the left eigenvalues for 2 to 20 mass-spaced rings are shown; adjacent to it are the eigenvalues for gradual ring shifts to equally-spaced samples, just as in Figure 7. In (b) and (c) experimental results are only included for 2 to 15 mass-spaced sampling rings. The vertical scale for ω^2 is linear in this figure.

just as we did in the former set of experiments with $a = 0.0$ shown in Figure 7. At least eleven of the discrete modes have well-defined frequencies which we can check against our SWLT predictions.

TABLE 3

S-modes		U-modes	
Predicted	Calculated	Predicted	Calculated
.3308	.3245	-.4001	-.3973
.0581	.0584	-.2839	-.2744
.0157	.0159	-.1942	-.1812
.0124	.0125	-.1265	-.1134
		-.0773	-.0641
		-.0425	-.0314
		-.0191	-.0100

Table 3. Comparison of SWLT predictions with the results of computer experiments. The squared frequencies for all discrete modes of $a = .15$ detected in the experiments are shown.

As shown in Table 3, the results are especially good for the stable modes. Also, in this particular case, no T-modes were found.

Let us now turn to the eigenfunctions corresponding to the various classes of modes. In Figure 13 we see the first four U-modes. The lower two give ample evidence of the mixing of s-waves and l-waves which takes place in this category of mode. The small arrowheads in the figure denote the cutoff radius, r_c; unlike the stable modes, the U-modes continue to have nodes beyond their outer boundary. This can

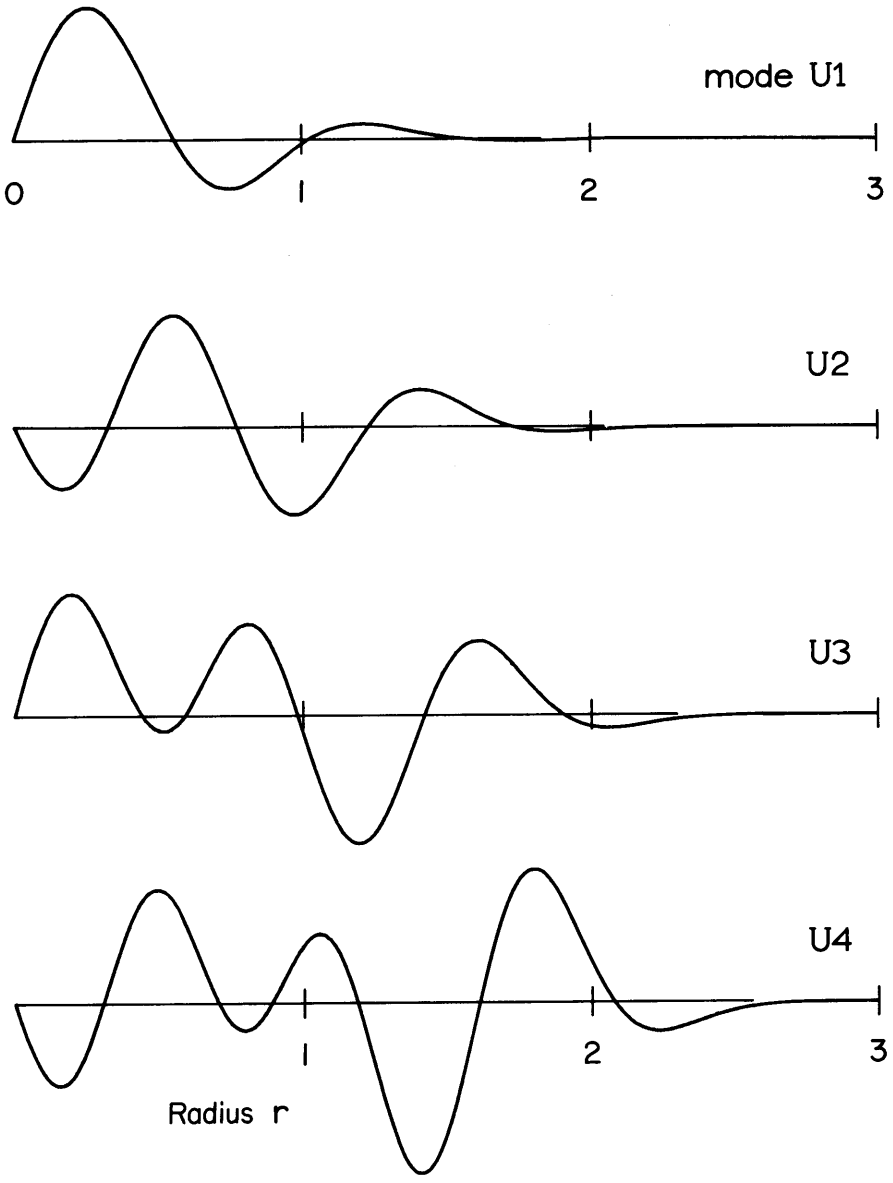

Figure 13. Eigenfunctions of the four most unstable modes for $a = 0.15$.

intuitively be understood in terms of the values of k at the boundary. For a resonance k is zero, and phase changes should have stopped; but for a cutoff k is at its maximum, $1/a$, and phase is still changing rapidly there. Indeed a visual inspection of the four curves shows that k in the decaying portions is indeed about $a^{-1} \simeq 6.66$; or in terms of wavelength $\lambda \simeq 0.94$.

Continuing our examination of the new classes of eigenfunctions, we show in Figure 14 five representations of a continuum wavefunction, a C-mode, for $\omega^2 = 0.50$. These five curves were found by using radially-spaced sampling points and adjusting the ring extent, r_e, so that one of the eigenfrequencies exactly coincided with $\omega^2 = 0.50$. For the first four curves, the ring extent was in the vicinity of $r = 3$, but for the last curve we moved r_e into about 2.3 to provide a higher concentration of samples over the important radial range of the eigenfunction, while still computing with a total of 30 samples.

Again SWLT is an excellent predictor of these wavefunctions. Just as with $a = 0.0$, we compute the SWLT-predicted nodal locations, and contrast them with the actual experimental nodes. As shown in Table 4 the agreement is excellent.

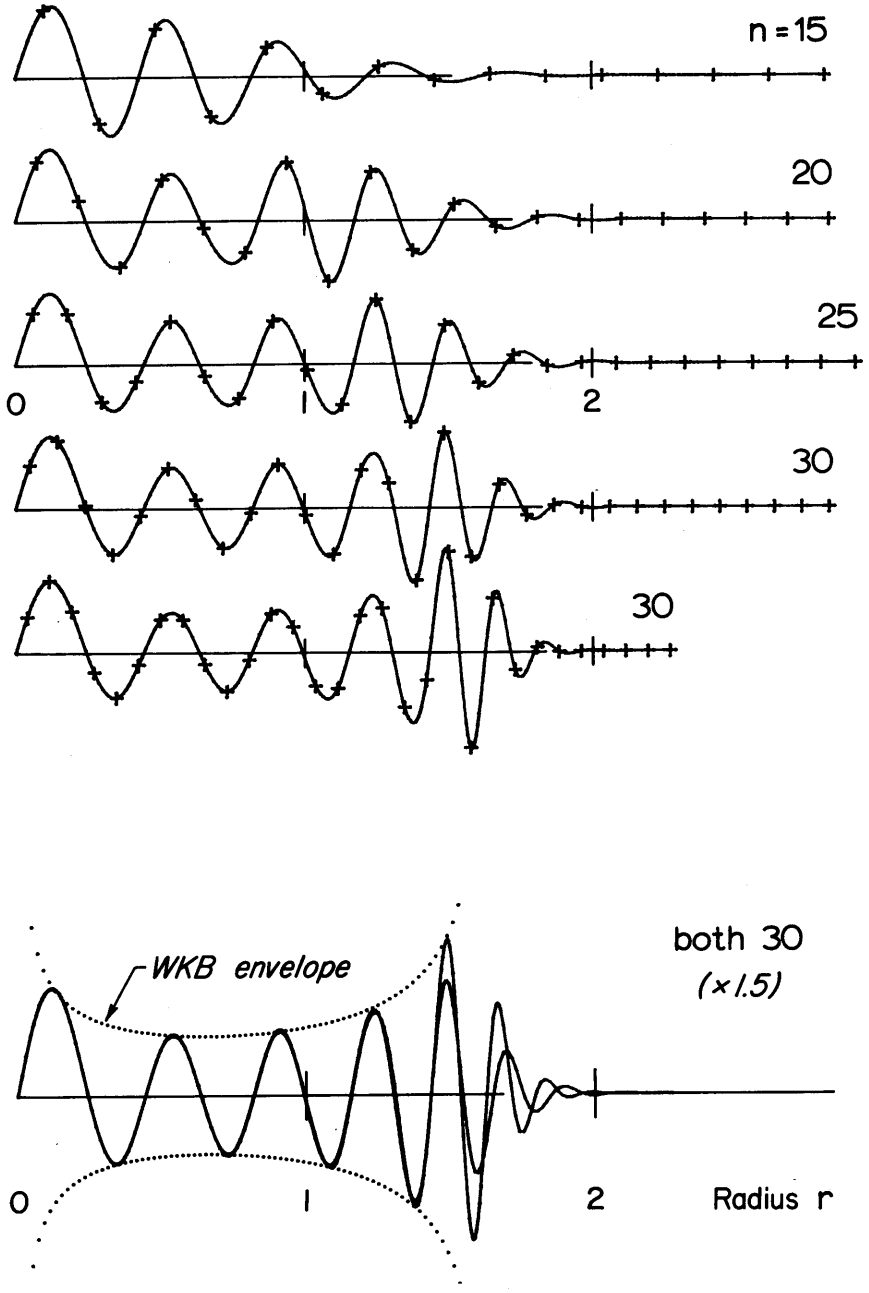

Figure 14. Five representations of the stable continuum wavefunction with $\omega^2 = 0.50$. Decreasing interring spacings were used in the upper sequence of curves to improve the accuracy of the wavefunction; sampling points are denoted by + . In the lower curve the two best curves are overlaid and expanded; superimposed is the prediction of the envelope height à la WKB.

TABLE 4

N and r_e

Node Number	15 2.905	20 2.986	25 2.973	30 2.870	30 2.305	WKB
1	0.237	0.242	0.243	0.242	0.243	0.243
2	0.427	0.445	0.443	0.442	0.442	0.441
3	0.622	0.639	0.636	0.635	0.634	0.633
4	0.818	0.842	0.826	0.820	0.818	0.817
5	1.018	1.014	1.003	0.996	0.993	0.991
6	1.217	1.175	1.172	1.158	1.155	1.152
7	1.419	1.331	1.314	1.310	1.303	1.299
8	1.616	1.485	1.445	1.434	1.434	1.430
9	1.821	1.637	1.576	1.545	1.536	1.540
10	2.012	1.788	1.703	1.651	1.627	1.624

Table 4. Comparison of SWLT nodal positions with calculated positions, computed with $a = 0.15$, $\omega^2 = 0.50$, and varying numbers of sampling points and extents.

In this example we can go even further with SWLT predictions. In the lower plot of Figure 14, the envelope of the amplitudes as generated by the group velocity arguments of Appendix III are overlaid as dotted lines on the final wavefunction. Peak heights match the envelope heights to within a few percent.

iii) Achievement of Stability

An initial rationale for the use of reduced gravity was that it made possible the achievement of stability in the local axisymmetric

sense. In Chapter I, we made an initial estimate of the value of a needed to achieve stability by finding the particular radius most unstable and then choosing a_{crit} to just stabilize that position. For our model this value is 0.2016. But the calculations with $a = 0.15$ demonstrated that the unstable modes do not form a continuum as implicitly required by this argument, but rather comprise a discrete set. This will alter our predictions.

Since a discrete mode requires that some sizable interval in radius be unstable we can expect that the finding of discrete U-modes will lower the SWLT estimate of the value of a needed to stabilize our model. While this is true, it is a peculiarity of the gaussian model that, near $a = a_{crit}$, a disk out to about $r = 2$ stabilizes locally almost simultaneously in a, as is visible in the kr diagram of Figure 11b. There the region containing the unstable frequencies has deflated to a thin area measuring less than one-half π. Thus at this value, $a = 0.2$, the last unstable U-mode has already just disappeared. However, since the region from center to about $r = 2$ changes from stability to instability in unison, we would expect that only a small decrease in a would expand this area to one-half π thereby implying that the newly-predicted a_{crit} is only slightly less than the locally determined a_{crit}.

To find the new a_{crit} we must determine the value of a for which ω^2 becomes equal to zero for the last remaining unstable mode. In Figure 15 we show the traces made by the squared frequency

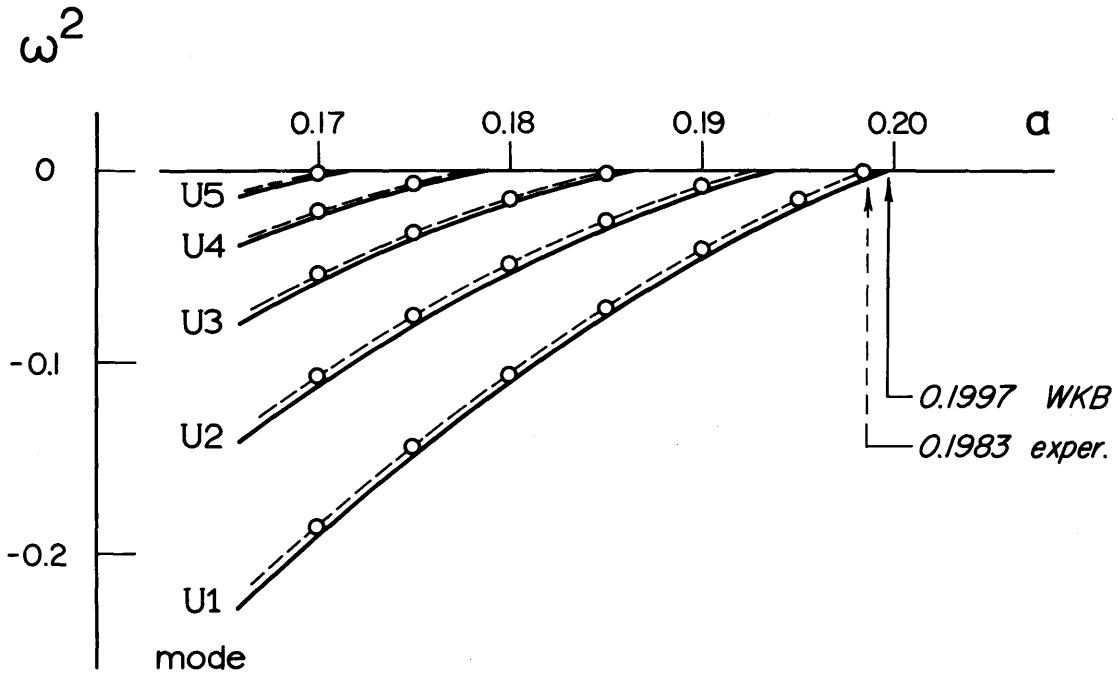

Figure 15. The approach to stabilization of the first five U-modes. The solid curves indicate the SWLT-predicted values as a function of a . The open circles mark experimental data points.

of the U-modes as a is varied, according to SWLT. The last unstable mode, labeled U1, intersects the line $\omega^2 = 0.0$ at $a = 0.1997$. This value compares very well with the values of a_{crit} estimated from the model computations, for various numbers of samples, which we give in Table 5.

TABLE 5

Number of Samples	a_{crit}
5	0.1979
10	0.1977
15	0.1979
20	0.1982
25	0.1983
30	0.1983

Table 5. Computed values of a for which stability is just achieved, for varying numbers of mass-spaced sampling locations.

The simultaneous stabilization of the gaussian model is made even clearer by following the last unstable eigenfunction to stability, as we show in Figure 16. Note how the radial extent of the unstable region grows as a approaches a_{crit}. Over that whole region we both expect and find a wave with wavenumber approximately equal to a^{-1}, as both the l-wave and s-wave components of the U-mode have wavenumbers close to that value.

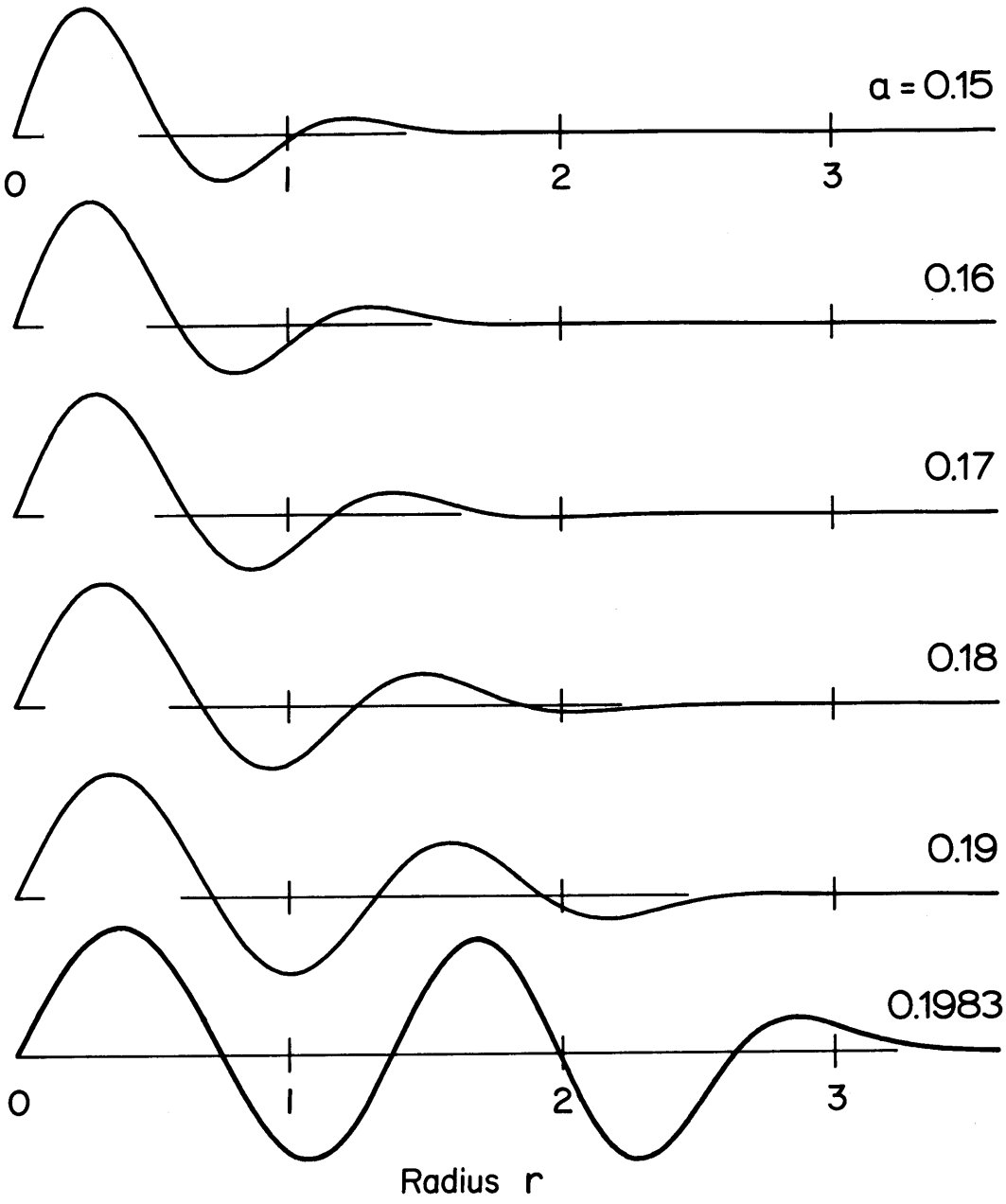

Figure 16. The eigenfunction of the U1 mode, shown for values of a below its stabilization and disappearance, and for the final value $a = a_{crit}$. The wavefunction increases in radial extent due to the flattening of the cutoff curve near a_{crit}, as shown in Figure 11b.

iv) Predictions and Calculations for $a > a_{crit}$

In contrast to the $a = 0.15$ spectra just studied, the spectra of completely stable disks are rather plain. Let us first consider the predictions of SWLT and then check them via the calculations.

By definition, the unstable U-modes have disappeared from the eigenvalue spectrum for values of $a > a_{crit}$. In Figure 11a, the $\omega^2 r$ diagram for $a = 0.15$, these modes occupied the area between the lower curve, the cutoff, and the $\omega^2 = 0$ line. As we see in Figure 11c, the counterpart of Figure 11a for $a = 0.25$, the cutoff curve has risen above the $\omega^2 = 0$ line, effectively eliminating the area available for unstable modes. But less obvious is the fact that the cutoff curve has also risen to pass through all but the highest S-mode eigenvalue. This essentially means that only one stable discrete eigenmode should appear at $a = 0.25$, according to SWLT. In the locale of the previous S-modes, a new type of mode now occurs in place of both the S-modes and the C-modes. This mode is simply described by referring to the kr plot for $a = 0.25$ in Figure 11. At lower values of ω^2, for example 0.20, there is no real solution to the dispersion relation inside some cutoff radius, r_c, which itself is the intersection of the $\omega^2 = 0.20$ line and the cutoff curve. Instead, in that inner disk the wavefunction must take on complex values of k, if indeed a wavenumber can be determined there.

On the other hand, in the region between r_c and r_L we would again expect l-waves and s-waves. At r_L there is a boundary condition on the l-wave component, and at r_c a matching condition on the two; however, since no boundary condition applies to the s-wave at r_L, there will be no condition demanding discreteness for modes at these frequencies. Thus only a continuum remains below ω_x^2 -- the maximum height of the cutoff.

The extinction of the S-modes therefore continues as a increases. By the time a reaches 0.35 the last S-mode disappears, according to SWLT.

The full disk calculations, at $a = 0.25$ and $a = 0.35$, confirm these spectral predictions nicely. In Figure 12 these spectra are each displayed for 2 through 15 mass-spaced samples. For $a = 0.25$, one discrete mode alone appears, at a squared frequency of .3460 (for 15 samples), which is to be compared with the SWLT prediction of .3480. Note that there seems to be a condensation of modes in the $a = 0.35$ spectrum near where the last discrete mode disappeared, but there is nothing that could even be tentatively identified as a discrete mode there.

The only novel eigenfunctions for these values of a is of course the N-modes, so-called because they exist in the "non-allowed" region. In Figure 17 we show progressively improving representations of the eigenfunction at $\omega^2 = 0.20$ for $a = 0.35$, constructed by

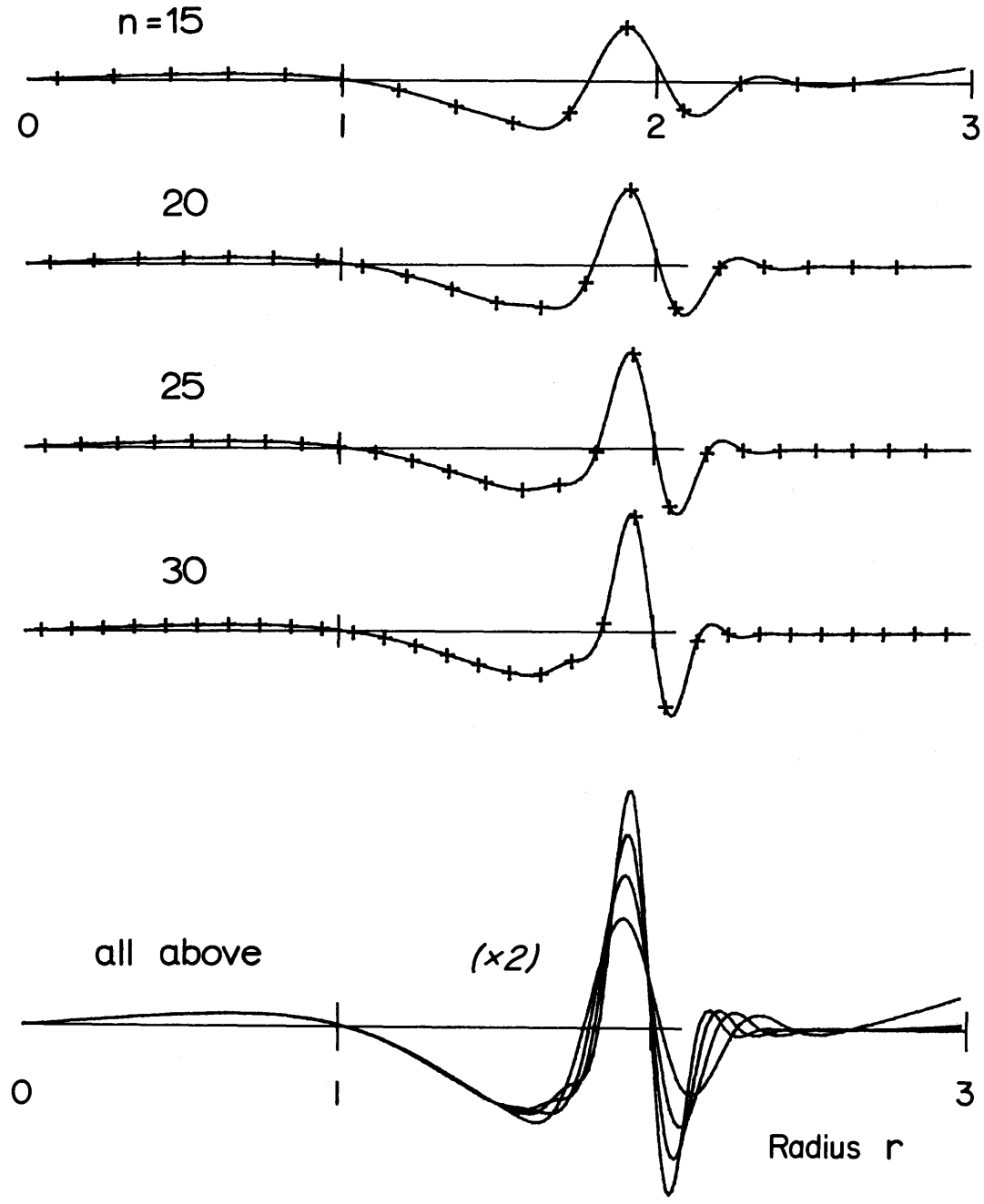

Figure 17. The eigenfunction of the continuum wavefunction at $\omega^2 = 0.20$, evaluated at a = 0.35. Above, four representations are shown, for 15 through 30 equally-spaced sampling rings. The interpolation points are denoted by +. Below, the four are superimposed to emphasize the limited radial extent of this class of mode.

the techniques used for Figure 14. Note that unlike other modal eigenfunctions, it is highly concentrated near the resonance radius. This might be loosely explained by noting that the increase of a weakens the gravitational forces needed to support large-scale modes.

The last item of interest is the manner of disappearance of the S-modes. The initial impression, provided by SWLT, is that the disappearance would be abrupt. However, this idea was immediately dispelled by the initial calculation of the eigenfunction for S1 at $a = 0.25$; the results were <u>not</u> free from a <u>short</u>-wavelength component, and in fact the eigenfunction was extremely sensitive to the location of the sampling points.

In order to decide whether there was intrinsic "corruption" in the S1 mode or rather if it was merely difficult to compute, we chose to perform a totally different series of computer experiments. In these we drove the disk with a forcing that would preferentially induce long-wavelength oscillations.

Two devices were available for this purpose. The most obvious was to simply use a forcing whose radical variation is smooth and of large scale, and similar in shape to the expected long-wave response. In line with this, we chose to use a force proportional to

$$r\, e^{-r^2/2}$$

which is of the rough shape of the earlier (low a) S1 eigenfunctions. Secondly, we chose to force the disk with an exponentially growing force,

$$re^{-r^2/2} \cos \omega t \, e^{\varepsilon t}$$

so that, at any instant, the response will be principally that caused by recent forcing, and less that caused by long-previous forcing. The latter response would naturally be expected to be more corrupted by short-wavelength functions, as the transformation, whether numerical or dynamical, from long-wavelength functions to short-wavelength functions has to have some finite time scale. Therefore we excluded the evolved responses to earlier forcings as much as possible with the device of growing forcing.

Intuitively, to avoid the evolution of long-wavelength responses, it might be thought a better idea to impulsively force the disk. However, this method of forcing necessarily induces responses at frequencies far from that of the S1 mode. These other frequencies necessarily correspond to short-wavelength oscillations, and so the overall response would have large amounts of non-long-wavelength eigenfunctions.

One very convenient form for writing the response of the disk to the above forcing is that obtained by decomposing it into modes. If the force is exactly given by

$$\varepsilon A e^{\varepsilon t} \cos \omega t \, re^{-r^2/2} \tag{69}$$

then the response is

$$x(t) = \sum_{i=1}^{n} E_{ci} \cos \omega t - E_{si} \sin \omega t \qquad (70)$$

where

$$E_{si} = \frac{2\varepsilon^2 A\, q_i\ \omega v_i(r)}{((\omega_i^2 - \omega^2) + \varepsilon^2)^2 + (2\varepsilon\omega)^2}$$

and

$$E_{ci} = \frac{\varepsilon\, A\, q_i\, ((\omega_i^2 - \omega^2) + \varepsilon^2)\, v_i(r)}{((\omega_i^2 - \omega^2) + \varepsilon^2)^2 + (2\varepsilon\omega)^2}$$

Here $v_i(r)$ is the eigenfunction of the ith mode, ω_i its frequency, and q_i the respective component of the forcing.

This response to forcing is not what we are looking for, the S1 eigenmode, but we can theoretically obtain this eigenmode from it. If we were able to take n large enough to form an accurate response function for each ε, we could obtain the S1 eigenfunction by allowing ε to go to zero. The part of the response which does not go to zero with ε, the first term of equation (70), the out-of-phase term, has as a limit

$$\frac{A\, q_{si}\, v_{si}\, (r)}{2\omega} \sin \omega t\ \delta(\omega - \omega_{s1}) \ , \qquad (71)$$

whose radial variation immediately gives the eigenfunction $v_{S1}(r)$. However, a practical maximum value of n prevents this limit in ε from being taken, but it will not prevent our verifying that indeed there is intrinsic corruption of the S1 mode at values of a as low as 0.25.

In Figure 18 we show the out-of-phase component of the response, as well as the in-phase component, for $a = 0.30$ -- which is just below the SWLT disappearance value of $a = 0.313$. For a very high rate of growth, $\varepsilon = \omega/2\pi$, little waviness appears in the eigenfunction. But as ε progresses to smaller values, down to our practical limit of $\omega/6\pi$, more and more radial variation arises. Thus the limit can definitely be expected to possess a short-wavelength component.

The pattern of growth of short-wavelength corruption as a increases is also interesting. Even though SWLT predicts that the discrete mode will disappear at $a = 0.313$ we can continue to follow the descendant of S1 to higher values of a by appropriately defining the frequency and performing the same limit procedure in ε. Any method of choosing frequency is somewhat arbitrary, just as was the choice of forcing function, but it is physically appealing to maximize the kinetic energy absorbed by the disk by choosing ω^2. We used this rule exclusively, and found that it agreed with SWLT, and worked without any difficulties in convergence up to at least $a = 0.40$.

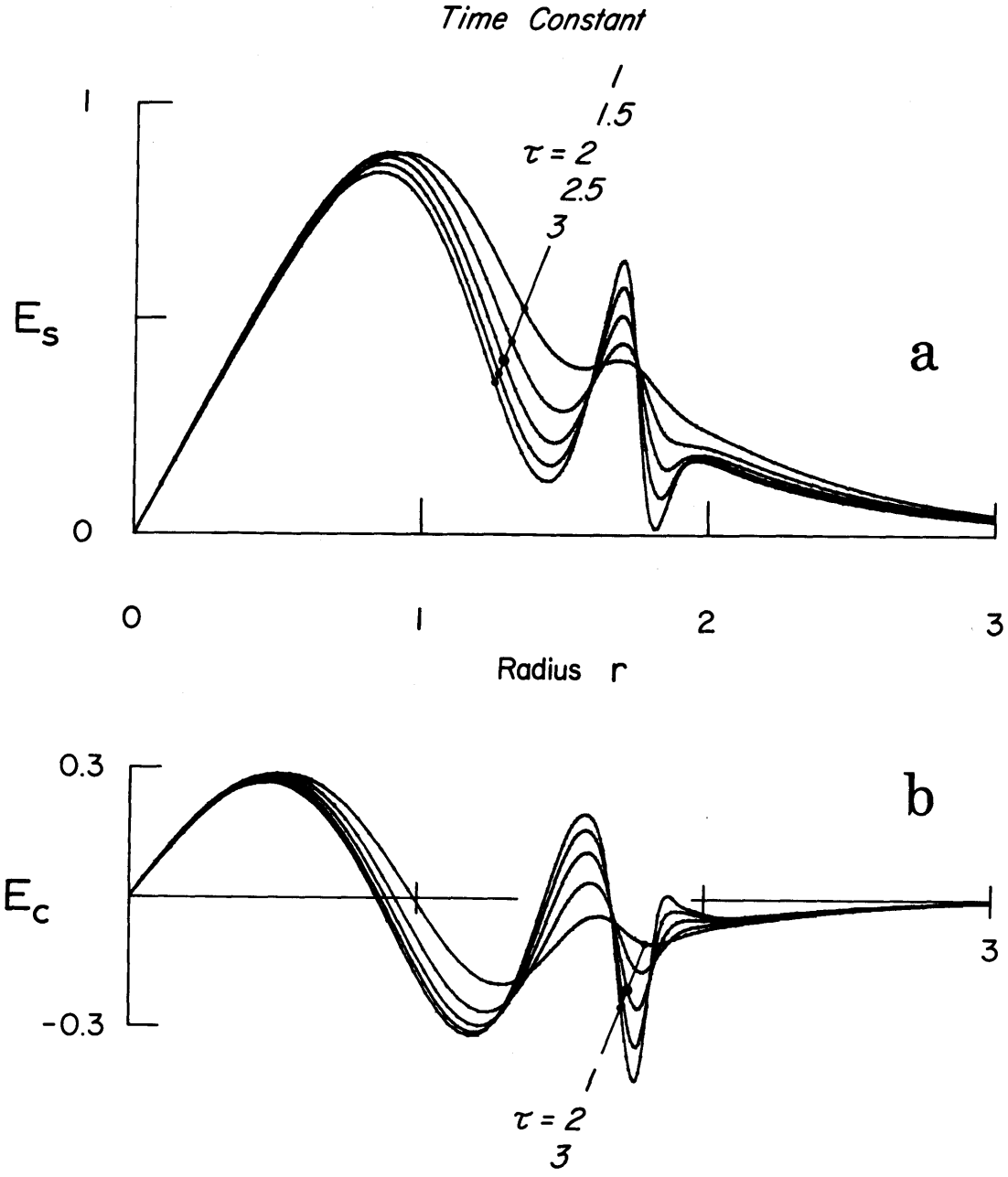

Figure 18. The out-of-phase, E_s, and in-phase, E_c, components of the response to a growing forcing, as functions of radius r. The curves shown are for growth rates equal to $\omega/2\pi\tau$, where ω is the frequency of forcing. Here ω is the frequency of the S1 mode at offset a = 0.3.

The gradual increase of the short-wavelength component in the "discrete mode" is pictured in Figure 18c, where we show the response function for $a = 0.20$ through $a = 0.40$ for an ε of $\omega/4\pi$. There is apparently no sudden increase in waviness upon passing through the SWLT-predicted disappearance value of $a = 0.313$

This gradually increasing "leakage" from long waves to short waves at a near a cutoff radius is quite understandable in view of the model equation of Appendix IV: near the cutoff radius the two saddle points in the asymptotic solution to the integral equation move close enough together so that they cannot be treated independently. At this point the discrete frequency mode has both long and short wave components. On the other hand, for $a < 0.313$, the 1-wave component must still alone satisfy the boundary conditions which caused it be discrete, and thus the spectrum but not the eigenfunctions give witness to the abrupt disappearance of the last discrete mode.

v) Summary

The variation of the modal spectrum with a can be very concisely summarized in one diagram. In Figure 19 we show the theoretical spectrum for values of a from 0.0 to 0.35. The lowest values of a, 0.0, is clearly seen as a very special case: the unstable eigenmodal traces tend to it only asymptotically, as they run out to negative infinity. The positive continuum of stable modes approaches arbitrarily close to $a - 0.0$ but does not occur there. Only the discrete stable

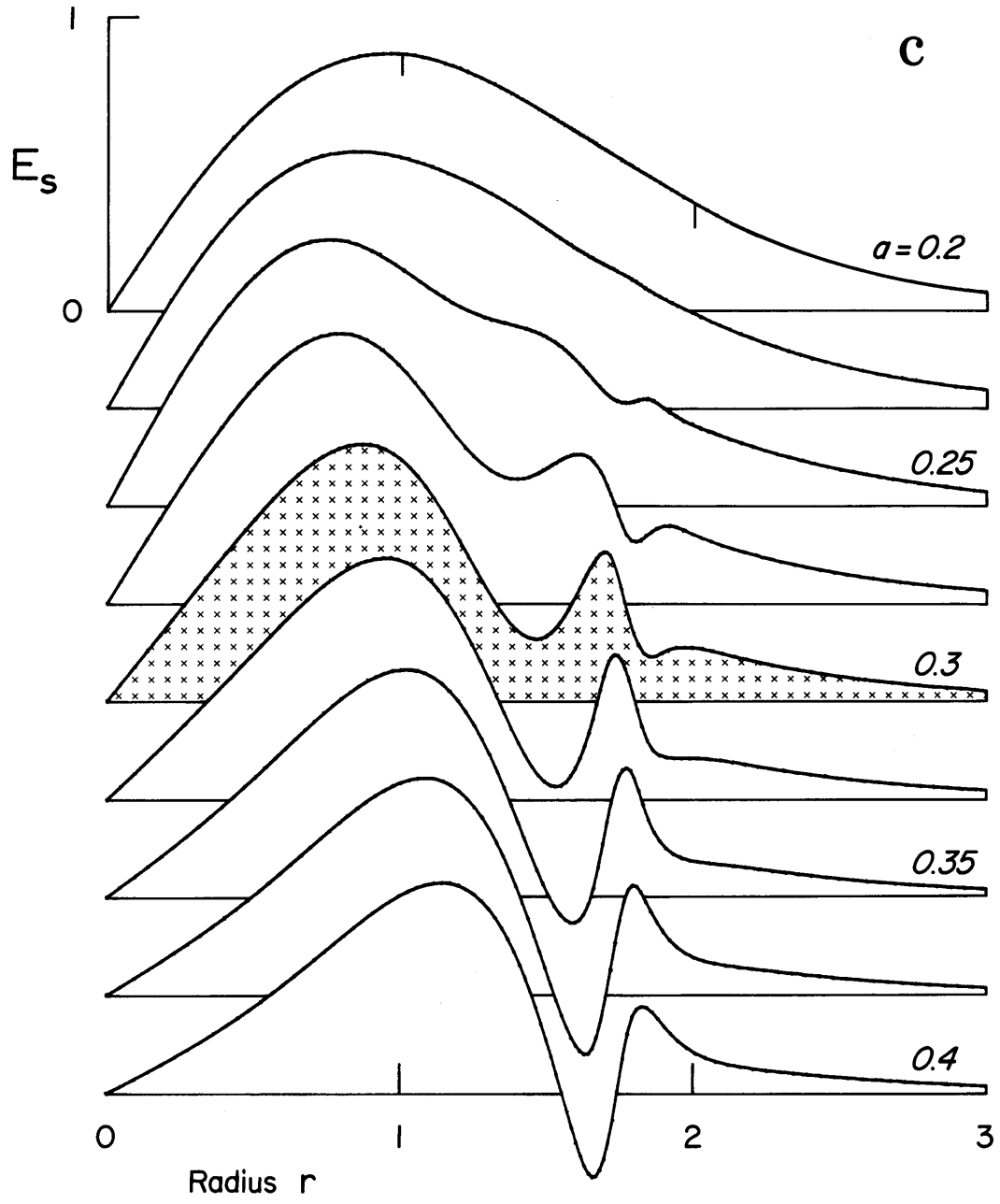

Figure 18c. The variation of the out-of-phase response to a growing forcing for various values of offset a, shown as functions of radius r. The composite wavefunction E_s shows increasing amounts of the short-wavelength component as a approaches and passes a_{crit}.

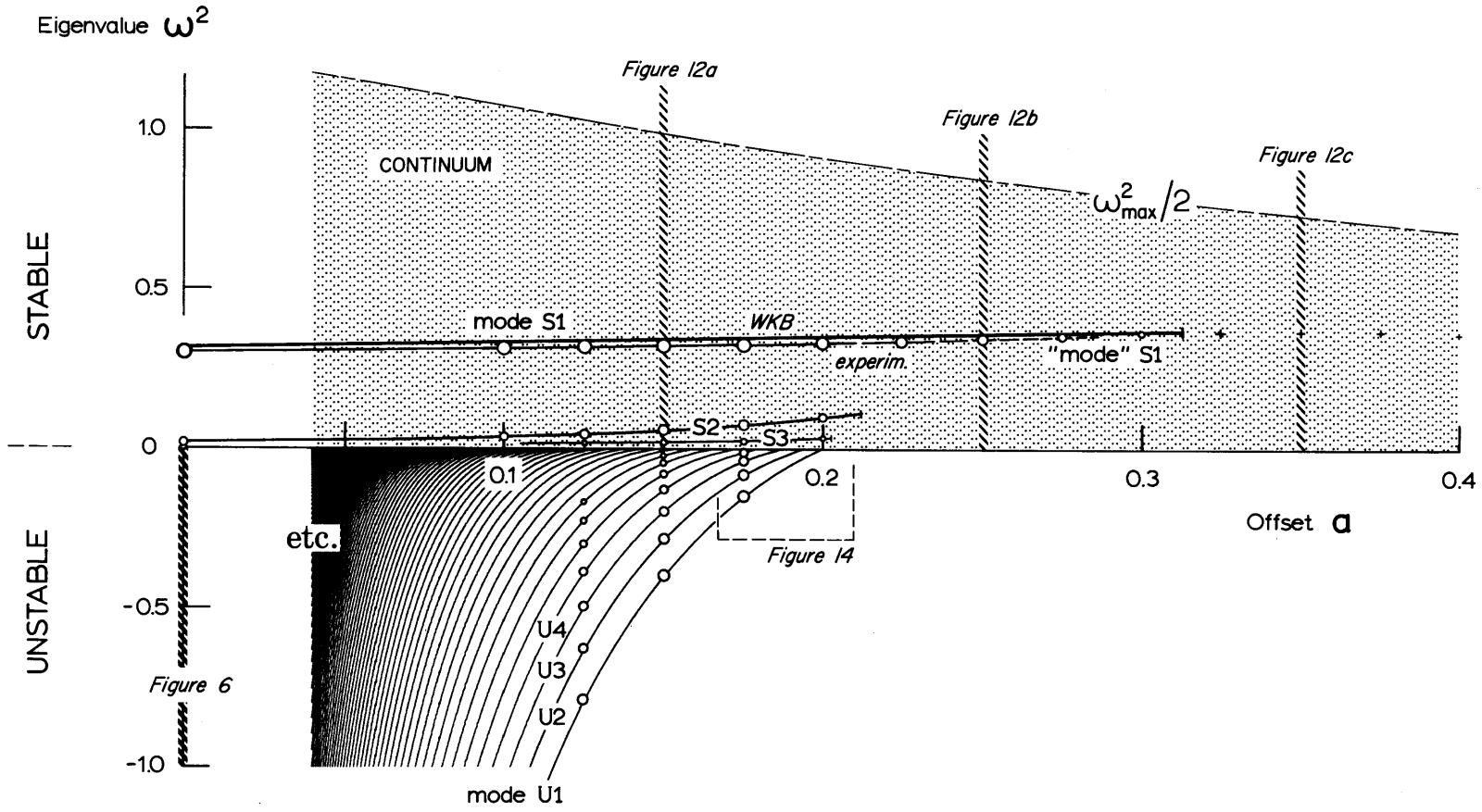

Figure 19. Axisymmetric Summary Diagram. Here SWLT-predicted spectra and experimental values are compared. On the bottom of the diagram, the thickly populated family of discrete unstable modes thins to only a few members as a approaches a_{crit}; on the top the small number of S-modes are seen buried in the stable continuum. The disappearance of the S-modes are indicated by the end of the solid lines, but the extension of S1 to higher values of a by our forcing arguments are indicated by + near the right edge of the figure.

frequencies extend from 0.0 into the remainder of the diagram. These latter modes disappear when they contact either the cutoff curve, or after they enter the thin band of frequencies along $\omega^2 \gtrsim 0$ not treatable by SWLT. Finally note that the continuum upper limit shrinks, being given exactly by $\kappa^2(0,a)$ which itself is a measure of the weakening of gravitational force brought on by the increase in a.

This entire set of predictions has been shown to be accurately borne out by the full disk calculations, as shown by the calculated eigenvalues superimposed in Figure 19. In brief, the axisymmetric modes are both quite complicated, but also quite understandable by asymptotic analyses.

IV. RESULTS OF NONAXISYMMETRIC CALCULATIONS

Nonaxisymmetric motions of our model galaxy are, expectedly, much more complicated than the axisymmetric motions we have just examined and, regretably, even with this simple model, less amenable to explanation. In this chapter we will principally describe several types of modal behavior that the disk can display, and secondarily, for a few modes, provide an interpretation in terms of a straightforward application of SWLT.

Before proceeding to the reportage, it is useful to recall that for the same number of sampling points, say n, there are twice as many nonaxisymmetric eigenfrequencies and eigenmodes, $4n$, for each non-zero azimuthal mode number m, as there are axisymmetric ones. Furthermore, while the axisymmetric eigenfrequencies always occur in pairs, the nonaxisymmetric may or may not. Real eigenfrequencies occur without any partner at all, while complex eigenfrequencies are accompanied by their conjugate. This last fact of course occurs because the modal eigenfrequencies are the characteristic values of a matrix, denoted by M in equation (55), which we were able to write in real form by a particular choice of phase for the displacements.

To accommodate these complex frequencies, we will be forced to display the spectrum in a plane rather than stretched along a line, and

consequently the illustrative pyramids of eigenvalues used in Chapter III are not as appropriate here. They might be used for the stable, i.e., real, frequencies, but even this usage has difficulties. The eigenfrequency calculations with smaller numbers of sampling rings, say 10 or less, simply are not accurate enough to reveal the discrete frequencies which stood out so clearly in the pyramids. For this reason, the only method we used for the detection of nonaxisymmetric discrete stable frequencies was the comparison of spectra for different sampling locations, each done with the maximum practical number of rings, 20. This latter limit arises because the computer used for the majority of the nonaxisymmetric calculations was able to process at most 80 by 80 matrices.

Even with 20 rings, accuracy problems still remain. For example, two nearly equal stable eigenfrequencies can, for some choices of sampling point location, turn into a complex conjugate pair of frequencies with a very small imaginary part. In spectra where this happened, an empty gap appears in the stable portion of the spectrum. We did not attempt to fill in such gaps as we considered that it would not be wholly reliable to separate out these mock unstable modes from the large number of actually unstable modes. Fortunately, this difficulty occurs only to a small number of modes in any single spectrum.

As noted in Appendix II, our nonaxisymmetric calculations were done only for non-zero offset a. In the reporting to follow we ex-

clusively concentrate on the values of a near the axisymmetric a_{crit}, namely from 0.15 to 0.25.

(a) General Characteristics of the Eigenvalue Spectra

Contrary to what might be initially imagined, the computed spectra for different values of azimuthal mode number, m, are remarkably similar. For a given value of a, the spectra for $m = 1$, 2, or 3 exhibit the same features, such as the existence of discrete modes, and also change in the same fashion when either a or the sampling locations are varied. This occurs despite the differences in both the location of resonances and cutoffs and also the frequencies for which they exist. Thus, comments made concerning a spectrum calculated for one value of m are usually valid for other values of (non-zero) m also.

In Figure 20 a typical spectrum is shown to illustrate the characteristic features. This particular spectrum corresponds to $m = 1$ and $a = 0.15$, that is, for a less than our axisymmetric a_{crit}. The sampling points were equally spaced out to $r = 3.2$, in the manner used for equispaced points in Chapter III. In the nonaxisymmetric calculations, we exclusively used this method of sampling, as we found the equal spacing of samples to be more convenient than the mass spacing, inasmuch as the former places more of the samples at larger radii where the disturbances are often most pronounced.

The outstanding feature of this pectrum is clearly the pair of sloping, almost straight lines projecting away from the $Re(\omega)$ axis.

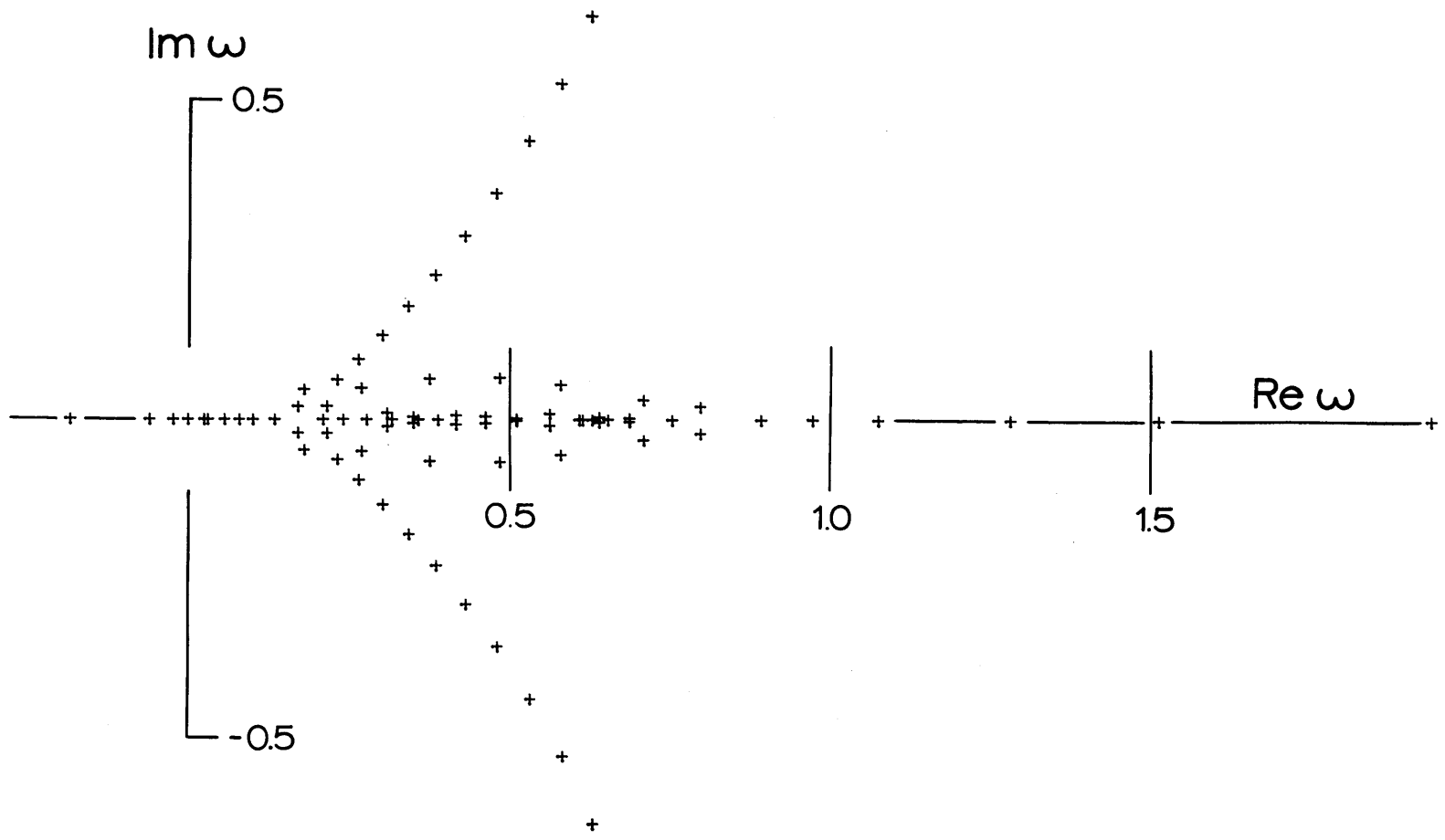

Figure 20. Nonaxisymmetric eigenvalue spectrum for periodicity m = 1, offset a = 0.15, showing the multitude of complex, i.e., unstable frequencies. The sampling points were spaced uniformly out to radius r = 3.2. Here N = 20.

These frequencies represent highly unstable modes, with growth or decay rates $Im(\omega)$ of the same magnitude as their oscillation rates $Re(\omega)$. The remaining unstable frequencies do not consistently form a simply shaped locus in the complex plane. However, in many cases, especially when the sampling locations are nearer the center, these points include one or more conjugate pairs of arch-shaped curves such as the one appearing in Figure 20.

(b) The Most Unstable Modes

Let us first concentrate on the "line" of unstable eigenvalues. Those points are more aptly described as beads on a necklace, as they are quite discrete, and the gaps between individual values remain empty for any location of sampling rings. The accuracy of these frequencies is demonstrated in Table 6, where we give the most unstable members of this category of mode for $m = 1$ and 2, and $a = 0.15$ and 0.25. In all four cases the first, or most unstable, eigenmode has a frequency accurately computed to about three digits, with the accuracy declining for less unstable modes.

Note that only two discrete frequencies are listed for $a = 0.25$ for both $m = 1$ and $m = 2$. These are the only well-determined frequencies, as the remainder have such low values of $Im(\omega)$ that they cannot be extracted from the non-"line" complex frequencies.

TABLE 6

r_e ring extent

	3.2	3.6	4.0	4.4
m = 1				
a = 0.15	.6344$\pm i$.6836	.6339$\pm i$.6332	.6331$\pm i$.6328	.6318$\pm i$.6322
	.5870$\mp i$.5275	.5857$\mp i$.5240	.5834$\mp i$.5196	.5799$\mp i$.5146
	.5358$\pm i$.4381	.5327$\pm i$.4352	.5284$\pm i$.4323	.5221$\mp i$.4295
	.4846$\mp i$.3559	.4801$\mp i$.3515	.4736$\mp i$.3467	.4658$\mp i$.3434
	.4350$\pm i$.2891	.4299$\pm i$.2855	.4248$\pm i$.2843	.4175$\pm i$.2829
a = 0.25	.2802$\pm i$.0695	.2801$\pm i$.0694	.2799$\pm i$.0694	.2797$\pm i$.0693
	.2058$\mp i$.0244	.2090$\mp i$.0261	.2092$\mp i$.0255	.2082$\mp i$.0264
m = 2				
a = 0.15	1.2642$\pm i$.6515	1.2645$\pm i$.6527	1.2651$\pm i$.6542	1.2660$\pm i$.6562
	1.1820$\mp i$.5458	1.1823$\mp i$.5460	1.1826$\mp i$.5469	1.1820$\mp i$.5485
	1.0970$\mp i$.4541	1.0961$\mp i$.4538	1.0970$\mp i$.4541	1.0858$\mp i$.4548
	1.0153$\mp i$.3777	1.0122$\mp i$.3793	1.0207$\mp i$.3859	
	.9326$\mp i$.3164	.9265$\mp i$.3270	.9410$\mp i$.3496	
a = 0.25	.8308$\pm i$.2177	.8308$\pm i$.2177	.8310$\pm i$.2176	.8310$\pm i$.2174
	.7136$\mp i$.1555	.7139$\mp i$.1560	.7137$\mp i$.1576	.7160$\mp i$.1597

Table 6. Discrete unstable frequencies found in comparing data for 20 equally-spaced sampling points, with different outer edges.

The unstable modes of the "line" have a characteristic planform, two examples of which are shown in Figures 21 and 22. The density maps are contoured at equal intervals, and only the non-negative contours

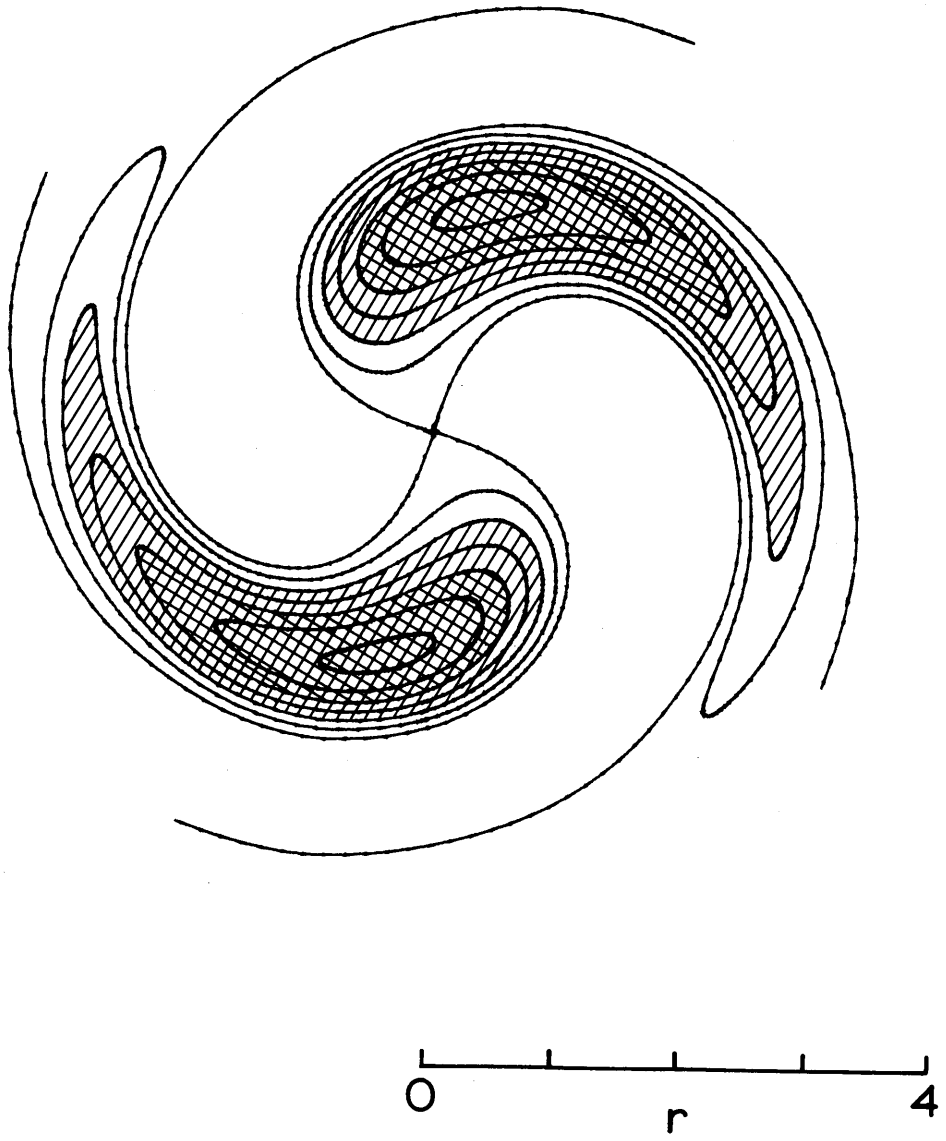

Figure 21. Density contours for the most unstable mode with periodicity $m = 2$ and offset $a = 0.15$. Only the contours with non-negative density are shown. The marked contours are equally spaced in density. The disk rotates counterclockwise; this spiral grows with time such that $\omega = 1.264 + i\ 0.653$.

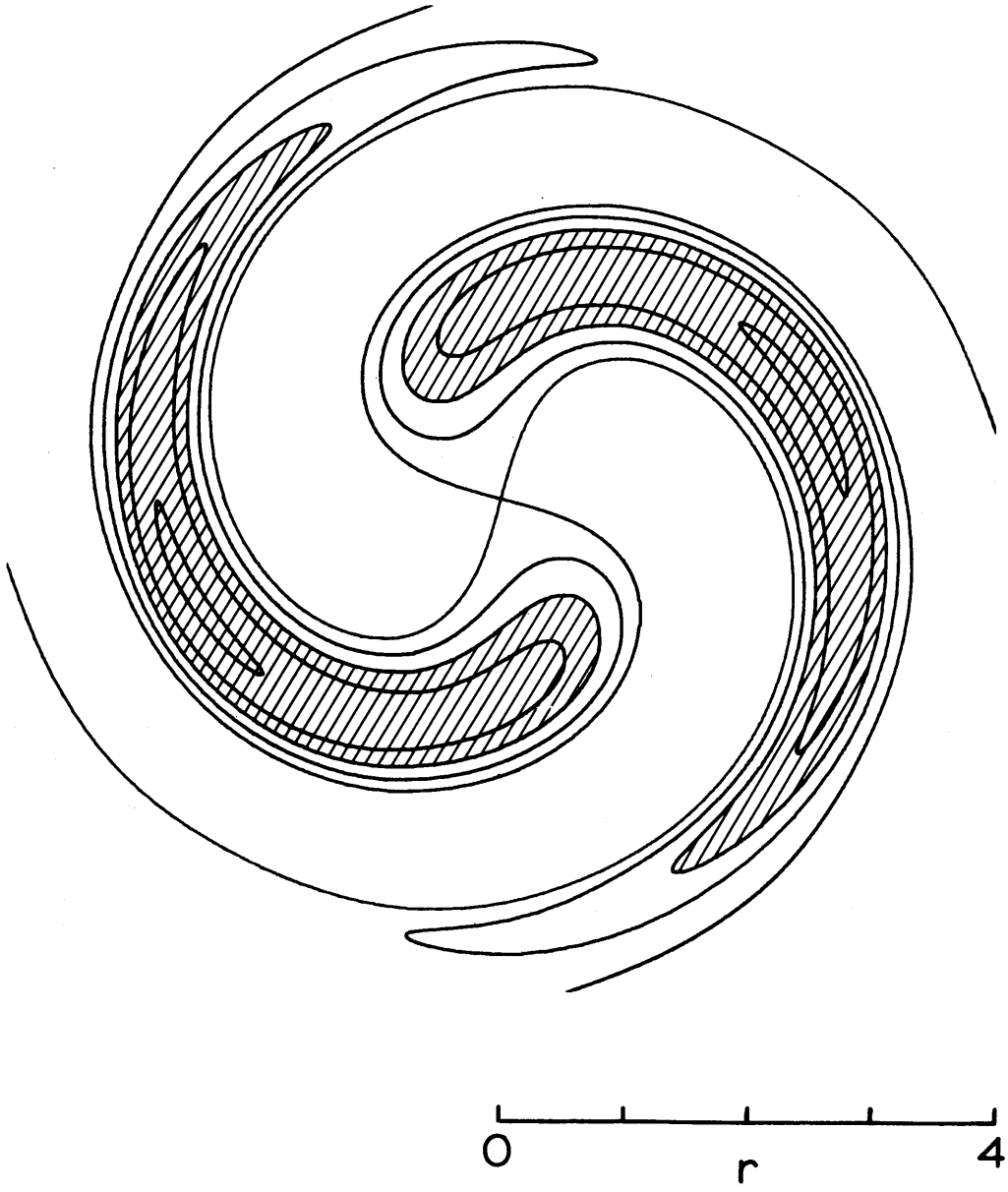

Figure 22. Density contours for the second most unstable mode with periodicity $m = 2$ and offset $a = 0.15$. Here $\omega = 1.182 + i\ 0.546$.

are shown. The modes are grand open-armed spirals of a trailing sense, with the winding becoming tighter as the imaginary part of the frequency declines. The two spirals shown are the two most unstable for $m = 2$ and $a = 0.15$.

Although these two spirals were computed for a less than the axisymmetric a_{crit}, spirals do persist well beyond this value, with a large, easily recognizable spiral existing for $m = 2$ and $a = 0.40$. On the other hand, there is a large reduction in the growth rates when a is increased. In Figure 23 we track the imaginary parts of the most unstable modes from $a = 0.15$ up to 0.25. The growth rate does decline by a factor of 3 but there is clearly no sign that nonaxisymmetric stabilization will occur with this model as easily as axisymmetric stabilization did. Other calculations show that even with the extreme value of $a = 1.0$, unstable modes continue to exist for $m = 2$.

(c) Stable Modes

To discuss the stable modes we will concentrate solely on describing the spectrum of one typical case: $a = 0.25$ and $m = 1$. As noted before, spectra with a greater than a_{crit} have less extensive lines of unstable points, however, this particular example has no special features absent in other spectra.

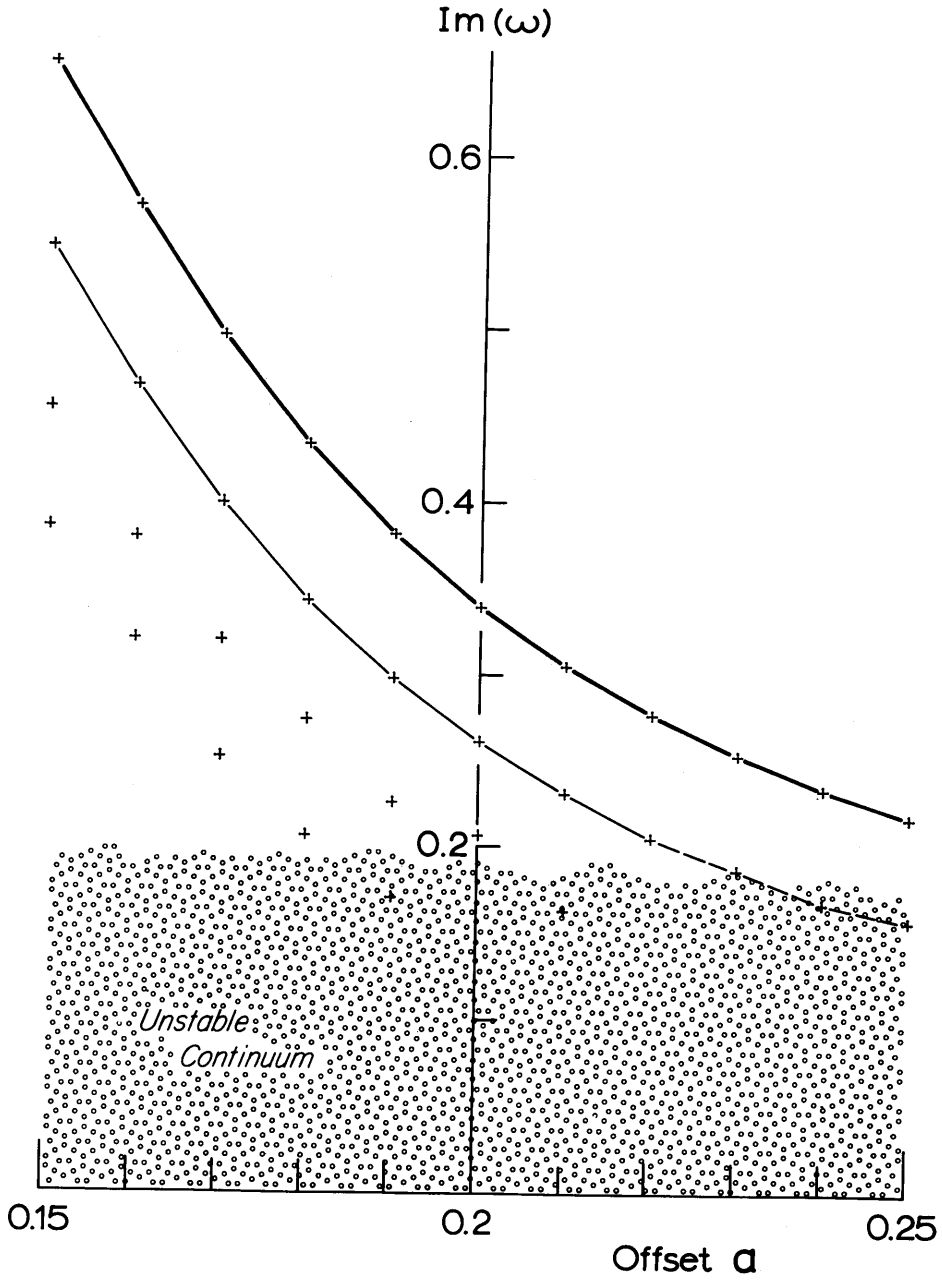

Figure 23. Variation of the imaginary parts of the most highly unstable modes as a function of offset a. The leftmost point of the darker line corresponds to the spiral pictured in Figure 21, the lighter line, Figure 22. Neither the grand spirals nor the other unstable modes stabilize near the axisymmetric stability point $a = a_{crit}$.

In Figure 24 we show the stable part of the spectrum, as calculated for four different sample spacings. Since the frequencies are very crowded near $\omega = 0.3$, a section of the spectrum has been shown expanded in scale on the right side of the figure. A number of discrete frequencies are apparent. Two stand out at $\omega = 0.422$ and 0.337, and a sequence of 7 stretch from $\omega = 0.102$ down to 0.017; some other tentative identifications might be added.

To even begin to interpret these frequencies we need to generalize SWLT to the nonaxisymmetric case. Rather than rederive the nonaxisymmetric dispersion relation for modified gravity, let us just note here instead, that if we make the assumption of tightly wound waves following Lin and Shu (1964), the derivation for our model goes through exactly as the axisymmetric one does with a single exception. In all the formulas relating to SWLT, the frequency, ω, is replaced by the frequency relative to the local convection, $\omega - m\Omega$. In other words, the dispersion relation merely changes to

$$\kappa^2 - (\omega - m\Omega)^2 = 2\pi G\mu k e^{-ak} , \qquad (72)$$

where again k is the radial wavenumber.

Before, discrete eigenvalues arose in connection with special radii, specifically, resonance and cutoff. These radii exist in abundance in the nonaxisymmetric cases. In contrast to the axisymmetric

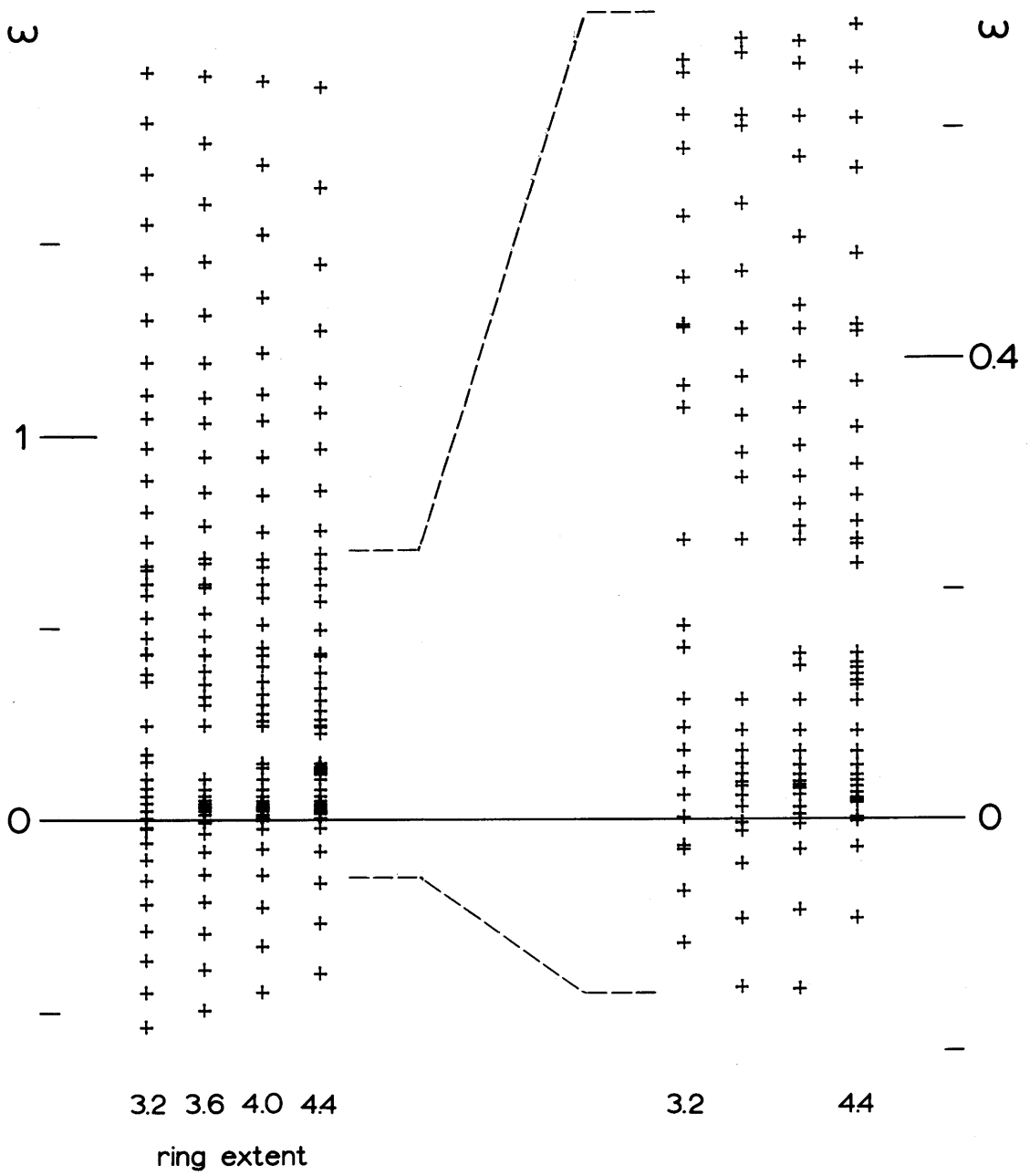

Figure 24. The real eigenvalues for offset a = 0.25, periodicity m = 1. On both left and right, the four columns correspond to four experiments with different ring spacings. In each, the sampling rings were equally spaced out to the ring extent, in the manner discussed in Chapter III. The right side is an expansion of a crowded portion of the multiple spectra shown on the left.

case, there are three distinct types of resonance when m is non-zero, rather than only one. Two are typically referred to as Lindblad resonances, and are radii where k takes on the two values 0 and ∞. Equivalently, they are defined as radii where

$$\omega = m\Omega \pm \kappa . \tag{73}$$

Here the positive and negative signs correspond to the "outer" and "inner" Lindblad resonances, L^+ and L^-, respectively. These resonances are clearly the analogs of the $m = 0$ resonance, when $\omega = \pm \kappa$.

The remaining resonance arises from an entirely different cause, involving the orbital frequency rather than the epicyclic frequency. Although the dispersion relation (72) itself does not reveal any special behavior when $\omega = m\Omega$, consideration of the particle displacements, X and Y of equation (24), does. In the present approximation they are related by the equation:

$$Y = X/(\omega - m\Omega) \tag{74}$$

which implies that tangential displacement Y may suffer a singularity at the "corotation radius" or "particle resonance radius", r_p, where $\omega = m\Omega\ (r_p)$.

The loci of these three resonances are plotted in Figure 25 for $a = 0.15$ and 0.25 and $m = 1$ and 2. The only qualitative difference between the sets of curves involves $\bar{L}$, which is negative only for $m = 1$. This negative value implies that a retrograde pattern speed exists at inner radii, as noted in Chapter II, and does not imply any difference in the character of the spectrum for $m = 1$, such as the $m = 0$ instability which occurs when the axisymmetric ω^2 can be negative.

In Chapter III we used the term "cutoff" to denote a radius where the wavenumber took the value, $1/a$, which corresponded to the lowest allowable frequency. Here such radii are also possible, but at any such radius there are two frequencies related to it.

By setting $k = 1/a$ in the dispersion relation (72) and solving for ω,

$$\omega = m\Omega \pm (\kappa^2 - 2\pi G\mu/ae)^{\frac{1}{2}} \qquad (75)$$

it becomes obvious that there will be either two or zero frequencies for which a certain radius r is the cutoff radius. If

$$\kappa^2(r) \quad 2\pi G \mu(r)/ae \quad , \qquad (76)$$

then two real frequencies are described by equation (75). Otherwise, cutoff might be said to occur only for a complex pair of frequencies.

As noted in equation (67) of Chapter III, the condition (76) for cutoff to exist is exactly the condition for the squared axisymmetric cutoff frequency, ω^2_{min}, at the radius r to be non-negative. Thus from the axisymmetric diagrams of Figure 11, we can deduce that for $a = 0.15$, cutoff would only exist for radii beyond $r = 2.81$, while for $a = 0.25$, cutoff occurs at all radii. We plot the frequencies of the cutoffs as C^+ and C^- on Figure 25a and c. As just noted, in both diagrams for $a = 0.15$, these lines intersect the particle resonance curve P at $r = 2.81$, where $\kappa^2 = 2\pi G\mu/ae$.

Recall that in the axisymmetric case, two types of discrete frequencies arose, one where the long and short waves met at cutoff, and the other with only long waves stretching from center to resonance. Let us develop the analogous waves for the nonaxisymmetric case $m = 1$, $a = 0.25$, which we are using as our typical representative.

As shown in Figure 25b, there is only one band of frequencies which might contain discrete modes analogous to the U-modes of Chapter III. Since the region between C^- and C^+ corresponds to frequencies where no real value of k will satisfy the dispersion relation, frequencies from $\omega = 0.16$ to $\omega = 1.14$ cannot support such a mode. Also, neither do the frequencies above the maximum L^+, 1.96, or below the minimum L^-, -0.66, have real solutions of relation (72). Lastly, the bands from 1.14 to 1.96 and -0.66 to 0.0 have no cutoff radius, leaving only the narrow region of 0.0 to 0.16.

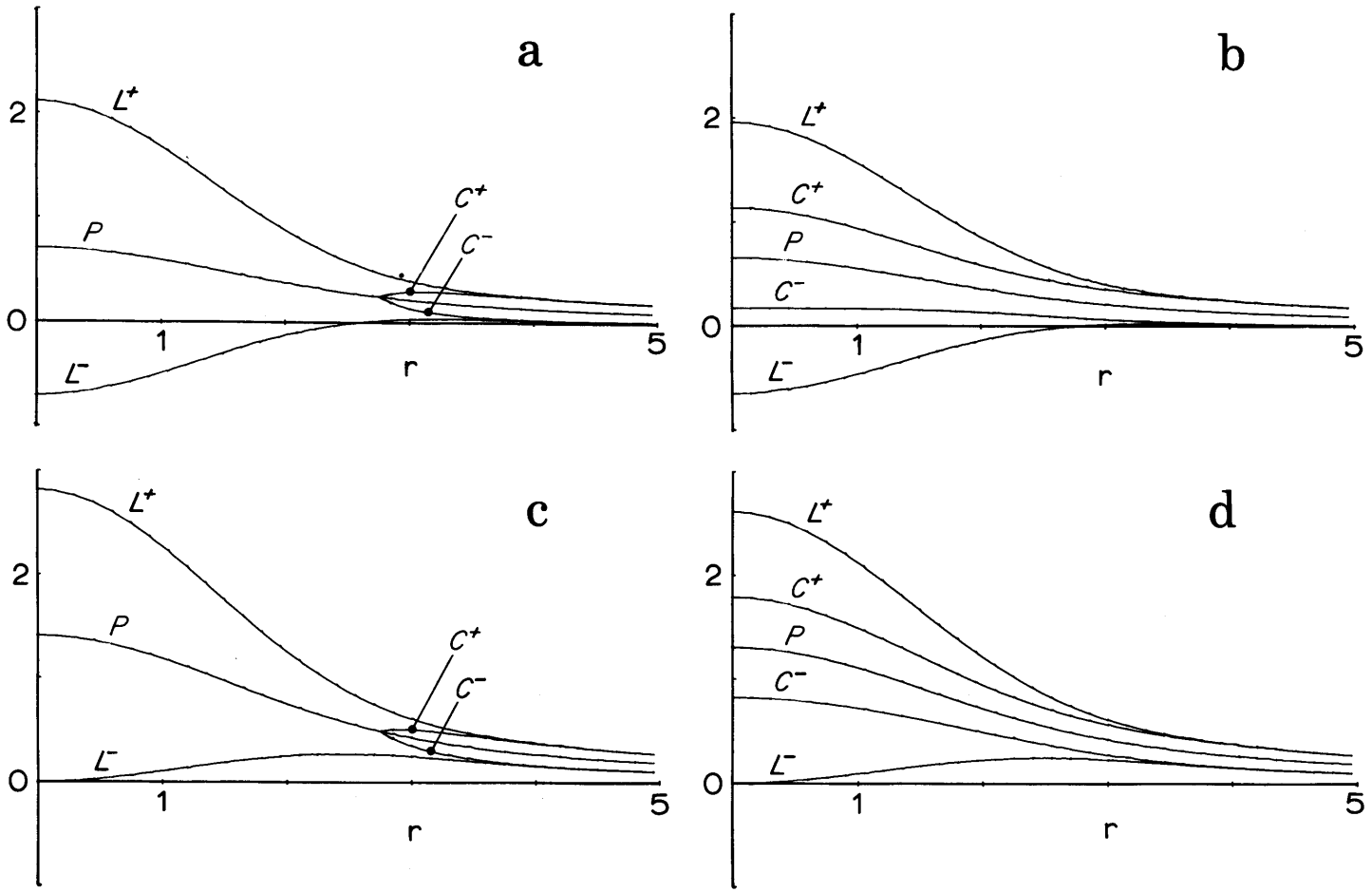

Figure 25. Frequency vs resonance and cutoff radii. Both inner Lindblad and outer Lindblad resonances, L^- and L^+, the particle resonance, P, and inner and outer cutoff radii, C^- and C^+, are shown for (a) m = 1, a = 0.15, (b) m = 1, a = 0.25, (c) m = 2, a = 0.15, and (d) m = 2, a = 0.25. For a less than the local value of a_{crit} (see Chapter III), the cutoff curves will not extend over all values of radius r.

Again we have the problem that in the lowest part of this band, from 0.0 to 0.03, two inner Lindblad resonance radii occur between the center and cutoff, analogously to the problem of the T-modes elaborated on in Chapter III. Since no real k exists in the interval of range between those two resonances, we cannot use SWLT to describe modes occurring in this band of frequencies. This leaves only the band from 0.03 to 0.16 to contain candidates for a SWLT explanation similar to that of the U-modes.

Using the rule of phase differences developed for axisymmetric U-modes, namely that the phase integrals for long and short waves differ by $(n - \frac{1}{2})\pi$, we can predict the frequencies at which U-modes should occur. For our representative case, six such frequencies occur, as indicated in Table 7. For comparison, the seven detected discrete frequencies are also listed, along with a debatably discrete frequency which appears near the highest predicted value. The lowest two detected discrete frequencies fall below the band in which SWLT is applicable -- these modes, if they are the analogous to the T-modes, "tunnel" from resonance to resonance with complex values of k.

Unfortunately, even such limited success does not occur with the analogues of the S-modes. SWLT applied to the example case indicates that a $1\frac{1}{2}$-π mode should exist at $\omega = 1.167$ with a wave stretching between the center and L^+, and another $1\frac{1}{2}$-π

TABLE 7

n	SWLT Predicted Frequency	Computed Frequency
1	0.139	0.142 ?
2	0.108	0.102
3	0.083	0.076
4	0.065	0.058
5	0.052	0.045
6	0.042	0.038
	-	0.028
	-	0.017

Table 7. Comparison of nonaxisymmetric predicted and calculated U-mode frequencies for $a = 0.25$ and $m = 1$.

mode should exist at $\omega = -0.064$, lying between the center and L^-. These frequencies do not correspond to any discrete frequency in Figure 24. Conceivably, the computational accuracy might be insufficient to indicate the discrete modes; however, the S1 modes were by far the best computed of all axisymmetric modes, as can be seen by the two-ring calculations in Figures 7 and 12. No other values of a or m fared any better.

Conversely, the other two discrete frequencies, 0.422 and 0.337, do not correspond to real wavenumbers for almost all radii. As seen in Figure 25, the only real solution of the dispersion relations at those frequencies exists between C^+ and L^+, but over this interval the SWLT-predicted integrals of long waves amount

to only fractions of π, thereby disallowing any SWLT explanation similar to the S-modes of Chapter III. Furthermore, as was discussed in connection with the N-modes of Chapter III, cutoff and resonance boundaries do not combine to form discrete modes, at least in the axisymmetric cases. Thus, a wholly unique explanation seems to be necessary for these two modes.

(d) Other Unstable Modes

Little can be said with confidence about the remaining unstable modes, as they have the least well-computed frequencies. There seem to be no unstable discrete frequencies other than those previously mentioned. However, the modes whose frequencies fall on an "arch", such as in Figure 20, appear to comprise a one-dimensional continuous spectrum. Because of the curious contours of the density plots we include, as Figure 26, the planform of a typical "arch" mode, computed for $a = 0.15$ and $m = 2$, the same values as Figures 21 and 22. Two sets of peaks and valleys occur in this mode rather than one. The large innermost peak is the higher one and is located just inside the corotation radius. The smaller peak has a maximum value of only 7% of the first peak, and lies just outside corotation. Other modes of the arch appear similar, but with differing peak density ratios and longer spiral-like tails. But as with the grand spirals, no SWLT explanation was found.

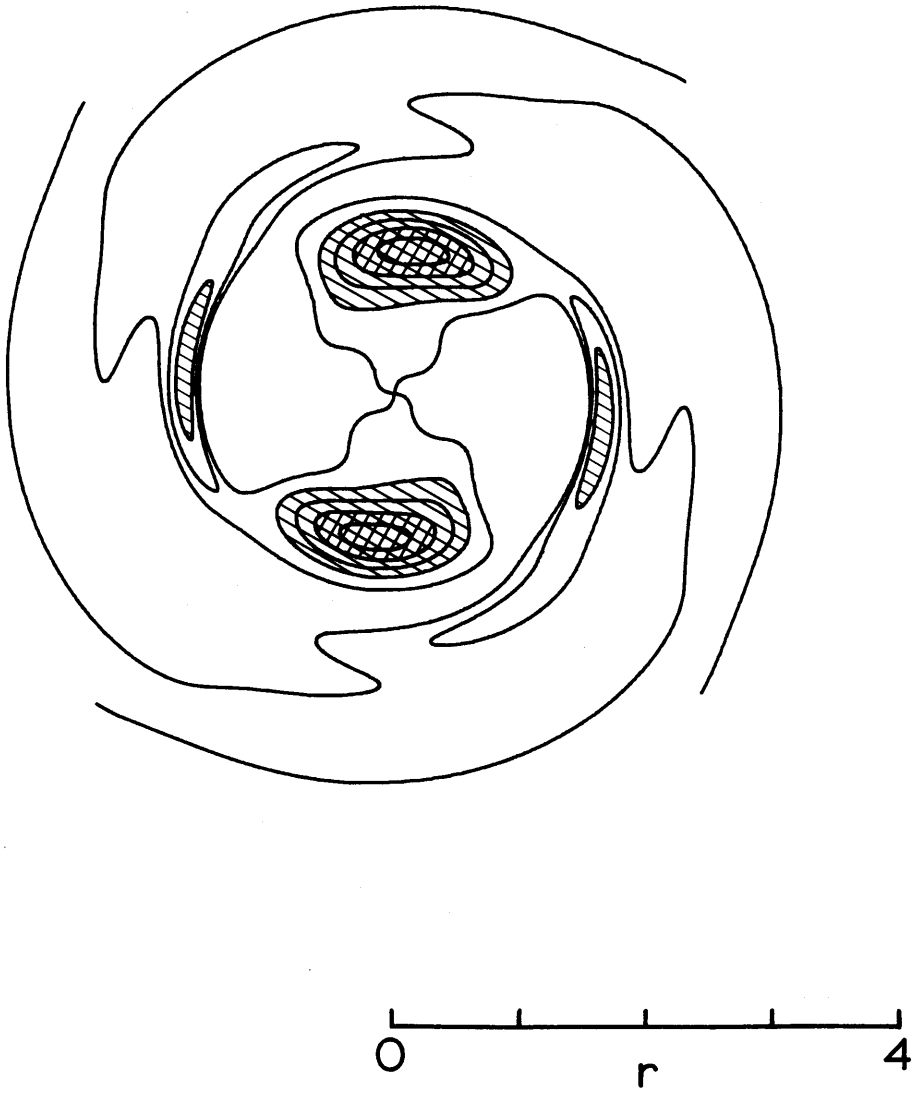

Figure 26. Density contours for a mode lying on the "arch" of unstable values. For this plot $a = 0.25$, $\omega = 1.214 + i\ .058$. The contours on each of the two hills are separately equispaced in density; however the maximum height of the second, outer peak is only seven percent of the maximum of the inner peak.

APPENDIX I. DISPLACEMENT INTERPOLATION

All the calculations done with the model were carried out by computing radial and tangential displacements of the equilibrium disk. The purpose of this appendix is to show why a direct interpolation of the displacements is inferior to a symmetrized interpolation of weighted displacements.

To do this we must consider the requirements of continuity. Let (x, y) be a rectangular coordinate system with the origin at the disk center. Any continuous displacement of the disk can be described by the function of x and y which describes the motion of the point (x, y) along the x and y axes. Denote these displacements by (X, Y). Assuming the displacements arbitrarily differentiable, these functions can be expressed as power series in x, y:

$$X = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} a_{ij} x^i y^j \qquad (A\text{-}1)$$

$$Y = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} b_{ij} x^i y^j$$

These expressions can be transformed into polar reference coordinates (r, θ) by making the substitution:

$$x = r \cos \theta$$
$$y = r \sin \theta \, , \tag{A-2}$$

and the displacements can be transformed into radial and tangential displacements, (R, T) by

$$R = X \cos \theta + Y \sin \theta$$
$$T = -X \sin \theta + Y \cos \theta \tag{A-3}$$

which lead to

$$R = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} r^{i+j} \left(a_{ij} \cos^{i+1} \theta \sin^{j} \theta + b_{ij} \cos^{i} \theta \sin^{j+i} \theta\right)$$

$$T = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} r^{i+j} \left(-a_{ij} \cos^{i} \theta \sin^{j+1} \theta + b_{ij} \cos^{i+1} \theta \sin^{j} \theta\right). \tag{A-4}$$

The trigonometric product $\cos^m \theta \sin^n \theta$ can be written in terms of trigonometric functions of multiple angles:

$$\cos^m \theta \sin^n \theta = \sum_{k=0}^{m+n} \left(A_k^{mn} \sin k \theta + B_k^{mn} \cos k \theta\right). \tag{A-5}$$

If n is odd, the left hand side is odd around $\theta = 0$, and therefore $B_k^{mn} = 0$. If n is even, $A_k^{mn} = 0$. Another identity is that $A_k^{mn} = B_k^{mn} = 0$ both for k odd if $m+n$ is even, and for

k even if $m+n$ is odd. This identity is easily proved by recursion. If it holds for $m = m_o$, $n = n_o$ it also holds for $m = m_o + 1$, $n = n_o$, and for $m = m_o$, $n = n_o + 1$ as can easily be seen by performing the multiplication and using the identities

$$\begin{aligned}
\cos\theta \cos k\theta &= \tfrac{1}{2}\cos(k+1)\theta + \tfrac{1}{2}\cos(k-1)\theta \\
\cos\theta \sin k\theta &= \tfrac{1}{2}\sin(k+1)\theta + \tfrac{1}{2}\sin(k-1)\theta \\
\sin\theta \cos k\theta &= \tfrac{1}{2}\sin(k+1)\theta - \tfrac{1}{2}\sin(k-1)\theta \\
\sin\theta \sin k\theta &= -\tfrac{1}{2}\cos(k+1)\theta + \tfrac{1}{2}\cos(k-1)\theta \,.
\end{aligned} \tag{A-6}$$

Since the identity is obvious for $m = 1$, $n = 0$ and $m = 0$, $n = 1$ the recursion shows it holds for all m and n.

In terms of multiple angles R and T become

$$\begin{aligned}
R = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} r^{i+j} \Big(a_{ij} \sum_{k=0}^{i+j+1} \{A_k^{i+1\ j} \sin k\theta + B_k^{i+1\ j} \cos k\theta\} + \\
b_{ij} \sum_{k=0}^{i+j+1} \{A_k^{i\ j+1} \sin k\theta + B_k^{i\ j+1} \cos k\theta\}\Big), \\
T = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} r^{i+j} \Big(-a_{ij} \sum_{k=0}^{i+j+1} \{A_k^{i\ j+1} \sin k\theta + B_k^{i\ j+1} \cos k\theta\} + \\
b_{ij} \sum_{k=0}^{i+j+1} \{A_k^{i+1\ j} \sin k\theta + B_k^{i+1\ j} \cos k\theta\}\Big) \,.
\end{aligned} \tag{A-7}$$

Inverting the series summation order produces

$$R = \sum_{k=0}^{\infty} \{\sin k\theta \sum_{i=i_1}^{\infty} \sum_{j=j_1}^{\infty} r^{i+j} (a_{ij} A_k^{i+1\ j} + b_{ij} A_k^{i\ j+1}) + \cos k\theta \sum_{i=i_1}^{\infty} \sum_{j=j_1}^{\infty} r^{i+j} (a_{ij} B_k^{i+1\ j} + b_{ij} B_i^{i\ j+1})\} \tag{A-8}$$

$$T = \sum_{k=0}^{\infty} \{\sin k\theta \sum_{i=i_1}^{\infty} \sum_{j=j_1}^{\infty} r^{i+j} (-a_{ij} A_k^{i\ j+1} + b_{ij} A_k^{i+1\ j}) + \cos k\theta \sum_{i=i_1}^{\infty} \sum_{j=j_1}^{\infty} r^{i+j} (-a_{ij} B_k^{i\ j+1} + b_{ij} B_k^{i+1\ j})\} ,$$

where $i_1 = \max(0, k-1)$ and $j_1 = \max(0, k-1-i)$.

Finally note that for $k = 0$ we have as the first non-zero term

$$r(a_{10} B_0^{20} + b_{10} B_0^{11} + a_{01} B_0^{11} + b_{01} B_0^{02}) \tag{A-9}$$

as a trigonometric identity causes the r^0 term to be zero. For $k = 1$ and higher, the first term is a trigonometric function times r^{k-1} as the inner summations have non-zero lower indices. Thus R and T in the m^{th} mode have power series starting with $r^{|m-1|}$, and we therefore interpolate $R/r^{|m-1|}$ and $T/r^{|m-1|}$ rather than

the unweighted displacements, so that continuity will be maintained. Also note that the second trigonometric identity forces both $R/r^{|m-1|}$ and $T/r^{|m-1|}$ to be power series in r^2 rather than r. This restriction on the expansion is handled by using only symmetric functions of r. We double the number of interpolation points and obtain their values by reflecting the original set over the $r = 0$ axis, thereby obtaining a symmetric set of interpolation points. This produces a symmetric spline curve for interpolation, and satisfies the second requirement for continuity.

APPENDIX II. COMPUTATIONAL METHODS FOR $a = 0.0$

As noted in Chapter III, the kernels of the force integrals of equation (51) are mildly singular for $a = 0.0$, and, as would be expected, our usual technique of integrating with equally-sized gaussian panels proved to be inaccurate even with twice the number of panels typically used for non-zero a. We therefore wished to find a convenient method of performing the integration which would be efficient, in the sense of minimizing the number of evaluations of the integrand required for a given accuracy.

Rather than explore the possibilities of transforming the variable of integration, or of using another type of integration in place of the gaussian, we decided as a first approach on merely optimizing the placement and sizes of a fixed number of our usual gaussian panels.

For reference, let us here recall that the integral of a function $f(x)$ from x_a to x_b is approximated according to many integration schemes by the weighted sum of several values of the function:

$$\int_{x_a}^{x_b} f(x)\, dx \simeq \sum_{i=1}^{n} w_i\, f(x_i) \qquad \text{(A-10)}$$

where

$$x_a < x_i < x_b,\ i = 1, \ldots . n.$$

The gaussian method differs from similar schemes in that the points x_i and weights w_i are chosen so that polynomials up to order $2n-1$ are perfectly computed. We make use of this method in the non-singular case by dividing up our total integration range into N_I equal panels, as noted in Appendix V, and by then using ten-point gaussian approximation on each of them.

Now, in the singular case we have $K_m(r,\rho)$ and $\partial/\partial r\ K_m(r,\rho)$ contributing singularities to the integrand of the form

$$K_m(r,\rho = r + \varepsilon) = - G \ln \varepsilon^2$$
$$\frac{\partial K_m}{\partial r}(r,\rho = r + \varepsilon) = - 2G/\varepsilon \quad . \tag{A-11}$$

Naturally, we should take advantage of the symmetry of these singularities, and place the panels in a symmetric fashion around the singular point $r = \rho$. The two obvious choices were to have the singular point be the center of one of the panels, or to have it be located at the boundary between two equally-sized panels. Since intuitively it seemed to us that a polynomial would fit one side of the singularity better than it would fit both sides simultaneously, we chose the latter course of action.

This left the question of how panel sizes should be arranged and allocated. However, since the majority of the contribution to the integral does not come in the vicinity of the singularity, but rather

from the more well-behaved regions, we cannot severely deplete these latter regions of panels. Yet the integral in the region of the singularity will not be well computed unless that region receives a larger than normal number of panels. With these considerations in mind, we concentrated on a family of arrangements which would satisfy these requirements and yet leave us some room for optimizing. The family studied was one in which each symmetric panel size on both sides of the singularity was increased by a constant ratio over the adjacent panel on the side toward the singularity, and, as shown in Figure A-1, we continued this pattern symmetrically away from the singularity until we came to one of the two boundaries, either the center or outer integration edge. The remaining region, shown here from A to E, was given a number of equally-sized panels. Within this scheme, two variables were still at our disposal, the ratio of adjacent panel sizes, and the number of panels allotted to the symmetric region OA as opposed to those in the remainder AE.

After a considerable amount of experimentation, we found working values for these two quantities which seemed superior to all others tried, in the sense that, for a given number of points, the integrals using these values more closely agreed with the integrals computed with very large numbers of panels. The final values used were: 1.5 for the adjacent panel ratio, and a panel distribution allotting the region AE half the number it would have received in the equal-sized distribution used for the non-singular cases. These values were then used for all of the $a = 0.0$ computations presented in this thesis.

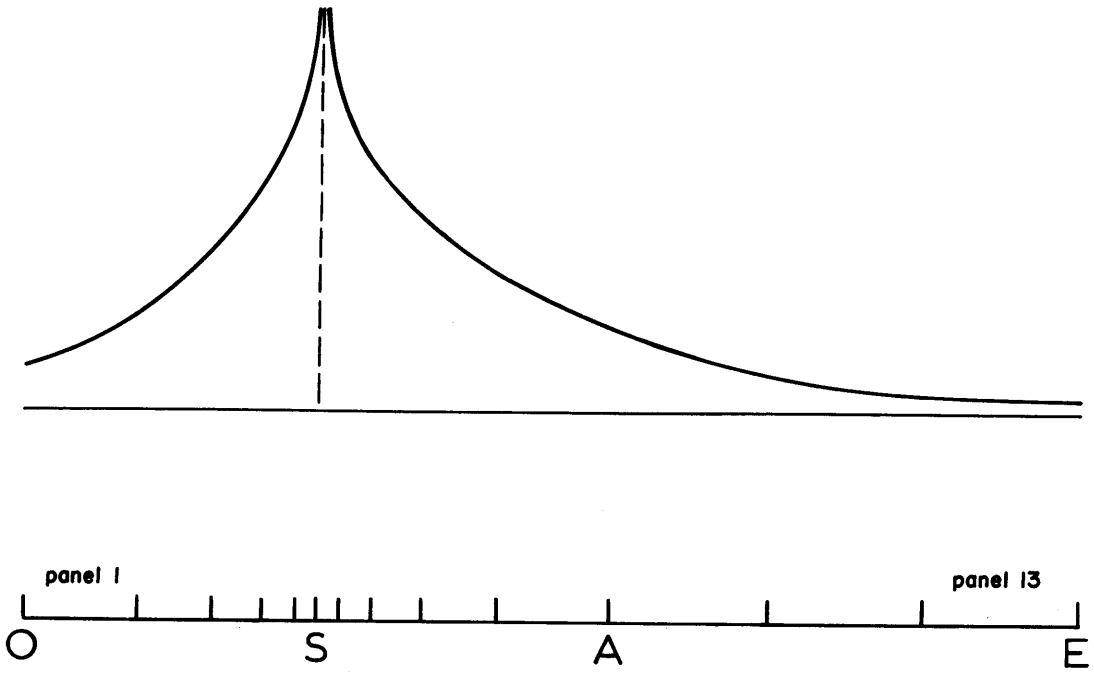

Figure A1. The arrangement of integration panels for the singular case a = 0.0 . Panels are progressively smaller nearer the singularity S in the region O (center) to A, and equally spaced from A to E (outer edge of integration).

Although this method is comparatively efficient, there is a higher computational cost involved in calculating the singular force matrix as opposed to the non-singular. In the latter we would use the same panel placement for each of the n^2 or $4n^2$ integrals (axisymmetric or non-axisymmetric), and this conveniently would allow us to compute many common quantities once for each evaluation location, rather than n^2 or $4n^2$ times. In the singular case, a total of n different panel arrangements were used so that these common quantities were computed n times as frequently as in the non-singular cases. Two other factors add to the cost of the singular cases, first, more panels are needed for a given accuracy, and second, on the small computer used for the majority of the calculations performed, the common quantities for the n different panel arrangements had to be swapped in and out of core n times each. For this compelling reason, no non-axisymmetric calculations were made in the singular case.

APPENDIX III. AXISYMMETRIC SHORT WAVELENGTH THEORY

From the abundant numerical agreements between theory and experiment listed in Chapter III, it is clear that SWLT accurately explains most of the behavior of axisymmetric perturbations. The purpose of this appendix is to derive formally the SWLT equations which we referred to in Chapter III. Phase and amplitude estimates will be discussed as well as the modifications to the theory necessary at the center. By itself, SWLT is not capable of predicting the behavior at resonance, and for this reason a special study was made of the resonance region. This study appears in the next appendix. The discussion of wave matching at cutoff is also considered in Appendix IV.

For the purpose of deriving SWLT, we will need to make two assumptions about relative magnitudes. The first is standard: wavelength is small compared to the scale of the disk. The second arises from our use of reduced gravity: the gravitational offset parameter a is also taken to be small compared with the scale of the disk, that is, of the same order as or smaller than the wavelength. Actually, these assumptions are formal and are useful in performing the derivation. As was seen in Chapter III, SWLT may still be surprisingly accurate even though the assumptions are not fulfilled.

Let us start with some definitions. Rather than carry out the derivation in terms of the displacements used in the calculations, we will use velocity variables as have been used in all previous SWLT formulations for thin disk galaxies, e.g. Lin and Shu (1964). From the outset, let us take the variation with time of all quantities to be proportional to $\exp(i\omega t)$, as we look only for modes of a fixed frequency. Let $\phi'(r,z)$ be the perturbation potential, at radius r and height z as computed with gravity reduction; let $u'(r)$ and $v'(r)$ be the radial and tangential perturbation velocities, and $\mu'(r)$ the perturbation density. As usual, equilibrium surface density is still denoted by $\mu(r)$.

First, we connect potential and surface density via Poisson's equation

$$\left(\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{\partial^2}{\partial z^2}\right)\phi' = 4\pi G \mu' \delta(z) \quad . \tag{A-12}$$

Here $\delta(z)$ is the Dirac delta function. Now let us make the short wavelength assumption by expanding ϕ' in terms of λ, a large parameter representing a typical wavenumber:

$$\phi' = \exp\left(i \lambda \{F_0(r,z) + \lambda^{-1} F_1(r,z) + \ldots\}\right) \quad . \tag{A-13}$$

By substituting this into Poisson's equation for $z \neq 0$ and ordering by powers of λ we obtain separate equations for $F_i(r,z)$. To

$O(\lambda^2)$ we find

$$\left(\frac{\partial F_0}{\partial r}\right)^2 + \left(\frac{\partial F_0}{\partial z}\right)^2 = 0 , \tag{A-14}$$

which has two separate, non-combinable solutions

$$F_0 = f_0(r + iz) \text{ and } F_0 = f_0^*(r - iz) \tag{A-15}$$

where f_0 and f_0^* are functions of one argument, as yet unspecified. Continuing, we find to $O(\lambda)$

$$-2\frac{\partial F_0}{\partial r}\frac{\partial F_1}{\partial r} + i\frac{\partial^2 F_0}{\partial r^2} + \frac{i}{r}\frac{\partial F_0}{\partial r} - 2\frac{\partial F_0}{\partial z}\frac{\partial F_1}{\partial z} + i\frac{\partial^2 F_0}{\partial z^2} = 0, \tag{A-16}$$

which reduces by cancellation into

$$\frac{\partial F_1}{\partial r} \mp i\frac{\partial F_1}{\partial z} - \frac{i}{2r} = 0 . \tag{A-17}$$

Here the upper sign is used when $F_0 = f_0$, the lower when $F_0 = f_0^*$. In the former case

$$F_1 = f_1(r + iz) + \frac{i}{2}\ln r \tag{A-18}$$

and in the latter

$$F_1 = f_1^*(r - iz) + \frac{i}{2} \ln r \qquad \text{(A-19)}$$

where f_1 and f_1^* are also unspecified functions.

We now make the connection with surface density by integrating Poisson's equation from 0^- to 0^+, which results in

$$\left.\frac{\partial \phi'}{\partial z}\right|_{0^+} - \left.\frac{\partial \phi'}{\partial z}\right|_{0^-} = 4\pi G \mu' \qquad \text{(A-20)}$$

and using the obvious symmetry above and below the disk

$$\left.\frac{\partial \phi'}{\partial z}\right|_{0^-} = 2\pi G \mu' \quad . \qquad \text{(A-21)}$$

Since SWLT is based on the existence of consistent local wavenumbers for all functions, we can write surface density in terms of the same wavenumber expansion as potential, with its own amplitude:

$$\mu' = \lambda^{n_1} (M_0 + \lambda^{-1} M_1 + \cdots) \exp i\lambda (F_0 + \lambda^{-1} F_1 + \cdots). \qquad \text{(A-22)}$$

In this expansion n_1 lending powers of λ have been appended to make possible a consistent set of first order equations, as we shall see shortly. With this expansion, equation (A-21) is, to leading

order in λ on both sides,

$$2\pi\ G\ \lambda^{n_1-1}\ M_0 \quad = \quad i\ \frac{\partial\ F_0}{\partial\ z} \quad . \tag{A-23}$$

Similarly, to the next order we find

$$2\pi\ G\ \lambda^{n_1-1}\ M_1 \quad = \quad i\ \frac{\partial\ F_1}{\partial\ z} \quad . \tag{A-24}$$

These can be written in terms of the two alternate solutions for ϕ':

$$M_0 \quad = \quad \frac{-1}{2\pi\ G\ \lambda^{n_1-1}}\ \frac{\partial\ f_0}{\partial\ r}\ , \quad M_1 \quad = \quad \frac{-1}{2\pi\ G\ \lambda^{n_1-1}}\ \frac{\partial\ f_1}{\partial\ r} \tag{A-25}$$

or

$$M_0 \quad = \quad \frac{1}{2\pi\ G\ \lambda^{n_1-1}}\ \frac{\partial\ f_0^*}{\partial\ r}\ , \quad M_1 \quad = \quad \frac{1}{2\pi\ G\ \lambda^{n_1-1}}\ \frac{\partial\ f_1^*}{\partial\ r}\ .$$

Now we can interpret the alternate solutions. To be physically realistic, we must have solutions for ϕ' which decay at large distances from the plane of the disk. In turn, the use of SWLT implies that potential is locally determined, and therefore potential must decay away from the plane of the disk even for small values of $|z|$. We can accomplish this by using the f_0 solution if

$$\text{Im}\left(\frac{\partial f_0}{\partial z}\, z\right) = \text{Re}\left(\frac{\partial f_0}{\partial r}\, z\right)$$

is positive and the f_0^* solution if

$$\text{Im}\left(\frac{\partial f_0^*}{\partial z}\, z\right) = \text{Re}\left(\frac{\partial f_0^*}{\partial r}\, z\right)$$

is positive, assuming that f_0 and f_0^* are mirror images of one another in the plane $z = 0$. This is equivalent to saying we will use the solution

$$f_0(r + is\,|z|)$$

at all times, where we choose s to be

$$s = \text{sgn}\left(\frac{\partial f_0}{\partial r}\right),$$

the sign of the radial derivative of f_0. Thus we no longer have to refer to the starred alternate solutions, but instead have a single form for ϕ' and μ':

$$\phi' = \exp i\lambda\,(f_0(r + is|z|) + \lambda^{-1}\left[f_1(r + is|z|) + \frac{i}{2}\ln r\right] + \ldots)$$

$$\mu' = \frac{-s\,\lambda}{2\pi\, G}\left(\frac{\partial f_0}{\partial r}(r) + \lambda^{-1}\frac{\partial f_1}{\partial r}(r) + \ldots\right). \qquad \text{(A-26)}$$

$$\exp i\lambda\,(f_0(r + is|z|) + \lambda^{-1}\left(f_1(r + is|z| + \frac{i}{2}\ln r\right) + \ldots)$$

Before leaving the solutions to Poisson's equations we have one last item to derive. According to our device of reduced gravity, the force that the particles feel is gauged by the derivative of the potential evaluated on a plane at a height a above the disk. Thus the force will be given by

$$-\frac{\partial \phi'}{\partial r} = i\lambda\left[\frac{\partial f_0}{\partial r}(r + isa) + \lambda^{-1}\left(\frac{\partial f_1}{\partial r}(r + isa) + \frac{i}{2r}\right) + \ldots\right]. \qquad \text{(A-27)}$$

$$\exp i\lambda\left[f_0(r + isa) + \lambda^{-1}(f_1(r + isa) + \frac{i}{2r}) + \ldots\right].$$

In order to be able to cancel the wavenumber exponent from the dynamical equations, we need to write this in terms of the original wavenumber plus a correction. Here we make use of second assumption, of the smallness of a, and write

$$f_0(r + isa) \simeq f_0(r) + isa\,\frac{df_0}{dr}(r) + \frac{(isa)^2}{2}\,\frac{d^2 f_0}{dr^2}(r) + \ldots \qquad \text{(A-28)}$$

and similarly for f_1 and their first derivatives. Let us rewrite the force expansion, keeping only those force terms which contribute to first or second order in λ^{-1}. We will assume that $a = O(\lambda^{-1})$ at most, as if it were much greater than λ^{-1} the perturbation force acting on the particles would be exponentially small. So we have

$$- \frac{\partial \phi'}{\partial r} = i\lambda\left[\frac{df_0}{dr}(r) + isa\frac{d^2f_0}{dr^2}(r) + \lambda^{-1}\left(\frac{df_1}{dr}(r) + \frac{i}{2r}\right)\right] \cdot$$

$$\exp i\lambda\left[f_0(r) + isa\frac{df_0}{dr}(r) - \frac{a^2}{2}\frac{d^2f_0}{dr^2}\right.$$

$$\left. + \lambda^{-1}\left(f_0(r) + isa\frac{df_1}{dr}(r) + \frac{i}{2}\ln r\right)\right] . \qquad \text{(A-29)}$$

Finally, we further expand this by writing the terms of $O(\lambda^{-1})$ in the exponent as a power series

$$- \frac{\partial \phi'}{\partial r} = i\lambda\left[\frac{df_0}{dr}(r) + isa\frac{d^2f_0}{dr^2}(r) + \lambda^{-1}\left(\frac{df_1}{dr}(r) + \frac{i}{2r}\right)\right] \cdot$$

$$\left[1 - i\lambda\frac{a^2}{2}\frac{d^2f_0}{dr^2}(r) - as\frac{df_1}{dr}(r)\right] \cdot$$

$$\exp i\lambda\left[f_0(r) + i\,sa\frac{df_0}{dr}(r) + \lambda^{-1}(f(r) + \frac{i}{2}\ln r)\right] .$$

(A-30)

We are now ready to examine the dynamical equations:

$$i\omega u' - 2\Omega v' = - \frac{\partial \phi'}{\partial r}$$

$$i\omega v' + \frac{\kappa^2}{2\Omega}u' = 0 \quad , \qquad \text{(A-31)}$$

which we combine, as usual, into

$$- \omega^2 u' + \kappa^2 u' = - i \omega \frac{\partial \phi'}{\partial r} \quad . \tag{A-32}$$

At this point we require an expansion of u':

$$u' = \lambda^{n_2} (U_0 + \lambda^{-1} U_1 + \ldots) \exp i\lambda (f_0(r) +$$

$$\lambda^{-1}(f_1(r) + \frac{i}{2} \ln r) + ..) , \tag{A-33}$$

which we immediately insert into the dynamical equation:

$$(\kappa^2 - \omega^2) \lambda^{n_2} (U_0 + \lambda^{-1} U_1 + \ldots) = \omega \lambda \left[\frac{df_0}{dr} (r) + i \, sa \frac{d^2 f_0}{dr^2} (r) +\right.$$

$$\left. \lambda^{-1} \left[\frac{df_1}{dr} (r) + \frac{i}{2r}\right]\right] \cdot \left[1 - i\lambda \frac{a^2}{2} \frac{d^2 f_0}{dr^2} (r) -\right.$$

$$\left. as \frac{df_1}{dr} (r)\right] \exp \left[-as \lambda \frac{df_0}{dr} (r)\right] , \tag{A-34}$$

Following the routine of SWLT, we decompose this in powers of λ:

$$(\kappa^2 - \omega^2) \lambda^{n_2 - 1} U_0 = \omega \frac{df_0}{dr} \exp (- as\lambda \frac{df_0}{dr} (r)) \tag{A-35}$$

$$(\kappa^2 - \omega^2)\,\lambda^{n_2-1}\,U_1 = \omega\left\{ i s a \lambda \frac{d^2 f_0}{dr^2}(r) + \frac{df_1}{dr}(r) + \right.$$

$$\frac{i}{2r} + \frac{df_0}{dr} \cdot \left\{ -i\lambda^2 \frac{a^2}{2} \frac{d^2 f_0}{dr^2}(r) - \right. \qquad \text{(A-36)}$$

$$\left.\left. as\,\lambda \frac{df_1}{dr}(r) \right\} \right\} \exp\left[-as\,\lambda \frac{df_0}{dr}(r) \right].$$

Having expanded both Poisson's law and the dynamical equations, it remains only to expand the continuity equation before we can balance and solve the entire set. The continuity equation:

$$i\omega\mu' + \frac{1}{r}\frac{d}{dr}(r\mu u') = 0 \qquad \text{(A-37)}$$

merely becomes

$$i\omega\lambda^{n_1-n_2-1} M_0 + \mu U_0\, i \frac{df_0}{dr} = 0 \qquad \text{(A-38)}$$

$$i\omega\lambda^{n_1-n_2-1} M_1 + \mu U_1\, i \frac{df_0}{dr} + \mu U_0\, i \frac{df_1}{dr} +$$

$$\frac{1}{r}\frac{d}{dr}(r\mu U_0) - \frac{1}{2r}\mu U_0 = 0\,. \qquad \text{(A-39)}$$

Let us use the first order equations to decide on the leading powers of λ:

Poisson's law	$O(G)\ O(\lambda^{n_1-1})$	=	$O(1)$
dynamics	$O(\lambda^{n_2})$	=	$O(\lambda)$
continuity	$O(\lambda^{n_1})$	=	$O(\lambda^{n_2+1})$,

where we have deliberately introduced G, the gravitational constant, into the balancing scheme as a device to satisfy all three equations simultaneously. The solution is thereby quite clear:

$$\begin{aligned} n_1 &= 2 \\ n_2 &= 1 \\ O(G) &= O(\lambda^{-1}) \ . \end{aligned}$$

With this order scheme, the solution of the total set of equations is now quite straightforward. For the first order solution we solve equation (A-37) for U_0 and replace U_0 in equation (A-35) with its equivalent:

$$(\kappa^2 - \omega^2) \left(\frac{-\omega M_0}{\mu \frac{df_0}{dr}} \right) = \omega \frac{df_0}{dr} \exp\left(-as\lambda \frac{df_0}{dr}(r)\right) . \qquad \text{(A-40)}$$

Then by using (A-25) to replace M_0 we find

$$(\kappa^2 - \omega^2) \left(\frac{-\omega \left(\frac{-s}{2\pi G\lambda}\right)}{\mu} \right) = \omega \frac{df_0}{dr} \exp\left(-as\lambda \frac{df_0}{dr}(r)\right) , \qquad \text{(A-41)}$$

or equivalently

$$\omega^2 = \kappa^2 - 2\pi G\mu \lambda s \frac{df_0}{dr} \exp\left(-as \lambda \frac{df_0}{dr}\right) . \qquad \text{(A-42)}$$

Of course, this is the dispersion relation, and it becomes more familiar by replacing $\lambda df_0/dr$ with k, a specific symbol for the wave-number:

$$\omega^2 = \kappa^2 - 2\pi G\mu |k| \exp(-a|k|) . \qquad \text{(A-43)}$$

This equation was the basis for most of our SWLT conclusions in the body of the thesis.

The second order solution is not much more difficult. If we make use of the first order result,

$$U_0 = \frac{\omega s}{2\pi G\lambda\mu} , \qquad \text{(A-44)}$$

equation (A-39) becomes

$$\frac{-i\omega s}{2\pi G\lambda} \frac{df_1}{dr} + \mu U_1 i \frac{df_0}{dr} + \frac{i\omega s}{2\pi G\lambda} \frac{df_1}{dr} + \frac{1}{r} \frac{\omega s}{2\pi G\lambda}$$

$$- \frac{1}{2r} \frac{\omega s}{2\pi G\lambda} = 0 \qquad \text{(A-45)}$$

or

$$\mu\, U_1\, i\, \frac{df_0}{dr} + \frac{1}{2r}\, \frac{\omega s}{2\pi\, G\lambda} \quad = \quad 0\ . \tag{A-46}$$

If we write equation (A-36) as

$$U_1 \quad = \quad \omega\, \left\{ i s a \lambda\, \frac{d^2 f_0}{dr^2} + \frac{df_1}{dr} + \frac{i}{2r} + \frac{df_0}{dr}\, \left[-i\, \lambda^2\, \frac{a^2}{2}\, \frac{d^2 f_0}{dr^2} - as\lambda\, \frac{df_1}{dr} \right] \right\}\ .$$

$$\left[\frac{\exp\, (-as\lambda\, \frac{df_0}{dr})}{\kappa^2\ -\ \omega^2} \right] \tag{A-47}$$

the rightmost term, in braces, can be replaced with

$$1/(2\pi\, G\mu\, \lambda\, s\, \frac{df_0}{dr})$$

by using the dispersion relation. This leads to

$$\frac{2\pi\, G\, \lambda\mu\, U_1\, \frac{df_1}{dr}}{\omega s} \quad = \quad i\ s\ a\ \lambda\, \frac{d^2 f_0}{dr^2} + \frac{df_1}{dr} + \frac{i}{2r} \quad +$$

$$\frac{df_0}{dr}\, (-\, i\, \lambda^2\, \frac{a^2}{2}\, \frac{d^2 f_0}{dr^2} - as\lambda\, \frac{df_1}{dr})\ , \tag{A-48}$$

but equation (A-46) can be rewritten as

$$\frac{2\pi\, G\lambda\mu\, U_1 \frac{df_0}{dr}}{\omega s} = \frac{i}{2r} \quad . \tag{A-49}$$

Thus equation (A-48) becomes

$$0 = i\, sa\, \lambda \frac{d^2f_0}{dr^2} + \frac{df_1}{dr} + \frac{df_0}{dr} \left(- i\, \lambda^2 \frac{a^2}{2} \frac{d^2f_0}{dr^2}\right.$$

$$\left. - a\, s\, \lambda \frac{df_1}{dr}\right) \tag{A-50}$$

which we solve for the unknown df_1/dr

$$\frac{df_1}{dr} = \left(\frac{- isa\, \lambda - i\, \lambda^2 \frac{a^2}{2} \frac{df_0}{dr}}{1 - as\lambda \frac{df_0}{dr}}\right) \frac{d^2f_0}{dr^2} \tag{A-51}$$

and integrate by rewriting as

$$\frac{df_1}{dr} = \frac{i}{2} \frac{\frac{d}{dr}\left(1 - as\lambda \frac{df_0}{dr}\right)}{\left(1 - as\lambda \frac{df_0}{dr}\right)} - \frac{i}{2} as\lambda \frac{d^2f_0}{dr^2} \quad . \tag{A-52}$$

Thus

$$f_1 = \frac{i}{2} \ln (1 - a\, s\, \lambda \frac{df_0}{dr}) - \frac{i}{2} as\lambda \frac{df_0}{dr} \qquad \text{(A-53)}$$

plus an inconsequential constant.

To second order, the perturbation velocity u' can be written now:

$$u' = \lambda\left(\frac{\omega s}{2\pi G\lambda\mu} + \lambda^{-1} \frac{i\omega s}{2\pi G\lambda\mu\ 2r \frac{df_0}{dr}}\right) \exp$$

$$i\lambda \left[f_0(r) + \lambda^{-1} \left(\frac{i}{2} \ln (1 - as\lambda \frac{df_0}{dr}) - \right.\right.$$

$$\left.\left. \frac{i}{2} as\lambda \frac{df_0}{dr} + \frac{i}{2} \ln r\right)\right] \qquad \text{(A-54)}$$

or

$$u' = \frac{\omega s}{2\pi G\mu} \left(1 + \frac{i}{2r\ \lambda \frac{df_0}{dr}}\right) \frac{1}{\sqrt{r(1-as\lambda\frac{df_0}{dr})}} \cdot$$

$$\exp \left(i\ \lambda\ f_0 + \frac{as\lambda}{2} \frac{df_0}{dr}\right) . \qquad \text{(A-55)}$$

In terms of the wavenumber $k = \lambda\ df_0/dr$, we can write this as

$$u' = \frac{\omega s}{2\pi G\mu}\left(1 + \frac{i}{2rk}\right)\frac{1}{\sqrt{r(1-a|k|)}}\exp$$

$$\left(i\int_0^r k\,dr' + \frac{a}{2}|k|\right) . \qquad \text{(A-56)}$$

Except for the $O(k^{-1})$ term in the first pair of parentheses, this formula can be derived in a fast, more intuitive way by appealing to conservation of energy. As noted by Toomre (1969), for axisymmetric waves, wave energy is maintained while propagating radially, and furthermore the velocity at which it propagates is $\partial\omega/\partial k$. From the dispersion relation, equation (A-43) we can find that

$$\partial\omega/\partial k = -2\pi G\mu \exp(-a|k|)(1-a|k|)\, s/2\omega . \qquad \text{(A-57)}$$

Since wave energy is proportional to the amplitude of kinetic energy, $\frac{1}{2}\mu u'^2$, we immediately deduce that

$$u' \propto \sqrt{\frac{2}{\mu r}\left[\frac{\partial\omega}{\partial k}\right]^{-1}}$$

by insisting that the energy passing through any circumference is constant. Thus we again find

$$u' \propto \frac{1}{\mu}\,\frac{1}{\sqrt{r}\sqrt{\exp(-a|k|)}}\,\frac{1}{\sqrt{1-a|k|}} \qquad \text{(A-58)}$$

just as before.

Now let us analyze the center of the disk. In the regular derivation just finished, we solved Poisson's equation under the assumption $r = O(1)$. Clearly the solution obtained is not valid in the center. However, a complete set of nonsingular solutions to Poisson's equation for potential in polar coordinates can be written as merely

$$\phi'(r,z) = C\, J_0(wr)\, e^{-wr} \tag{A-59}$$

where C and w are arbitrary constants. We will use a single solution of this form as the inner solution to our set of equations, valid over a small region in the center where wavenumber is approximately constant. Then, we must match this solution with the one just derived which is valid for larger radii. The center solution, written asymptotically in r, is

$$\phi'(r,z) \simeq C\, e^{-wz} \sqrt{\frac{2}{\pi wr}} \sin(wr - \pi/4)$$

$$= C\, e^{-wz} \sqrt{\frac{1}{2\pi wr}}\, i\, (e^{i(wr-\pi/4)} - e^{-i(wr-\pi/4)}) . \tag{A-60}$$

Thus two of the outer solutions can be summed to match this, resulting in

$$\phi'(r,z) = C \frac{1}{\sqrt{2\pi \frac{df_0}{dr}(0)\, r}} i \left(\exp\left(i\lambda f_0 + i f_1 - i\pi/4\right) - \right.$$

$$\left. \exp\left(-i\lambda f_0 + i f_1 + i\,\pi/4\right)\right) \qquad \text{(A-61)}$$

where we note that the coefficient of f_1 does not change when f_0 is replaced by $-f_0$ because f_1 is a real function of $|f_0|$. We can replace the leading constants with unity by a proper choice of C, leaving the solution for $z = 0$ as

$$\phi'(r,0) = \frac{1}{\sqrt{r}} \frac{1}{\sqrt{\exp\left(-a|k|\right)}} \frac{1}{\sqrt{1-a|k|}} \cdot$$

$$\cos\left(\int_0^r k\, dr' - \pi/4\right) . \qquad \text{(A-62)}$$

Similarly, the velocity u' becomes

$$u'(r) = \frac{\omega}{2\pi G\mu} \frac{1}{\sqrt{r}} \frac{1}{\sqrt{1-a|k|}} \frac{1}{\sqrt{\exp(-a|k|)}} \cdot$$

$$\left(\sin\left(\int_0^r k\, dr' - \pi/4\right)\right.$$

$$\left. + \frac{1}{2rk} \cos\left(\int_0^r k dr - \pi/4\right)\right) . \qquad \text{(A-63)}$$

APPENDIX IV. CONCLUSIONS FROM A MODEL EQUATION

In this appendix we will be studying a particular approximation to the axisymmetric dynamical equations, designed specifically to determine the phase condition at the resonance radius. For this purpose, no attention will be given to the global variation of parameters. Instead, we will be drastically modifying the dynamical equations so as to make them nearly soluble analytically.

1. Approximation

The model equation which we will shortly be deriving can most concisely be described as the resultant equation in the limit of zero perturbation density. The dispersion relation indicates that as the density participating in the perturbation declines, the wavelengths involved in the solution will proportionately decline, assuming of course that we similarly diminish our gravity offset a. Suppose that we rescale distances on the basis of either wavelength or a, all the while keeping the resonance position fixed in space. First of all the center of the disk will recede to infinity, making curvature effects disappear. Second, equilibrium density will approach a constant. And finally, $\kappa^2(r) - \omega^2$ will become linear with a single zero at

our fixed resonance point. For convenience, we may take that resonance radius to be the center of the stretched coordinate, which we label r to indicate its original purpose.

Under these three conditions, the dynamical equations are greatly simplified and much more amenable to analysis. However, the approximation process has so drastically altered the equations that we expect no possible comparison of prediction between the model equation solution and SWLT for the original disk. Instead, we can expect that certain invariant conditions, such as the boundary condition at resonance, will remain the same, and now be derivable -- this is the purpose of the modeling.

Because the assumption we make concerning $\kappa^2(r)$ is nonphysical -- that it be linear over an infinite range -- we cannot necessarily assume that a solution of the model equation exists. Instead, to interpret the results of the formal solution we found it necessary to apply a slowly growing force to the model, and then allow the growth rate to diminish. The usefulness of this device will be evident later when we examine the formal solutions of the equations.

2. Derivation

The fundamental equation for the model is of course the dynamical equation of Chapter III, which we write now as

$$\frac{\partial^2 x'}{\partial t^2} + \kappa^2(r)\, x' = -\frac{\partial \phi'}{\partial r} + F \qquad \text{(A-64)}$$

Here $x'(r,t)$ is again the real radial displacement, ϕ', the corresponding potential and F, an applied force. We will be examining the response of our model equation to an applied force with wavenumber s^* and with time dependence of the form

$$F = \mathrm{Re}\{A\, e^{is^*r}\, e^{i\omega t + gt}\} \tag{A-65}$$

where ω is a real frequency and g a small growth rate. Because of the linearity of equation (A-64), we will want to assume that x' and ϕ' have the same time dependence in their solutions

$$\begin{aligned} x'(r,t) &= \mathrm{Re}\{x(r)\, e^{i\omega t + gt}\} \\ \phi'(r,t) &= \mathrm{Re}\{\phi(r)\, e^{i\omega t + gt}\} \ . \end{aligned} \tag{A-66}$$

Owing to the spatial inhomogeneity in our model equation, these two functions will by no means have the same wavenumber dependence as F.

Let us now introduce the assumption that

$$\kappa^2(r) = \omega^2 + Dr \tag{A-67}$$

which puts resonance at $r = 0$ in the limit of small g. D is here a positive constant. Then equation (A-64) becomes

$$(Dr + 2i\omega g + g^2)\, x = - \, d\phi/dr + A\, e^{is^*r} \ . \tag{A-68}$$

Having the dynamics completed, we can turn to the evaluation of the potential. Since we have approximated away all curvature effects in the model equation, the perturbation force $-d\phi/dr$ can be regarded as the sum of the forces from a set of infinitely long thin rods, oriented perpendicularly to the coordinate r, and each possessing a total line mass density $\bar{\mu}(r)dr$:

$$-\frac{d\phi}{dr} = -\int_{-\infty}^{\infty} \bar{\mu}(\rho)\, d\rho\, \frac{2G\,(r-\rho)}{a^2 + (r-\rho)^2} \qquad \text{(A-69)}$$

Of course $\bar{\mu}(r)$ is the surface density. Note also that we are here using reduced gravity with offset a. Owing to our neglect of curvature forces, $\bar{\mu}(r)$ is merely $-\mu \frac{dx}{dr}$ where μ is the constant equilibrium density participating in the perturbation. By simply integrating by parts we can write the above as

$$-\frac{d\phi}{dr} = 2G\,\mu \int_{-\infty}^{\infty} x(\rho) \left\{ \frac{a^2 - (r-\rho)^2}{\{a^2 + (r-\rho)^2\}^2} \right\} d\rho \quad . \qquad \text{(A-70)}$$

Then for convenience, let us designate the term in braces as the kernel $K(r-\rho)$. Thus the final dynamical equation emerges as

$$(Dr + 2\,i\omega g + g^2)\,x = 2\,G\,\mu \int_{-\infty}^{\infty} K(r-\rho)\, x(\rho)\, d\rho + A\, e^{is^*r} \quad . \qquad \text{(A-71)}$$

3. Formal Solution

The convolution form of equation (A-71) is well-suited to be treated by Fourier transforms. Let us denote the transform of $x(r)$ by $\underline{X}(s)$, defined so that

$$x(r) = \int_{-\infty}^{\infty} \underline{X}(s)\, e^{irs}\, ds \qquad \text{(A-72)}$$

and use the integral to replace $x(r)$ in equation (A-71):

$$\int_{-\infty}^{\infty} (Dr + 2i\omega g + g^2)\, \underline{X}(s)\, e^{irs}\, ds = 2G\mu \int_{-\infty}^{\infty} K(r-\rho) \int_{-\infty}^{\infty} \underline{X}(s)\, e^{irs}\, ds\, d\rho + A\, e^{is^*r}. \qquad \text{(A-73)}$$

In textbook fashion we modify the first term on the left-hand side by integrating by parts

$$\int_{-\infty}^{\infty} Dr\, \underline{X}(s)\, e^{irs}\, ds = -i\, \underline{X}(s)\, De^{irs}\Big|_{-\infty}^{\infty} + D \int_{-\infty}^{\infty} e^{irs}\, \frac{d\underline{X}}{ds}\, ds\,, \qquad \text{(A-74)}$$

and then carry on with the solution by temporarily assuming that $\underline{X}(s)$ goes to zero at $\pm\infty$.

Similarly, the standard modification for the convolution term is merely

$$2G\mu \int_{-\infty}^{\infty} K(r-\rho) \int_{-\infty}^{\infty} e^{i\rho s}\, \underline{X}(s)\, ds d\rho = 2G\mu \int_{-\infty}^{\infty} \underline{X}(s)\, e^{irs}\, ds \int_{-\infty}^{\infty} e^{i(\rho-r)s} K(\rho-r)\, d(\rho-r)\,. \tag{A-75}$$

The inner integral on the right-hand side

$$\begin{aligned} \int_{-\infty}^{\infty} e^{iys}\, K(y)\, dy &= \int_{-\infty}^{\infty} e^{iys}\, \frac{a^2 + y^2}{(a^2 + y^2)^2}\, dy \\ &= \int_{-\infty}^{\infty} e^{iys} \left(\frac{2a^2}{(a^2 + y^2)^2} - \frac{1}{a^2 + y^2} \right) dy \end{aligned} \tag{A-76}$$

is easily evaluated. Of the pair of terms, the second is soluble in terms of contour integration as

$$\int_{-\infty}^{\infty} e^{iys} \left(\frac{1}{y + ia} - \frac{1}{y - ia} \right) \frac{1}{2ia}\, dy\,. \tag{A-77}$$

If Re s is greater than zero we close the contour in a semicircle of arbitrarily large radius above the Re y axis, incorporating the pole at ia. If Re s is less than zero, we close it below. Consequently, the second integral has a value of

$$\frac{\pi}{a} e^{-a\sigma s}$$

where σ = sgn Re s. For real s we can write this as

$$\frac{\pi}{a} e^{-a|s|}\,.$$

The first integral is evaluatable as a derivative in a of the first, giving us a final value for the whole integral of

$$\pi \, |s| \, e^{-a|s|}$$

for real s and a similar formula for complex s.

This in turn leads to a form suitable for replacement in the transformed dynamical equation:

$$2\pi \, G \int_{-\infty}^{\infty} \underline{X}(s) \, e^{irs} \, ds \, \pi \, |s| \, e^{-a|s|} \, . \tag{A-78}$$

Lastly, in the same spirit of Fourier transforms we write the applied force term as

$$A \, e^{is^*r} \quad = \quad A \int_{-\infty}^{\infty} e^{irs} \, \delta(s-s^*) \, ds, \tag{A-79}$$

and this enables us to write equation (A-71) as a single Fourier transform:

$$\int_{-\infty}^{\infty} e^{irs} \left[iD \frac{d\underline{X}}{ds} + (2i\omega g + g^2) \, \underline{X} - 2\pi \, G\mu \, |s| \, e^{-a|s|} \, \underline{X} - A \, \delta(s-s^*) \right] ds \quad = \quad 0, \tag{A-80}$$

which is equivalent to

$$i\,D\,\frac{d\underline{\overline{X}}}{ds} + (2i\omega g + g^2)\,\underline{\overline{X}} - 2\pi\,G\mu\,|s|\,e^{-a|s|}\,\underline{\overline{X}} = A\,\delta(s-s^*), \tag{A-81}$$

a differential equation for the transformed variable.

The solution of equation (A-81) is straightforward, and leads to a free or unforced solution:

$$\underline{\overline{X}} = \underline{\overline{X}}_0 \exp \frac{i}{D}\{(g^2 + 2i\omega g)\,s - 2\pi\,G\mu\,\mathrm{sgn}\,s\,\left(1-e^{-a|s|}\cdot\frac{|s|}{a} + \frac{1}{a^2}\right)\} \tag{A-82}$$

Here $\underline{\overline{X}}_0$ is a constant of integration. Now this solution leads to a non-convergent integral for the Fourier inversion for any choice of g, as there is a real multiple of g in the exponent. Even for zero g we are faced with an inversion integral whose behavior at large $|s|$ is given by

$$\int \underline{\overline{X}}_0 \exp\left(irs - \frac{2\pi G\mu}{D}\right) ds \tag{A-83}$$

which is related to a delta function at $r = 0$. In fact, this $\delta(r)$ behavior can clearly be seen to arise in the direct equation (A-71) with the right-hand side set equal to zero.

Fortunately a forced solution can be found without these nuisances. The forced solution of equation (A-81) can immediately be obtained from the free solution by an integration. Since the forcing is a delta function in wavenumber, the forced solution is immediately available:

$$\overline{X} = A\, H(s-s^*) \exp \frac{i}{D} \{(g^2 + 2ig)\, s - 2\pi G\mu \operatorname{sgn} s \left(1-e^{-a|s|} \left(\frac{|s|}{a} + \frac{1}{a^2} \right)\right)\} \qquad \text{(A-84)}$$

Where $H(s-s^*)$ is a unit step function occurring at $s = s^*$. The inversion integral of this function clearly converges for any finite s^* and positive g. Furthermore, the form of equation (A-84) makes plain the difficulty with the free solution. Suppose s^* is zero, or in other words we decide to force the mode uniformly in space. Only positive wavenumbers take on non-zero amplitudes. In general, disturbances at one wavelength only turn into disturbances at algebraically large wavelengths, never at smaller ones -- the flow in wavenumber space is unidirectional. Thus we see that the free solution does not converge simply because it is required to have disturbances at arbitrarily large negative wavenumber, which will continually supply the flow; and the sum of these disturbances is non-integrable.

Let us make the final point that the forced solution is identical with the free solution except for the non-convergent tail at arbitrarily large negative wavenumbers. For our purposes we can simply use the forced solution as we are concerned primarily with the phase conditions

on the long wavelength waves near the resonance radius. These waves are clearly not affected by waves with arbitrarily large wavenumbers, as is demonstrated by the asymptotic inversion of equation (A-84).

4. Long Waves at Resonance

Before proceeding with the investigation of the phase conditions at resonance, let us transform the variables into a slightly more meaningful set. First we will scale both wavenumber and distance by a, writing

$$u = as \qquad u^* = as^*$$

$$R = r/a \qquad R_c = 2\pi\, G\mu/a^2\, e\, D$$

where R_c is the cutoff radius where the long and short wavelength branches of the dispersion relation have the same wavenumber, $u = 1$. Then we have

$$x(R) = \frac{A}{a} \int_{u^*}^{\infty} \exp i \left\{ Ru + \frac{1}{D} (g^2 + 2i\omega g)\, u - R_c\, e\, \operatorname{sgn} u \left[1 - (1 + |u|)\, e^{-|u|} \right] \right\} du \,. \tag{A-85}$$

This integral will be evaluated by the method of steepest descent for some values of R. Consider the relation for the zeroes of the derivative of the exponent:

$$i\,\{R + \frac{1}{D}\,(g^2 + 2i\omega g) - R_c\, e|u|\, e^{-|u|}\} = 0 \qquad \text{(A-86)}$$

for $g = 0$, this has a solution for real $|u|$ only for R between 0 and R_c, just where SWLT, applied to the model equation would be valid. In order to find the phase condition, we therefore should evaluate this integral for small positive R to obtain the long-wavelength branches of the solution.

At this point there is no difficulty in making an asymptotic approximation to the long-wavelength contribution, x_L, to the integral (A-85). Two terms occur in this contribution, amounting to steepest descent evaluation at $\pm u = \pm(u_r + iu_i)$, where $u_r > 0$:

$$x_L(R) \simeq \frac{A}{a}\left\{\frac{\exp i\,\{Ru + \left(\frac{g^2}{D} + \frac{2i\omega g}{D}\right)u + R_c\, e\,\left[1 - (u+1)\exp(-u)\right]}{|R_c\, e(1-u)\exp(-u)|^{\frac{1}{2}}}\right.$$

$$\sqrt{2\pi}\,\exp i\{-\pi/4 - 1/2\ \text{phase}\left[(1-u_r-u_i^2) - i\,(2u_i - u_r\,u_i)\right]\}$$

$$+\ \frac{\exp i\,\{-Ru - \left(\frac{g^2}{D} + \frac{2i\omega g}{D}\right)u - R_c\, e\,\left[1-(u+1)\exp(-u)\right]}{|R_c\, e\,(1-u)\exp(-u)|^{\frac{1}{2}}}\cdot$$

$$\left.\sqrt{2\pi}\,\exp i\{\pi/4 - 1/2\ \text{phase}\,\left[(1-u_r-u_i^2) - i\,(2u_i - u_r\,u_i)\right]\}\right\}. \qquad \text{(A-87)}$$

Since problems of convergence are now behind us, we can take the limit of $g \to 0$, after which $x_L(R)$ appears as

$$x_L(R) \simeq \sqrt{2\pi}\,\frac{A}{a}\,\frac{2}{\left|R_c\, e\, (1-u)\, \exp\,(-u)\right|^{\frac{1}{2}}}\cos\{Ru - R_c\, e\left[1 - e^{-u}\,(1+u)\right] - \pi/4\}\,. \qquad \text{(A-88)}$$

At resonance, $R = 0$, the phase of the long-wave contribution is $-\pi/4$, which is to be compared with $\pi/2$ for the first node in the direction of cutoff. Thus the resolution of the phase question is that a phase change of $3\pi/4$ exists between the last node and the resonance radius for longer waves.

Transferred back to the original axisymmetric problem, this phase condition determines the rule for discrete modes: Total phase change from center to resonance must be $n\pi$, with contributions of $5\pi/4$ from center to first node arising from Bessel function behavior at the center, $(n-2)\pi$ in internodal distance, and $3\pi/4$ from last node to resonance. If no nodes occur between center and resonance, the $\pi/4$ surplus from the center combines with the $\pi/4$ deficit from the resonance to make the total phase change merely π.

5. Mixing at Cutoff

We can also use the linear κ^2 model as a vehicle to study the solution near cutoff, as the scaling was chosen to leave the dis-

tance to cutoff, R_c, finite. The use of asymptotics should go through without change from resonance to cutoff, inasmuch as the formal solution is identical between 0 and R_c. For radii beyond R_c, however, a modification will be needed as the steepest descent points now move into the complex plane, even in the limit of $g \to 0$.

By using the above procedure of asymptotic evaluation of the integral now for the shortwave contribution, we find the combined solution in the wave-like region to be

$$x(R) = \sqrt{2\pi}\,\frac{A}{a}\,2\left\{\frac{\cos\left\{Ru - R_c e\left[1 - e^{-u}(1+u) - \pi/4\right]\right\}}{\left|R_c e\,(1-u)\exp(-u)\right|^{\frac{1}{2}}}\right. \tag{A-89}$$

$$\left. + \frac{\cos\left\{R\bar{u} - R_c e\left[1 - e^{-\bar{u}}(1+\bar{u}) + \pi/4\right]\right\}}{\left|R_c e\,(1-\bar{u})\exp(-\bar{u})\right|^{\frac{1}{2}}}\right\}.$$

Here we have used the symbol $\bar{u}$ to represent the short wavelength solution of equation (A-86), after the limit $g \to 0$ has been taken. Note the principal point of interest here, at cutoff, $R = R_c$, we have $u = \bar{u}$ and the phases of the two contributions differ by exactly $\pi/2$, although the value of these phases is not a fixed value but a function of R_c. This condition on phase difference combined with the center boundary condition will provide a specification of the discrete frequencies of the unstable modes, containing an obvious mixture

of both constituents. Make note however, that this condition is much less exact than that at resonance, as we are evaluating the contributions from two nearby steepest descent points as if they were widely separated.

For radii beyond $R = R_c$, equation (A-86) has no real solution -- this of course is one of the defining conditions for cutoff. However, we can still use steepest descent by finding the solutions in the complex plane. Let $u = u_r + iu_i$ with u_r positive -- then we desire the solutions to

$$\begin{aligned} e^{-u_r}(u_r \cos u_i + u_i \sin u_i) &= R/R_c e \\ e^{-u_r}(u_i \cos u_i - u_r \sin u_i) &= 0 \quad , \end{aligned} \tag{A-90}$$

which is the complex form of the equivalent of equation (A-86). These two equations are even and odd in u_i respectively, thereby implying that if $u_r + iu_i$ is one solution so is $u_r - iu_i$. Inasmuch as the actual equation (A-86) involves not u_r but $u_r \operatorname{sgn} u_r$, two further solutions are $-u_r + iu_i$, $-u_r - iu_i$, making four corners of a rectangle in the complex u plane.

The locus of the complex values of u traces out a symmetrical pair of U-shaped curves as R moves further beyond R_c. For R inside R_c, and u_r positive, there are two solutions of equation (A-86) on the positive half of the real line, one inside $u = 1$ and

one outside. By $R = R_c$ these two points have met and they then move perpendicularly off the real u axis; moving along curves which bend back toward the Im u axis to intersect it when $u = \pi/2\, i$ which corresponds to a radial point $R_s = \frac{1}{2}\pi e\, R_c$.

Because of the symmetry of equation (A-86) in u_r, the loci of negative values of u corresponding to real solutions of this equation must be mirror images of those for positive u_r. Thus the continuous change in wavenumber must stop when $R = R_s$, which might be called a second cutoff, as it marks the edge in R-space of decaying (and growing) sinusoidal waves, just as R_c marks the edge of non-decaying sinusoidal waves.

Since u is the effective local wavenumber, it is plausible that only solutions with positive u_i are realistic, as the lower two correspond to radially growing waves. Therefore to evaluate the Fourier inversion integral for radii beyond R_i, we will want to take a contour through the two points with positive u_i. Under this assumption, the use of steepest descent asymptotics immediately leads to the result:

$$x(R) \simeq 2\frac{A}{a}\sqrt{2\pi}\exp\left\{-\mathrm{Im}\left[Ru + R_c e\,(1 - \exp(-u)(1+u))\right]\right\} \cdot \frac{\cos\left\{\mathrm{Re}\left[Ru + R_c e\,(1 - \exp(-u)(1+u) - P/2 + \pi/4\right]\right\}}{\left|R_c e\,(1-u)\exp(-u)\right|^{\frac{1}{2}}} \tag{A-91}$$

where P = phase $(1-u)$ - phase (u).

Differentiating the real exponent with respect to R will produce an effective rate of exponential decay, which can only be equal to the imaginary part of the effective wavenumber, $\mathrm{Im}\, u$. This decay rate is zero exactly at cutoff and increases as u_i rises to the maximum value of $\pi/2$ when $R = R_s$. Beyond R_s our straightforward use of asymptotic evaluation is not valid, but we might surmise that the exponential decay turns into a decay function of another form, perhaps algebraic.

APPENDIX V. CONTROL OF ERRORS IN COMPUTATION

In order to eliminate the possibility that computational errors could corrupt the data calculated for the model, checks were made at each separate stage on the accuracy of computation. Basically there are three main steps in the computational procedure: first, the evaluation of the force matrix and the equilibrium frequencies, $\kappa^2(r)$ and $\Omega^2(r)$, evaluated at the sampling radii; second, the determination of the eigenvalues of the dynamic matrix; and third, the evaluation of the eigenfunctions corresponding to each eigenvalue. Each of these was verified by an independent procedure.

As noted in Chapter II, the force integrals take the form

$$\int_0^\infty Q_{m-\frac{1}{2}}\left(\frac{a^2 + r^2 + \rho^2}{2r\rho}\right) S(\rho)\, A(r,\rho)\, d\rho \qquad \text{(A-92)}$$

where $Q_{m-\frac{1}{2}}$ is a Legendre function of order $m - \frac{1}{2}$, S is the density function calculated by interpolation between values calculated at sampling radii and $A(r,\rho)$ is an algebraic function of r and ρ. Difficulties could conceivably arise in the accurate calculation of the Legendre function, the interpolation procedure, or in the integration itself.

Legendre functions were calculated by evaluating two expansions, a singular one for arguments near 1, and an asymptotic one for large arguments. The asymptotic formula is the one given by Gradshteyn and Ryzhik (1965) for toroidal functions

$$Q^{k}_{m-\frac{1}{2}}(\cosh \eta) = (-1)^{k} \frac{2^{k}\, \Gamma(m + k + \frac{1}{2})\sqrt{\pi}}{\Gamma(m + 1)} \sinh^{k} \eta\, e^{-(m+k-\frac{1}{2})} \cdot F(k+\tfrac{1}{2},\ k+m+\tfrac{1}{2};\ m+1;\ e^{-2\eta}) \qquad \text{(A-93)}$$

where $Q^{k}_{m-\frac{1}{2}}(\cosh \eta)$ is the associated Legendre function of the second kind, of order $m - \frac{1}{2}$ and k. By substituting $z = \cosh \eta$ and letting $k = 0$, we obtain

$$Q_{m-\frac{1}{2}}(z) = \frac{\Gamma(m+\frac{1}{2})}{\Gamma(m+1)} \sqrt{\pi}\, e^{-(m+\frac{1}{2})} F\left(\frac{1}{2},\ m + \frac{1}{2};\ m + 1;\ \frac{1}{(z + \sqrt{z^2-1})^2}\right). \qquad \text{(A-94)}$$

$F(a, b; c; \rho)$ is the hypergeometric function.

For the inner expansion we use the expansion of $Q^{\mu}_{\nu}(z)$ in terms of $\frac{z - 1}{z + 1}$ of Erdélyi et al (1953):

$$Q_\nu^\mu(z)\, e^{-i\mu\pi} = 2^{-1-\nu}\,\Gamma(\mu)\,(z+1)^{\nu+\frac{1}{2}\mu}\,(z-1)^{-\frac{1}{2}\mu}\cdot$$

$$F(-\nu, -\nu+\mu; 1-\mu; (z-1)/(z+1)) +$$

$$2^{-1-\nu}\,\Gamma(1+\nu+\mu)\,\Gamma(-\mu)\,(z+1)^{-\frac{1}{2}\mu+\nu}\cdot$$

$$(z-1)^{\frac{1}{2}\mu} / \Gamma(1+\nu-\mu)\cdot \qquad \text{(A-95)}$$

$$F(-\nu, -\nu+\mu; 1+\mu; (z-1)/(z+1)) .$$

It is necessary to take the limit of this expansion as $\mu \to 0$. The resulting expression,

$$Q_\nu(z) = 2^{-\nu}\,(z+1)^\nu \left[\{-\psi(1+\nu) - \tfrac{1}{2}\ln\left((z+1)/(z-1)\right) - \gamma\}\cdot F\left(-\nu, -\nu; 1; (z-1)/(z+1)\right) - \frac{1}{\Gamma^2(-\nu)} \sum_{k=0}^{\infty} \frac{\Gamma^2(-\nu+k)}{(k!)} \left(\frac{z-1}{z+1}\right)^k \left(\sum_{j=1}^{k} \frac{1}{-\nu+(j-1)} - \frac{1}{1+(j-1)}\right)\right]$$

(A-96)

is particularly easy to evaluate as the hypergeometric series and the latter expansion have almost identical terms. ψ is the psi (digamma) function and γ, Euler's constant.

These formuli and the programs calculating them were verified by comparing them with one another and with the tables of Abramowitz

and Stegun (1964). Accuracy of better than 10^{-9} was maintained at all times. It was empirically found that a crossover point of $z = 1.4$ gave the most efficient calculation at this accuracy.

The initial interpolation procedure used was that of cubic splines (Ahlberg, Nilson and Walsh, (1967)). These functions are piecewise cubic and continuous everywhere, with specified discontinuities in the derivatives at the points being interpolated, through which the interpolation curve is made to fit. The simple cubic spline has continuous first and second derivatives, with the third derivative discontinuous. This spline can be visualized at the curve taken by a very thin beam, which is forced to pass through the points being interpolated by forces perpendicular to the beam.

Since interpolation of the displacements between the sampling radii is one of the major techniques in our approximate calculation of eigenfrequencies, the question should be raised as to why splines were used, rather than another set of functions, for example, Chebyshev or Lagrange polynomials. Polynomial methods have great advantage in interpolating smooth functions, but the functions typically interpolated here are those in which displacements are zero at all sampling radii but one, where it is unity. This particularly non-smooth function causes the polynomial interpolations to have large contributions away from the displaced point. On the other hand, the splines tame the displacements between sampling locations far from the displaced point

rather well. This localization of a unit displacement means that the forces arising from mass displacement at a single sampling ring are not greatly influenced by the concurrent displacement of mass caused by the use of interpolation.

This argument can be carried further to ask if there are interpolation curves superior to simple cubic splines in respect of localizing a unit displacement. One candidate interpolation procedure is the spline-with-tension (Schweikert, (1966)), which is the spline which would be obtained from a thin beam, constrained to go through the interpolation points, but with a tension applied to the ends, rather than allowing them to be free linear extensions beyond the endmost interpolation points. Between interpolation abscissa, x_i and x_{i+1}, the spline-with-tension curve takes the form

$$y = C_{1i} \sinh \sigma(x_{i+1}-x) + C_{2i} \sinh \sigma(x-x_i) + C_{3i}(x_{i+1}-x) + C_{4i}(x-x_i), \tag{A-97}$$

where the coefficients C_{ji} are chosen to produce continuity of the curve and its first two derivatives.

The programs which were used to calculate the splines were validated by running them with a simple main program which provided input and output, and observing that the output passed through the interpolation points, and, in the case of the spline-in-tension, that

the tension parameter behaved as expected. The spline-in-tension program, run with very small tension, duplicated the output of the spline program, thus providing another check. Since the spline program computes polynomials and the spline-in-tension program hyperbolic sines, it is unlikely that agreement would occur unless both programs operated correctly.

By increasing σ, the tension parameter, it is possible to reduce the deviation of the curve from zero between non-displaced points to as small as desired. The limit curve, when $\sigma \to \infty$, is the piecewise linear curve obtained by connecting adjacent interpolation points with straight lines. Figure A2 shows the interpolation curves using splines and splines-in-tension which result from displacing one ring out of twenty and allowing the remaining nineteen to remain at rest. Notice that a tension of 50 makes a large change in the interpolation curve, virtually eliminating any interring displacement more than one ring away from the displaced ring, and reducing the displacement peaks next to the displaced ring to about one-third of that of the spline curve.

To get an estimate of what effect the interpolation scheme has on the spectra, we ran several cases twice, using first splines and then splines-in-tension with $\sigma = 50$. Little difference was apparent, with the eigenvalues differing in both real and imaginary parts by less than 0.001. Following this result, splines alone were used for all computations.

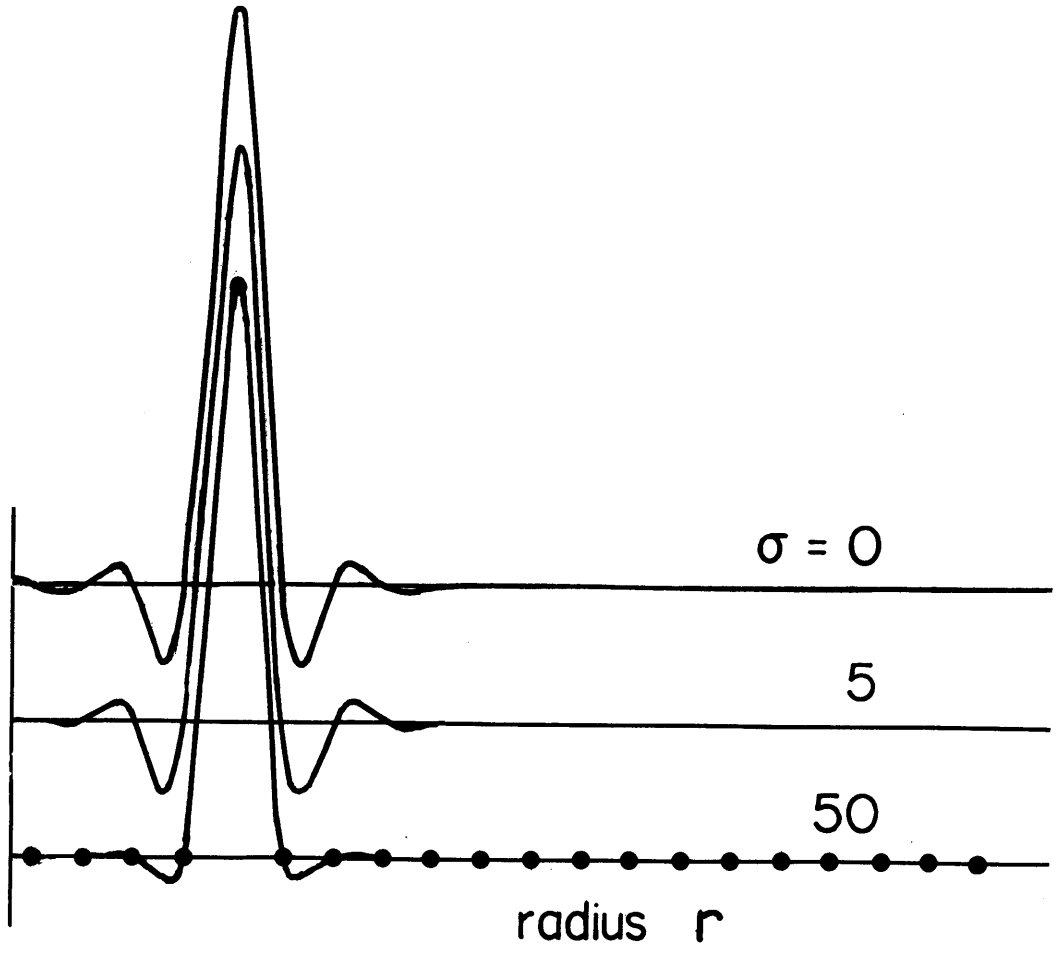

Figure A2. Effect of tension parameter σ on 20-point splines extending out to radius $r = 4$. In each spline, the fifth point is displaced a unit amount, while the remainder are unmoved, just as in the calculational scheme used.

The numerical integration of the force integrals was carried out by using Gauss' method of integration (Fröberg, (1965)). The quadrature formula approximates

$$\int_{-1}^{1} f(x)\,dx \quad \text{by} \quad \sum_{k=1}^{n} A_k\, f(x_k)$$

where weights A_k and abscissas x_k are chosen so that the integration is exact for polynomials up to degree $2n-1$. Rather than go to a high value of n and integrate using one application of Gauss' formula from the center to the outer edge of integration, r_e, we divided the integration interval into n_I panels, where the i th panel covers the subinterval

$$\left(\frac{i-1}{n_I}\, r_e,\ \frac{i}{n_I}\, r_e \right)$$

and applied Gauss' formula with $n_I = 10$ to each panel individually. This allowed an excellent test of integration accuracy to be made by varying n_I, the panel number. The integration routine was also tested on a number of analytically integratable functions, such as $\exp(x)$ and $\sin(x)$, before applying it to the force integrals.

In the majority of calculations n_I was set at about 15. Comparison of the eigenvalue results for $n_I = 15$ and $n_I = 16$ indicated excellent agreement. For the case of $m = 2$ and $a = .4$, the largest relative error among the real frequencies amounted to .025%, which occurred at the smallest real eigenvalue. Larger discrepancies, up to .1%, occurred in the imaginary parts of slightly complex eigenvalues.

A more sensitive parameter than the panel number was detected in the course of the integration tests, however. The outer edge of the integration interval is not fixed by any physical object or phenomena in the model and must be chosen arbitrarily. Choosing it close to the outermost sampling radius will make the model ignore its outer portion, and choosing it far from the outermost sampling ring will give the outermost one or two sampling points more effect than they would have if more points were available to add on beyond the outermost. We empirically determined that the integration edge could be placed anywhere in a wide band beyond the outermost sampling point without making any major change in the eigenfrequencies, and that placing the integration edge beyond that band caused sizable changes in the arch frequencies which had the largest real components.

The case $m = 2$, $a = .4$ was used for experimentation, with the sampling rings equally spaced between the center and $r = 3.6$, the innermost at 0.09 and the outermost at 3.51. Little difference in eigenvalues was found between runs with integration limits of 3.6

and 3.8, amounting to less than one-half percent in the real parts of all eigenvalues and less than one-half percent in the imaginary parts, except for the four outermost unstable eigenvalues of the arch, which have up to ten percent variation in the imaginary parts.

Changing the integration edge to 4.2 causes much larger variation in these outer arch eigenvalues, but again the remaining unstable and all stable eigenvalues are only slightly changed, of the order of one percent or less.

These arch eigenvalues are all characterized by having their maximum displacements near the outer edge of the rings, which accounts for their sensitivity to the integration edge. However, serious alteration of even these eigenvalues only occurs with extreme values of the integration edge, compared to the outermost ring radius. In general, the ratio of these two quantities did not exceed 1.1 in any run, except for the test just mentioned.

The only other functions besides the force integrals which enter into the dynamic matrix are the equilibrium disk frequencies, $\Omega(r)$ and $\kappa(r)$. For the gaussian disk, these can be simply calculated by power series for all radii of interest:

$$\Omega^2(r) = e^{-r^2/2} \sum_{i=1}^{\infty} r^{2i-2} q_i$$

$$\kappa^2(r) = e^{-r^2/2} \sum_{i=1}^{\infty} r^{2i-2} (2(i+1) q_i - q_{i-1}) \qquad \text{(A-98)}$$

where

$$q_0 = 0$$

$$q_i = \sum_{j=1}^{\infty} p_j \prod_{k=1}^{i-1} \frac{(2k - 1) - (j - 1)}{4k\ (k + 1)}$$

and

$$p_j = \sqrt{\frac{\pi}{8}}\, a^{j-1} \prod_{k=1}^{j-2} \frac{(k + 2)}{(k + 1)\ k} \qquad j \text{ odd}$$

$$= -\, a^{j-1} \prod_{k=1}^{j-2} \frac{(k + 2)}{(k + 1)\ k} \qquad j \text{ even} \quad .$$

Since they are power series which converge terms may be calculated until a required tolerance is reached. Generally, 10^{-6} was demanded.

Having disposed of the possible causes of computational error in the calculation of the dynamical matrix, we consider errors which arise from the calculation of its eigenvalues. The procedure used for this was the QR method, recommended by Wilkinson (1965) as the most effective of known methods. To ensure against programming errors, two implementations of this method were used and checked against each other. The shorter of the two was the pair of programs HSBG and ATEIG

from the Scientific Software Package of IBM (1968). These two transform the general matrix into a Hessenberg (almost triangular) matrix (HSBG) and then compute the eigenvalues of the almost triangular matrix (ATEIG).

The same steps are included in the algorithm of Grad and Brebner (1968), but their routine is more careful, involving a prescaling operation on the rows and columns. Both programs were checked independently, using randomly generated matrices and then were also checked against one another. The less complicated program of IBM provided sufficient accuracy when run on actual dynamic matrices, judged by the discrepancy between the two.

An indication of the accuracy of the eigenvalue program was gained by computing the determinant of

$$M - (\omega_i + \varepsilon) I$$

where M is the dynamic matrix of equation (55), and ω_i is an eigenvalue. These determinants were calculated by the method of Chió (see Kunz (1957)), which is designed to reduce roundoff error in the value of a determinant by minimizing the number of multiplications required. The resulting determinants showed that the calculated value was within 10^{-8} of that the value which would cause the determinant to be zero. The computer used carried 20 digits in double precision arithmetic, which was used on the more critical portions of the eigenvalue routines.

Further evidence of the accuracy of the eigenvalue methods, as used on dynamical matrices, is offered by the fact that the eigenvalues change only slightly when the panel number is changed, which induces small perturbations in the force integrals. Eigenvalue algorithms computing grossly false results would be expected to be unstable to perturbations, as the error would arise from a small divisor somewhere in the process.

The last major calculation associated with the model is the eigenfunction calculation. The Grad and Brehner algorithm has an eigenfunction subroutine associated with it, and for comparison a separate program was written to accompany the IBM pair of subroutines. Each worked adequately on randomly generated matrices.

The eigenfunction routines were also tested by evaluating

$$A \vec{v}_i$$

where $\vec{v}_i$ is the eigenfunction, and comparing the result with $\omega_i \vec{v}_i$. Again the two agreed to several digits accuracy in all cases tested.

Having eliminated the sources of computational error, any problems remaining with the routines must fall into two categories: programming errors in the basic model, for example computing the wrong density, or model deficiencies, such as too small a number of samples.

If such an error occurred in the axisymmetric programs, disagreement would surely have arisen between the model results and the SWLT theory predictions. Since the theory does not provide many corroborations with the nonaxisymmetric calculations, this check is not valid for these programs. However, to guard as much as possible against such errors, a high degree of commonality was maintained between the axisymmetric and nonaxisymmetric routines. The differences between the two were so thoroughly checked that such an error is extremely unlikely.

REFERENCES

Abramowitz, M., and Stegun, I. 1964, Handbook of Mathematical Functions (Washington: U.S. Government Printing Office).

Ahlberg, J.H., Nilson, E.N., and Walsh, J.L. 1967, The Theory of Splines and their Applications (New York: Academic Press).

Alexander, S. 1852, Astron. J., 2, p.95.

Chamberlin, T.C. 1901, Astrophys. J., 14, p.17.

Erdélyi, A. et al 1953, Higher Trancendental Functions, vol. 1 (New York: McGraw-Hill).

Fröberg, C.-E. 1965, Introduction to Numerical Analysis (Reading, Mass.: Addison-Wesley).

Grad. J., and Brebner, M.A. 1968, Comm. Assoc. Comp. Mach., 11, p.820.

Gradshteyn, I.S., and Ryzhik, I.M. 1965, Table of Integrals, Series and Products (New York: Academic Press).

Hildebrand, F.B. 1952, Methods of Applied Mathematics, (New York: Prentice-Hall).

Hockney, R.W., and Hohl, F. 1969, Astron. J., 74, p.1102.

Hohl, F. 1971, Astrophys. J., 168, p.343.

Hunter, C. 1963, Monthly Notices Roy. Astron. Soc., 126, p.299.

———·· 1965, Monthly Notices Roy. Astron. Soc., 129, p.321.

———·· 1969, Studies Appl. Math., 49, p.55.

———·· 1972, in Annual Review of Fluid Mechanics (Palo Alto: Annual Reviews) p.219.

Hunter, C., and Toomre, A. 1969, Astrophys. J., 155, p.747.

IBM 1968, System/360 SSP Programmer's Manual (White Plains: IBM).

Jeans, J.H. 1929, Astronomy and Cosmogony (Cambridge: Cambridge Univ. Press).

Kalnajs, A.J. 1965, unpublished Ph.D. thesis, Harvard University

———————, 1970, in Proc. I.A.U. Symp. No. 38 (New York: Springer-Verlag) p.318.

1971, Astrophys. J., 166, p.275.

———————, 1972, Astrophys. J., 175, p.63 .

Kunz, K.S. 1957, Numerical Analysis (New York: McGraw-Hill).

Lin, C.C. 1967, in Proc. I.A.U. Symp. No. 31 (New York: Springer-Verlag) p.313.

————————, 1970, in Proc. I.A.U. Symp. No. 38 (New York: Springer-Verlag) p.377.
1971, in Highlights of Astronomy, vol. 2 (Dordrecht, Holland: D. Reidel) p.88.

Lin, C.C. and Shu, F.J. 1964, Ap. J., 140, p.646.

————————————————, 1966, Proc. N.A.S., 55, p.229.

Lin, C.C., Yuan, C., and Shu, F.J. 1969, Ap. J., 155, p.721.

Lindblad, B. 1927, Arkiv Mat. Astron. Fysik, 20A.

———————, 1948, Handbuch der Physik, p.214.

———————, 1957, Stockholms Obs. Ann., 20, no. 6.

———————, 1961, Stockholms Obs. Ann., 21, no. 8.

———————, 1963, Stockholms Obs. Ann., 22, no. 5.

Lynden-Bell, D., and Kalnajs, A.J. 1972, Monthly Notices Roy. Astron. Soc., 157, p.1.

Mathewson, D.S., van der Kruit, P.C., and Brouw, W.N. 1972, Astron. and Astrophys., 17, p.468.

Mikhlin, S.G. 1957, Integral Equations and their Applications to Certain Problems in Mechanics, Mathematical Physics and Technology (New York: Pergamon).

Miller, R.H. 1971, Astrophys. and Space Sci., 14, p.73.

Miller, R.H. and Prendergast, K.H. 1968, Astrophys. J., 151, p.699.

Miller, R.H., Prendergast, K.H., and Quirk, W.J. 1970, Astrophys. J., 161, p.903.

Miyamoto, M. 1969, Pub. Astron. Soc. Japan, 21, p.319.

Rehm, R.G. 1965, unpublished Ph.D. thesis, M.I.T.

Schweikert, D.G. 1966, J. Math. Phys., 45, p.312.

Shu, F.J. 1968, unpublished Ph.D. thesis, Harvard Univ.

_____. 1970, Astrophys. J., 160, p.89.

_____. 1970, Astrophys. J., 160, p.99.

Shu, F.J., Stachnik, R.V., and Yost, J.C. 1971, Astrophys. J., 169, p.465.

Spitzer, L. 1962, Physics of Fully Ionized Gases (New York: Interscience).

Stix, T.H. 1962, The Theory of Plasma Waves (New York: McGraw-Hill).

Toomre, A. 1963, Astrophys. J., 138, p.385.

_____. 1964, Astrophys. J., 139, p.1217.

_____. 1969, Astrophys. J., 158, p. 899.

Wilkinson, J.H. 1965, The Algebraic Eigenvalue Problem (Oxford: Clarendon).