Ann. Rev. Astron. Astrophys. 1977. 15: 437–78

THEORIES OF SPIRAL STRUCTURE

×2119

Alar Toomre
Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

> Much as the discovery of these strange forms may be calculated to excite our curiosity, and to awaken an intense desire to learn something of the laws which give order to these wonderful systems, as yet, I think, we have no fair ground even for plausible conjecture.
>
> Lord Rosse (1850)
>
> A beginning has been made by Jeans and other mathematicians on the dynamical problems involved in the structure of the spirals.
>
> Curtis (1919)
>
> Incidentally, if you are looking for a good problem ...
>
> Feynman (1963)

1 INTRODUCTION

The old puzzle of the spiral arms of galaxies continues to taunt theorists. The more they manage to unravel it, the more obstinate seems the remaining dynamics. Right now, this sense of frustration seems greatest in just that part of the subject which advanced most impressively during the past decade—the idea of Lindblad and Lin that the grand bisymmetric spiral patterns, as in M51 and M81, are basically compression waves felt most intensely by the gas in the disks of those galaxies. Recent observations leave little doubt that such spiral "density waves" exist and indeed are fairly common, but no one still seems to know why.

To confound matters, not even the N-body experiments conducted on several large computers since the late 1960s have yet yielded any decently *long-lived* regular spirals. By contrast, quite a few such model disks have developed bar-like or oval structures that have endured almost indefinitely in their interiors. This finding is undoubtedly a major step toward understanding various bar phenomena that complicate the observed spirals, but it also reminds us rudely just how little we really comprehend: To avoid the rapid nonaxisymmetric instabilities associated with this bar-making, the experimental "stars" require (or otherwise they acquire) random velocities proportionately much larger than those observed among stars near the

Sun, or than those suggested by edge-on views of other galaxies. It is this dilemma, of course, that has fueled much of the recent speculation about massive but largely unseen halos—though the alternative that some very hot inner disks or spheroids might already cure those instabilities seems not yet to have been excluded firmly. Whatever that answer may be, it is clear that the spiral dynamics gets no easier when even the most rapid but least visible members of a galaxy need to be suspected of helping distort the axial symmetry.

Still, one must begin somewhere. Happily, this remains a subject where it makes sense to start almost at the beginning.

2 SOME OF LINDBLAD'S IDEAS

As Oort (1966) remarked in a graceful obituary, Bertil Lindblad's "long series of papers, beginning already in Uppsala in 1927, and continued until his death, bears witness to the struggle with the spiral problem. It is natural that in this field, on which at that time nothing was ripe for harvesting, he did not immediately find the right path." Oort was referring mostly, of course, to Lindblad's long obsession with *leading* spiral arms (cf his 1948 Darwin lecture). Lindblad believed such arms to consist of material shed recently from unstable circular orbits just outside a spinning central nebula that was shrinking and flattening with time in the manner envisaged by Jeans (1929) for the entire Hubble sequence. In that setting, unlike Chamberlin (1901) or Jeans himself with tidal forces from outside, Lindblad sought an *intrinsic* reason for the frequent two-armed symmetry of spirals. Already in his earliest papers he thought he had found one in the well-known dynamical instability of the Maclaurin spheroids, which generates a growing triaxial shape that rotates in space, in a wavelike fashion, only half as fast as the fluid itself. Lindblad expected the spillage to occur mostly from the two ends of the longer equatorial axis, which he felt would in reality rotate even more slowly.

Such was the origin of Lindblad's lifelong interest in waves in galaxies. During the many years that he continued to believe in the transient leading arms, this interest developed mostly into a fascination with bars. He often wrote of the latter in terms like "vibrations" or "modes" or "density waves"—although the last phrase for him usually meant only motions like the rotation of a lumpy rigid body or of a Jacobi ellipsoid, hardly true waves at all. In retrospect, none of Lindblad's extensive labors to calculate such bar modes seems worth recounting here. Still quite remarkable, however, are his repeated guesses such as "the most promising approach to a solution of the problem of spiral structure appears to be by way of the barred spiral nebulae," or that "in the ordinary spirals . . . a plainly developed barred structure may not appear, but there is likely to be a characteristic density wave $s = 2$, which causes a departure from rotational symmetry and which incites the formation of spiral structure by its disturbing action on the internal motions in the system" (Lindblad 1951).

Much of this early work obviously dates from an era before 21-cm astronomy and even Baade's explicit recognition of the two stellar populations. Yet that alone does not explain why, as Oort put it, the attempts at spiral-making via gravity

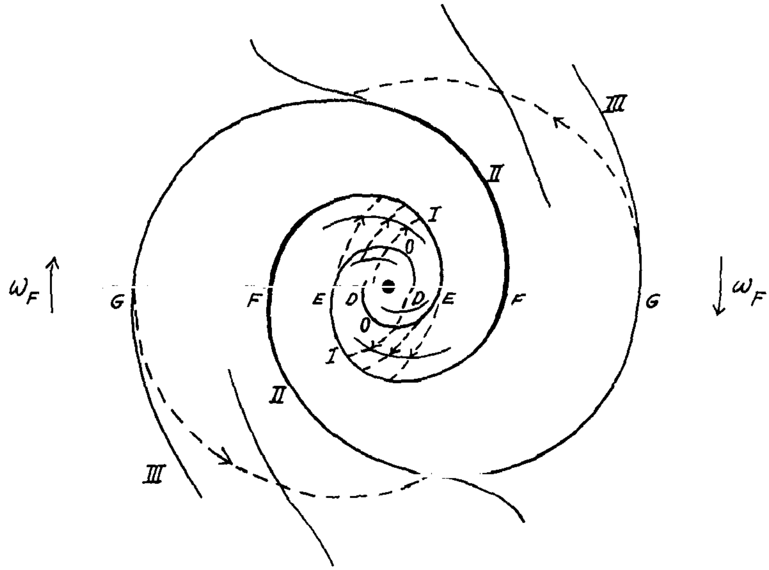

Figure 1 Lindblad's (1964) sketch of his "circulation theory." (Reprinted with permission from *Astrophys. Norv.*)

"remained for 35 years practically a monopoly of the Stockholm Observatory." A subtler reason was that by the 1950s it had become a little too obvious, to Baade (1963, p. 67) and many others, not only "that the primary phenomenon in the spiral structure is the dust and gas," but also "that we could forget about the vain attempts at explaining spiral structure by particle dynamics. It must be understood in terms of gas dynamics and magnetic fields." Amidst such a tide of facts and opinion, Lindblad stood almost alone in not renouncing his hopes for basically gravitational explanations of spiral structure—though he did finally concede, in a review article (Lindblad 1962) still worth reading, that his old leading-arm models are "not reconcilable with modern evidence." Rather more reconcilable, it seems now, are two further notions that occurred to Lindblad during his last decade.

One of these, stemming from the mid-1950s, was the idea of "dispersion orbits," which we return to in the next section. The other notion is pictured in Figure 1, taken from the second of two papers in which Lindblad (1963, 1964) speculated explicitly "on the possibility of a quasi-stationary spiral structure."

To give "the typical spiral structure . . . a certain steadiness in time" despite the differential rotation, Lindblad proposed that the two gaseous arms in Figure 1 are simply a *pattern* that rotates rigidly in space with a clockwise angular speed ω_F. Outward of the two (corotation) points F, where the orbital speed equals ω_F, gas elements move in the retrograde direction, as viewed in the rotating frame; inside F they tend to overtake the pattern. Nevertheless, Lindblad argued, the pattern persists. The self-gravity of the upper arm II, for instance, ensures that any interstellar matter that joins it at, say, the 12 o'clock position will drift slowly outward in radius *along the arm* as its differential rotation carries it a quarter-turn to the leftward point G, where at last it will part company with arm II/III. Then, "thanks to the attraction from [the other arm] II, [such material] will return towards that arc in a shower of orbits, of which one is indicated by the dotted curves." Inside the corotation radius, Lindblad envisaged analogously that the gas would drift inward while sliding clockwise along either of the spiral arcs I. Subsequently it would tend "to be assimilated in an outward motion" toward the next arm, to complete the cycle. Also drawn in Figure 1 are some inclined interarm "branches," which Lindblad suspected to be characteristic of the material in transit.

With this simple sketch Lindblad displayed an intuitive grasp of the gas trajectories involved in spiral shock waves that was not to be surpassed markedly until the work of Fujimoto (1968) and Roberts (1969). Nor was his reasoning here entirely qualitative: If only in a rough and laborious manner, Lindblad reckoned that enough extra force to maintain arms of 18° pitch angle could be provided by an arm mass totalling one-tenth of the entire mass of a galaxy. In short, this "circulation theory" was a big step forward from anything that preceded it, including even Oort's (1956) suggestion that "perhaps one side of an arm collects material while the other side evaporates it, so that the arm holds a constant position."

As it happened, apart from those gas orbits, this hypothesis of quasi-stationary spiral structure was in turn outdistanced soon by the QSSS hypothesis of Lin & Shu (1964, 1966) deliberately bearing the same name. Already these two papers furnished much more detail and clarity in support of their own conjecture that somehow "the matter in the galaxy can maintain a [spiral] density wave through gravitational interaction in the presence of the differential rotation," and that "this density wave provides a spiral gravitational field which *underlies* the observable concentration of young stars and the gas." Lin and Shu took the big extra step of including—and from their 1966 paper onwards, of also calculating far more plausibly—the gravitational forces to be expected from tightly wrapped collective waves among ordinary disk stars. Contrary to what is often supposed, however, these major improvements made or inspired by Lin had remarkably little to do with any "density waves" ever considered by Lindblad. Instead, they relate much more to that other idea which Lindblad termed "dispersion orbits." It is just those orbits, as we will now see following mostly Kalnajs (1973), that can be turned (literally) into a quick introduction to the basic mechanics still envisaged by Lin (e.g. 1971, 1975) and Shu (1973) for the bisymmetric spiral waves.

3 NEARLY KINEMATIC WAVES

The classic paper in which Maxwell (1859) explored the stability of Saturn's rings focused especially on the topic of infinitesimal disturbances to a configuration of N equal small masses rotating in centrifugal equilibrium at the vertices of an N-sided regular polygon, around a heavy mass imagined *fixed* at the center. Maxwell separated this problem into $N/2$ simpler problems, each involving a sinusoidal [$\sin(m\theta)$ or $\cos(m\theta)$] dependence of the radial and tangential displacements of the various ring particles upon their longitudes θ. For each integer wavenumber $0 < m < N/2$, he demonstrated that such a polygon can exhibit four distinct modes of distortion confined to its own plane—and that all four are pure waves travelling around the ring, provided those satellite masses are sufficiently small.

Lindblad (1958, 1961) extended this analysis, with some minor extra approximations, to the $m = 1$ and $m = 2$ perturbations of an essentially continuous ring of particles rotating with angular speed Ω in the fixed, non-Keplerian force field of an idealized galaxy. He likewise obtained four basic modes for each m, all of which are indeed wavelike when the ring mass is almost nil. Two of these modes, the ones he again called "density waves," then reduce to nearly frozen sinusoidal disturbances to the ring density, rotating in space with almost exactly the material speed Ω. As such they need not concern us further here (though the reader may wish to emulate Maxwell by resolving the little paradox of why the obvious tangential attractions from such orbiting mass concentrations are not immediately destabilizing). Much more relevant is the slower of the two remaining modes, or "waves of deformation": As Figure 2 illustrates for wavenumber $m = 2$, the limiting zero-mass version of that mode rotates in space with angular speed $\Omega - \kappa/m$, where κ is the so-called epicyclic frequency, or the angular rate of radial oscillation of each individual particle in the given central force field.

It is essentially this slowly advancing kinematic wave (rather than its fast partner with speed $\Omega + \kappa/m$), composed of many separate but judiciously-phased orbiting test particles, that Lindblad meant by his term "dispersion orbit." That name itself now seems only a quaint relic of his hope from the mid-1950s that the gas layer

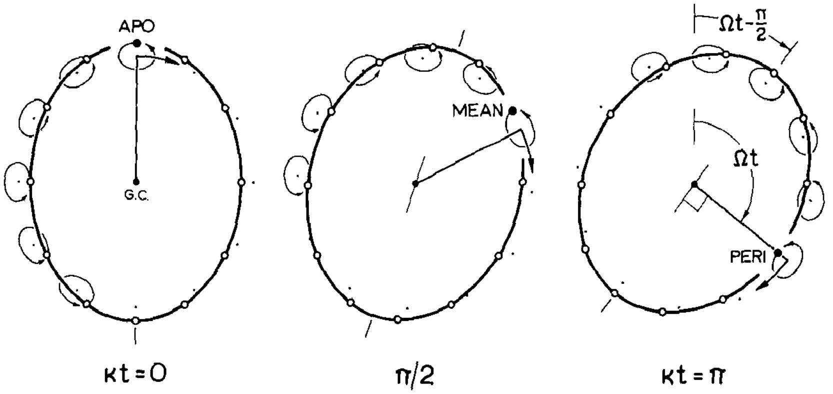

Figure 2 Slow $m = 2$ kinematic wave on a ring of test particles, all revolving clockwise (like the 12 shown) with mean angular speed Ω in strictly similar and nearly circular orbits. The small elliptical "epicycles," traversed counterclockwise in the above sequence of snapshots separated in time by exactly one-quarter of the period $2\pi/\kappa$ of radial travel along each orbit, depict the apparent motions of these particles relative to their mean orbital positions or "guiding centers." Drawn for the case $\kappa = \sqrt{2}\Omega$—or one where the rotation speed $V(r) = r\Omega(r) = \text{const}$ at neighboring radii—the diagram emphasizes that the oval *locus* of such independent orbiters advances in longitude considerably more slowly than the particles themselves. That precession rate equals $\Omega - \kappa/2$, as one can verify at once by comparing the last frame with the first.

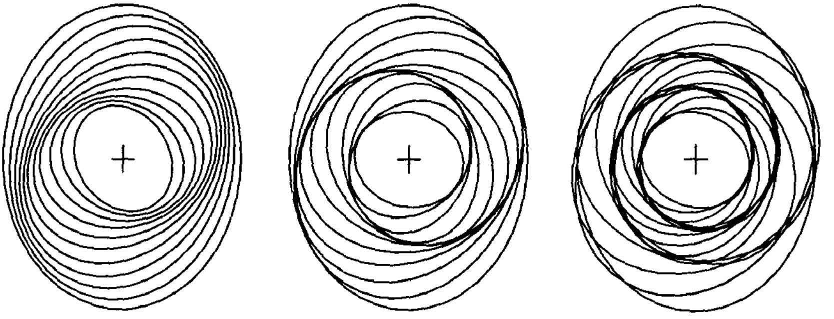

Figure 3 Three kinematic density waves of spiral form (Kalnajs 1973), obtained by merely superposing some kinematic ring waves such as shown in Figure 2.

in a galaxy might spontaneously tend to aggregate or "disperse" into just a few such distorted rings (an idea that might still apply somewhat to the shearing away of star associations born into eccentric orbits). But not at all quaint is the main reason for Lindblad's interest in this dispersal: He had just noticed that, whereas the circular speed $\Omega(r)$ varies markedly with radius r in a typical galaxy, the wave speed $\Omega(r)-\kappa(r)/m$, *in the bisymmetric case* $m=2$ *only*, remains fairly constant over a considerable range of intermediate radii. This fact greatly intrigued Lindblad—who did not need to be told that strict constancy would banish wrapping-up worries or that the nicest spirals tend to have two arms. Yet astonishingly, that is about as far as he ever got. Per Olof Lindblad agrees that somehow it never occured very explicitly to his father—despite writing in 1958 that "it is likely that the spiral arms are largely osculating to dispersion orbits"—to combine already those "orbits" into any long-lived *spiral* patterns such as he himself later proposed in Figure 1!

This logical gap between Lindblad's orbits and the density waves of Lin and Shu was closed emphatically only by Kalnajs (1973), who presented the three striking geometric examples reproduced in Figure 3. Each of these pictures consists of just 11 ellipses shifted systematically in longitude according to the rule $2\theta = \alpha \ln$ (major axis). None of them pretends to contain any actual dynamics, nor any trace of the helpful extra densities expected near all the major axes already from Figure 2. Kalnajs devised these three diagrams simply to dramatize how easy it is, in retrospect, to concoct trailing or leading spiral loci of modest or even very pronounced crowding from nothing more than *ad hoc* superpositions of individual finite-amplitude ring waves resembling Lindblad's $m=2$ dispersion orbits. Such spiral patterns are inherently wavelike and thus, quite apart from any deeper questions of origin or the preference for one shape over another, it is clear that these beautiful patterns would last forever *if* all the separate ovals could somehow be persuaded to precess not just with the small and nearly equal speeds $\Omega-\kappa/2$ but (*a*) at exactly the same rate and (*b*) without change of amplitude. The big task is to do all this persuading.

4 LIN-SHU DENSITY WAVES

When Lin and Shu almost succeeded in providing such assurances with their WKBJ estimates a decade ago, they concentrated largely—and understandably—upon condition (*a*), that of avoiding all wave shear. Even more understandably, they focused upon tightly wrapped waves, as a major analytical shortcut. Unlike Lindblad with his few or single rings, Lin and Shu reasoned that the disturbance gravity forces should be predominantly *radial* in any real two-armed spiral that is wound about as tightly as the one shown in the last frame of Figure 3. And these forces, they reckoned, ought to affect the collective oscillations of the disk stars and gas in any given orbiting neighborhood very nearly as much as they would there have affected a strictly axisymmetric density wave of the same (large) radial wavenumber.

4.1 *Axisymmetric Vibrations*

Even today, this reviewer knows of no analytical theory for interactive radial vibrations with amplitudes approaching those in Figure 3. However, the theory for short axisymmetric waves of *infinitesimal* amplitude goes back at least to Safronov (1960). Among other things, Safronov recognized almost explicitly that the self-gravity of any thin, cold disk of surface mass density μ tends to *reduce* the local frequency ω of short axisymmetric waves of wavenumber k from the epicyclic value κ to that implied by

$$\omega^2 = \kappa^2(r) - 2\pi G\mu(r)|k|. \tag{1}$$

In providing this formula, Toomre (1964) in essence only repaired a coefficient error made by Safronov. Thereupon, he very nearly contributed one of his own: To avoid the severe Jeans instabilities implied by Equation (1) for all wavelengths $\lambda = 2\pi/k$ shorter than

$$\lambda_{\rm crit}(r) = 4\pi^2 G\mu(r)/\kappa^2(r), \tag{2}$$

the root-mean-square radial velocity σ_u of any thin disk consisting only of stars with a Schwarzschild distribution of random velocities must exceed

$$\sigma_{u,\rm min}(r) = 3.358 G\mu(r)/\kappa(r), \tag{3}$$

as Kalnajs kindly corrected this writer (by about 30%) before publication. It then emerged, ironically, that Kalnajs in 1961 had omitted a factor 2π from an analogous estimate—and this had dampened his own interest in such collective effects!

Kalnajs (1965, p. 50) redeemed himself with a flawless first derivation of the (implicit) mathematical relationship between the local wavenumber k and frequency ω for an axisymmetric wave involving stars with an arbitrary rms speed $\sigma_u = Q\sigma_{u,\rm min}$, rather than just $\sigma_u = 0$ as in the simple Equation (1). Lin & Shu (1966) soon rederived the same as the crucial $m = 0$ subcase of their own—far more widely known—dispersion relation. In its original integral form, their version looks rather different

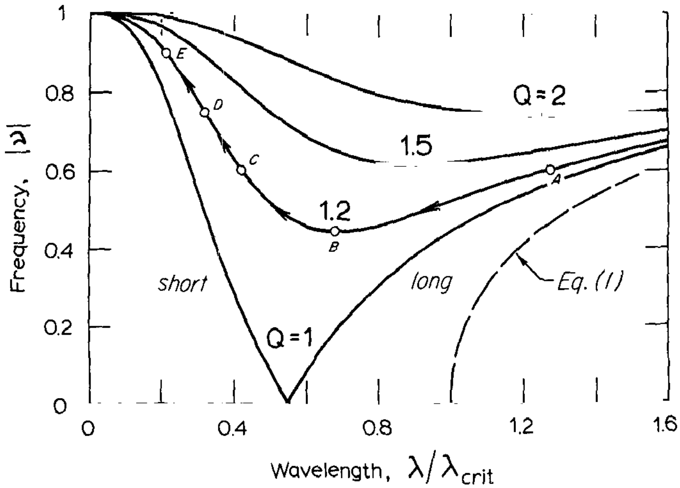

Figure 4 The Lin-Shu-Kalnajs dispersion relation for axisymmetric density waves of local frequency $\omega = \nu\kappa$ and modest radial wavelength λ in a thin, rotating disk of stars endowed with Q times the minimum random motions required by Equation (3).

from the series expression by Kalnajs [cf Toomre 1969, Equation (11)], but both imply the identical curves summarized in Figure 4 for several values of the stability parameter $Q \geqq 1$ at which the Jeans instabilities are no longer possible.

The key dynamical fact in Figure 4 is that the local self-attraction can again only reduce the frequency ω of the small-amplitude *collective* vibrations to some fraction $|\nu|$ of the epicyclic frequency κ, which itself continues to characterize the finite-amplitude radial and tangential excursions of the *individual* stars whose net density changes comprise that wave. This lowering of the wave frequency, as all three authors recognized, is greatest at intermediate wavelengths and tends to zero at both extremes. The important return of ω to κ as $\lambda \to 0$ occurs because, although all stars obviously sense the sinusoidal force field, only those whose maximum radial speeds U are less than about κ/k manage to contribute to the wave density a net amount greater than 60% of what they would have contributed if they had merely orbited in concentric circles before the wave came along. That ratio of cooperativeness indeed reaches zero already at $U = 1.841\ \kappa/k$ for $|\nu| = 1$, and at $2.405\ \kappa/k$ for $\nu = 0$; moreover, from stars with $U = \pi\kappa/k$—or ones whose apocenter and pericenter distances differ by exactly one wavelength—one even expects a bit of cussedness amounting roughly to a *negative* one-tenth contribution to the wave density (Kalnajs 1964, private communication).

4.2 *Matching of Precession Rates*

Together with various kin involving gas and/or thickness corrections, the axisymmetric LSK dispersion relation from Figure 4 has played a fundamental role in nearly all of the "asymptotic" explorations of tightly wrapped (i.e. almost axisymmetric) $m = 2$ spiral waves that Lin, Shu and several collaborators have conducted during the past decade. That role has been remarkably direct: As Lin & Shu (1964) first illustrated using the much less plausible Equation (1), the above reduction of that essential frequency from κ to $|\nu|\kappa$ implies an analogous change, from $\Omega \pm \kappa/2$ to $\Omega \pm |\nu|\kappa/2$, in the speeds of precession of the *net* density contributions expected from the many separate stars belonging to any given mean orbital radius but gyrating around various epicycles. And it was precisely in the seeming nuisance that ν depends upon the local radial wavelength λ (or upon the local pitch angle of the trailing or leading spiral) that Lin and Shu, already in 1964, spotted the grand opportunity of simply *adjusting* those more realistic precession rates to the strict constancy with radius that had eluded Lindblad.

Among much else, it needed to be demonstrated, of course, that this desired "tuning" of $\Omega \pm |\nu|\kappa/2$ could actually be accomplished simultaneously, for any single pattern speed Ω_p, over long enough stretches of radii and with plausible spiral arm spacings. That this is indeed possible without running badly afoul of observations was first shown for the Galaxy by Lin & Shu (1967; see also Lin et al. 1969) and for M33, M51, and M81 by Shu, Stachnik & Yost (1971; see also Tully, 1974, Roberts et al. 1975, Rots 1975, Rogstad et al. 1976).

The essential points of those feasibility studies may be gleaned from Figure 5b. The five curves in that diagram describe, for $N = 2$, 3, 4, 5, 6, the usual signed fractions

$$\nu(r; \Omega_p) = 2[\Omega_p - \Omega(r)]/\kappa(r) \equiv \nu_N(r) \tag{4}$$

of the local epicyclic frequency whereby an intended $m = 2$ wave of hypothetical pattern speed $\Omega_p = \Omega(Nr_0)$ would appear "Doppler shifted" to an observer orbiting with speed $\Omega(r)$ in one rather typical model galaxy. That model is a thin disk of surface mass density $\mu(r) = \mu_0 \exp(-r/r_0)$ analogous to the simple law that Freeman (1970a) among others has stressed to be characteristic of many observed *light* distributions. Its dimensionless rotation curve $\tilde{V}(r)$ in Figure 5a, based on a formula given by Freeman, mainly just corroborates that the Sun belongs somewhere near $r = 3r_0$ in this crude facsimile of our Galaxy.

Given that $|\nu| < 1$, Figure 5b implies first that only pattern speeds from an intermediate range, roughly between $\Omega(3r_0)$ and $\Omega(5r_0)$, make it even plausible to contemplate any single $m = 2$ spiral structure extending from the deep interior well beyond our make-believe Sun. The higher speed limit is here set by the unwillingness of the self-gravitating disk to carry waves that precess even more rapidly than the fast kinematic waves corresponding to $\nu = +1$, whereas the lower limit arises from the similar expectation that the collective effects near each radius can only raise the slow wave speeds above Lindblad's value $\Omega - \kappa/2$. As usual, the radii at which a prescribed Ω_p exactly equals $\Omega \pm \kappa/2$ are here referred to as the outer and inner Lindblad resonances.

A second type of basic restriction is represented by the shaded $|\nu| < 0.44$ strip in Figure 5b. It warns that already when the stability index Q exceeds unity by as little as 20%, there exists a sizable forbidden annulus around each proposed corotation radius. Lin-Shu waves of obviously low apparent frequency might still "tunnel" across any such region (cf Toomre 1969, Figure 3), but they will there not find any *local* hospitality in the WKBJ sense. Clearly that gap only widens for $Q = 1.5$ and 2.0, since Figure 4 then forbids $|\nu| < 0.61$ and 0.74, respectively.

This second restriction has been much underplayed in all of the wave fittings cited above, which have in essence just postulated that $Q = 1$, upon including those gas corrections, thickness effects, etc. Observations of nearby disk stars suggest very roughly that $1.2 \lesssim Q \lesssim 2$ (cf Toomre 1974a); elsewhere Q remains totally unmeasured and well-nigh unmeasurable. On the theoretical side, quite apart from any global instabilities, one interesting difficulty with presuming that $Q \cong 1$ in the gas-rich parts of a galactic disk is that the effective masses of any large chunks of interstellar matter—or, loosely speaking, the various bits and pieces of arms present in even the grandest spirals—should then be strongly enhanced by cooperative density "wakes" from passing stars (cf Figure 13). The result, as Julian (1967) perhaps overcautioned, can be a growth of the stellar random motions (or Q itself) at a rate well in excess of any estimated by Spitzer & Schwarzschild (1953; see also Woolley & Candy 1968) from encounters with "cloud complexes." If only for the last reason, it seems prudent to concentrate mostly on those features of Figure 5b that would survive even if $Q = 1.5$, or at the very least if $Q = 1.2$.

Contrary to what Lindblad (1964) surmised in Figure 1, such features evidently include next to nothing from beyond any of the corotation radii $r = Nr_0$, since the upper strip $0.44 < \nu < 1$ in Figure 5b indicates for each N that one can hope to tamper with the fast kinematic waves only over very limited stretches of radii, of order $r_0/2$, when $Q \geqq 1.2$. This nominal extent seems absurdly small, compared even with the "short" waves of lengths λ of order $2r_0$ implied there by Figures 4 and 5a. What is more, even when $Q = 1$, the ranges of radii corresponding to $0 < \nu < 1$ remain modest also in models such as shown in Figures 2 and 4 of Shu, Stachnik & Yost (1971)—who themselves proceeded to ignore those modified fast waves on the more enigmatic grounds that "if Lin (1970) is correct, only the region where $\Omega - \kappa/2 < \Omega_p < \Omega$ shows an organized spiral pattern." In a sequel, Roberts, Roberts & Shu (1975) likewise took no active interest in that short outer segment, now apparently because "the theoretical behavior of spiral density waves near corotation is more complicated, even in the linear regime, than has been previously assumed." One way or another, this reviewer agrees that the $\nu > 0$ waves do not seem destined for much glory.

This leaves only the bottom half of Figure 5b—or more exactly, only its lowest quarter or so—if we are to continue this exercise in pessimism with $Q = 1.2$ and 1.5. Down there at last we meet one very reassuring fact in the sizable range of radii over which especially the ν_5 curve still manages to remain within even the narrow $Q = 1.5$ corridor $-1 < \nu < -0.61$. Of course, this welcome news leans heavily upon the familiar near-constancy of $\Omega - \kappa/2$. After all, not much help is needed anyway from the self-gravity when (*a*) those slow kinematic speeds are as nearly uniform as Lindblad noted them to be near the "turnover" radii of most rotation curves, and provided also that (*b*) one contemplates pattern speeds like $\Omega(5r_0)$ that exceed those elementary speeds only slightly. In such circumstances, the comparatively modest help available from the LSK dispersion relation when $Q = 1.2$ or 1.5 can still be worth a great deal.

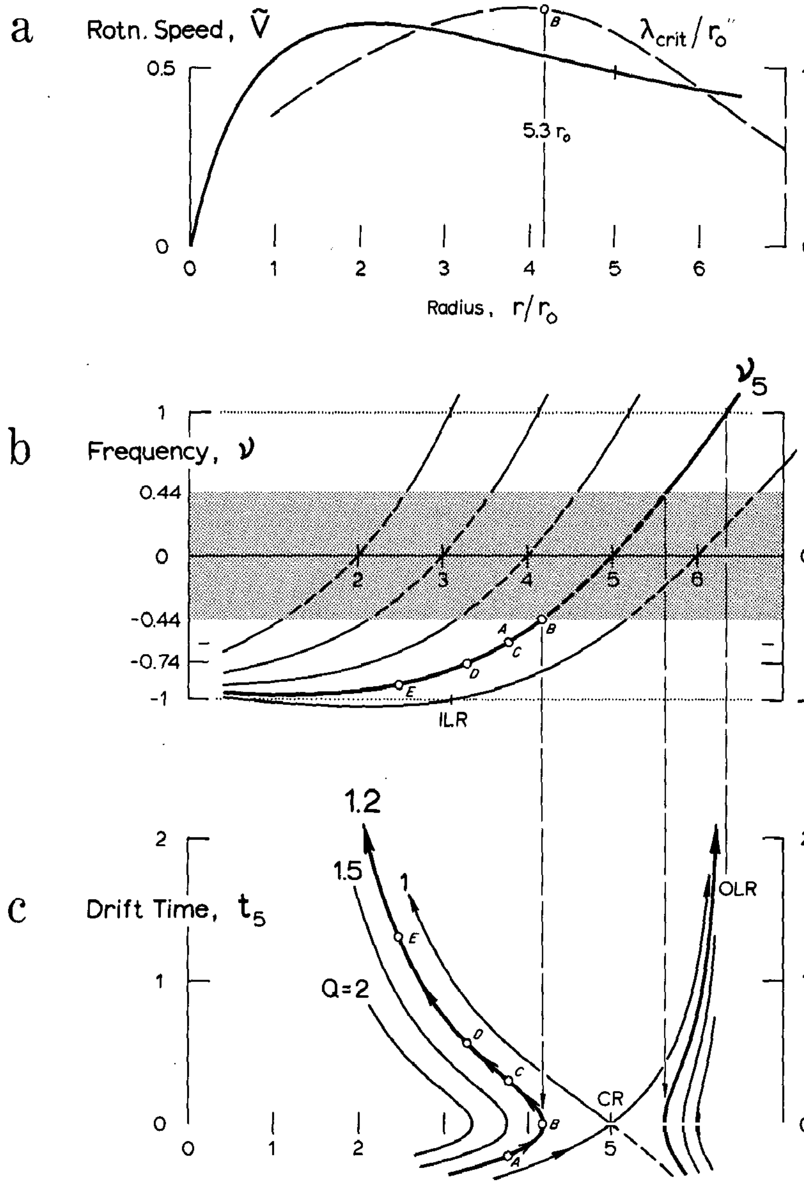

Figure 5 Theoretical data for a thin exponential disk of scale length r_0: (*a*) Rotation speed $\tilde{V}$ and stability length λ_{crit}, (*b*) dimensionless local frequencies $\nu_2, \ldots, \nu_6$, (*c*) group velocity drift times $t_5(r; Q)$, all plotted as functions of r. The shaded frequencies are locally unattainable when $Q = 1.2$. CR means corotation; ILR and OLR are some inner and outer Lindblad resonances. The points marked A through E appear also in Figure 4.

In short, at least when $Q \geqq 1.2$, the only reasonable overall wave speeds Ω_p for two-armed spiral structures as *extensive* as those seen in M51 and M81—and less obviously, perhaps also in our Galaxy—seem to be those that are slow enough to place corotation on, but not beyond, the outskirts of those systems. [This conclusion, by the way, appears to be completely independent of any "concept of the initiation of density waves by the Jeans instability in the outer regions of normal spiral galaxies (Lin 1970)" that Shu, Stachnik & Yost (1971) felt they were testing.] Almost as a corollary, however, the exact radius of inner Lindblad resonance (ILR), if indeed any, seems much more jittery. It must also be quite sensitive to modifications of the rotation curve itself in the interior.

As for other angular wavenumbers, notice that neither such ILR sensitivity nor any extensive near-resonance with $\Omega - \kappa/m$ could have arisen for $m > 2$, since a mere rescaling of the ν axis in Figure 5b reveals that the ILR and OLR radii would lie much closer to each other already when $m = 3$ or 4. This is the simple but compelling reason why Lin (1967) urged that no one should expect *multi*-armed structures to be capable of rotating rigidly as very extensive wave patterns—though he obviously slipped in phrasing it that "only *two*-armed spirals can be expected." In fact, just a halving of all labels on the ν axis in Figure 5b explains why Lin, Yuan & Shu (1969) wrote later that "the possibility that one-armed galaxies exist cannot be ruled out by the present discussion" for $Q = 1$. Curiously, however, at least the $Q \gtrsim 1.3$ exclusions ought indeed to rule out all local prospects for slow $m = 1$ waves.

Finally, as regards wavelengths, only one thing seems fairly certain. It is that the large but typical values of the reference length $\lambda_{crit}(r)$ shown in Figure 5a imply that the observed interarm distances can be accommodated in disks of dominant mass only with waves belonging to the *short* side of Figure 4—or to the side dominated by the random epicyclic motions, rather like the $|\nu| = 0.6$ choice C instead of A in that diagram. This important insight stems uniquely from Lin & Shu (1966, 1967). Indeed the analogous axisymmetric waves were known already to Kalnajs (1965), but it was Lin and Shu who first appreciated and pointed out vividly that just those short collective waves can be remarkably *useful*.

A contrary view was voiced by Marochnik, Mishurov & Suchkov (1972). These authors stressed, in effect, that the original long-wave hopes of Lin & Shu (1964) may yet deserve resurrecting. To illustrate that even such waves can be made tight enough, Marochnik et al. immobilized everything in the Galaxy except a uniform "population I" disk with $\mu = 40\ M_{\odot}\mathrm{pc}^{-2}$. That use of the old trick of reducing λ_{crit} seems a bit extreme (and indeed it was technically weak in admitting neither any shock dissipation nor the refraction of some yet shorter waves back from corotation), but it does provide one good closing reminder: Given our poor knowledge of galactic halos and/or hot disks, no one should be hugely surprised if the *effective* length λ_{crit} from examples such as in Figure 5 should need to be reduced by factors of order 2, at least for tightly wrapped waves and especially toward the interior. This would still not upset the impression that the most prominent real waves are apt to be "short" ones. However, it does make it almost meaningless to debate now whether one model or function $Q(r)$ or pattern speed Ω_p results in a photogenic diagram that fits observations better than another.

4.3 *Group Velocity*

After this long discussion of what might be feasible, it needs to be said plainly that none of these WKBJ studies of what are sometimes still called "self-sustained" density waves has yet come close to demonstrating the long-term survival of any such spiral patterns. Even today, it is not just the *origin* but even the *persistence* (cf Oort 1962) of these waves that remains distressingly unexplained; all we know for sure now is that the two topics are nearly inseparable!

In an early volume of this series, Lin (1967) wrote instead that his work with Shu "essentially amounts to establishing the possible existence of certain density waves of the spiral form, sustained by self gravitation. These are collective modes in the combined gaseous and stellar system." Even at that time, it was clear that the WKBJ approximations underlying the basic Lin-Shu theory become invalid near (and beyond) any radius where $|\nu| = 1$—and that there, as Prendergast (1967) remarked in an excellent short review, "the solution gets very difficult, and one has not yet succeeded in guiding it properly through the Lindblad resonance." Nevertheless, it seemed to Lin that his success with Shu in adjusting the speeds $\Omega \pm |\nu|\kappa/2$ over most of the radii *between* those resonances had in large measure already established a "theory free from the kinematical difficulty of differential rotation."

The fallacy of such reasoning became evident in 1969, when it was noticed that all this tuning of precession speeds had concerned only the *phase* velocity of what might logically be just a traveling wave (in the usual technical sense) rather than a genuine global mode. Lin and Shu had completely overlooked that in repairing one serious defect they had actually created another: An inevitable price for altering those speeds of precession in a wavelength-dependent manner via the (very sensible) radial forces is a *group* velocity, likewise directed radially. And it is this second kind of wave transport, familiar enough from countless other situations, that sooner or later even here plays havoc with the amplitudes—which the kinematic ovals in Figures 2 and 3 at least had the decency to keep invariant. Lin (1970) agreed that this "consideration . . . brings the problem even into sharper focus," and Shu (1970) conceded that the "amplification of trailing waves" announced by Lin & Shu (1966; see also Lin 1967) had resulted from a similar misunderstanding.

The demonstration that the wave amplitudes—or more exactly, the action densities—are indeed conveyed radially is fairly complicated, and seems best left to the papers of Toomre (1969), Shu (1970), Dewar (1972), and Mark (1974a). To dispel any needless mystery, however, let us quickly rederive the parallel effect of the group velocity upon the slow precession rates

$$\Omega_p(r,t) \equiv \Omega(r) - |\nu|\kappa(r)/2 \tag{5}$$

for a trailing, two-armed spiral wave pattern in which Ω_p varies mildly from radius to radius. In that case, the slight wave shear $\partial\Omega_p/\partial r$ near some radius r will slowly loosen or tighten that part of the spiral pattern, at a rate $\partial k/\partial t = -2\,\partial\Omega_p/\partial r$, where again $k(r,t)$ is the radial wavenumber. More to the point, as k changes with time, so must Ω_p itself at that radius, according to Equation (5), at the rate

$$\partial\Omega_p/\partial t = -\tfrac{1}{2}\kappa(r)[\partial|\nu|/\partial k]\,\partial k/\partial t = c_g\,\partial\Omega_p/\partial r, \tag{6}$$

where

$$c_g(r,t) = \kappa(r)[\partial\,|\,v\,|/\partial k]_{r,\Omega=\text{const}}. \tag{7}$$

The latter is just the group velocity, here reckoned positive when inwards. It deserves that name because Equation (6) implies that a traveler moving with the (gradually varying) speed c_g radially toward the center would find Ω_p changing not at all in his vicinity. Conversely, the Ω_p's expected in the not-too-distant future at any fixed radius are just the current values at nearby exterior radii, all as if prerecorded on a magnetic tape that is about to be played.

The severity of this radial drifting of *all* local properties of a trailing $m = 2$ wave pattern of approximate speed $\Omega_p \cong \Omega(5r_0)$ within the exponential disk is indicated by the four pairs of "characteristic" curves in Figure 5c. For each choice of Q, these curves report the radii which the imagined traveler would have gotten to at various times t_5 reckoned in multiples of the full rotation period $2\pi/\Omega(5r_0)$. Given that Ω_p is conserved along each curve, the dimensionless frequency v and wavelength $\lambda/\lambda_{\text{crit}}$ at successive radii are implied by Figures 5b and 4 exactly as before.

Since these curves are so similar, let us concentrate now on just the inner member of the $Q = 1.2$ pair. Starting off, for instance, at the point marked C (with $v = -0.6$, $\lambda = 0.42\ \lambda_{\text{crit}}$, and in terms of Figure 3, $\alpha = 11$) at $r = 3.8\ r_0$, notice that just one rotation period seems to be required for all this information to drift to point E (with $v = -0.9$, $\lambda = 0.21\ \lambda_{\text{crit}}$ and $\alpha = 16$) at $r = 2.5\ r_0$. A halving of λ_{crit} everywhere would obviously double such drift times, but otherwise this relatively tight spiral actually errs on the tranquil side: Had we kept $\lambda_{\text{crit}}(r)$ unchanged but selected instead the generally more open wave whose $\Omega_p \cong \Omega(4r_0)$, similar WKBJ estimates would have yielded a total elapsed time of 1.9 in the present units for the entire drift from the equivalent of point B to the very center. Except versus the old straw man of material arms, such a "persistence" seems not very impressive. Even the kinematic spirals from Figure 3 would last about that long with the speeds $\Omega - \kappa/2$ suggested by Lindblad.

4.4 *Forcing and/or Feedback*

Serious though they were, these group velocity criticisms were directed only at the faulty demonstration of longevity, not at the basic *idea* of nearly permanent spiral waves. As this reviewer has remarked before, no such radial propagation actually excludes really long-lived spiral wave patterns in a galaxy. If those patterns are to persist, the above simply means that fresh waves must somehow be created to take the place of older waves that drift away and presumably disappear.

That these short waves must indeed decay eventually—by Landau damping in an ILR region, at least in *linear* theory, and provided such a resonance is both (*a*) accessible and (*b*) unsaturated—was established firmly by Mark (1971, 1974a) and Lynden-Bell & Kalnajs (1972), who also corrected Toomre's (1969) inept remark that there should be similar damping "already in transit." At larger amplitudes, however, as Kalnajs (1972a; see also Simonson 1970) seems to have been the first to notice, most of the short-wave damping may well occur en route, via the gas-dynamical shocks yet to be discussed in Section 5.

Incidentally, either form of damping argues that these density waves must be trailing. The short leading waves admitted equally well by the original analyses of Lin and Shu would have the opposite (or outward) group velocity, and we would then be stuck with a task analogous to trying to create water waves that travel *away* from a beach!

It is hard enough to tell what causes the incoming waves of a trailing sense. Where could such fresh and relatively open spiral waves conceivably originate in an isolated galaxy? Broadly speaking, the safest answer today remains as obvious (and as painfully vague) as it did eight years ago: They still seem most likely to be a "by-product of some truly large-scale distortion or instability involving an entire galaxy," or the "consequences of some yet more basic density asymmetries—e.g. disturbances such as mildly bar-like waves or oval distortions which may be hard to detect but which cannot even remotely be approximated as tightly wrapped waves" (Toomre 1969; see also Lin 1970, especially p. 389, and Feldman & Lin 1973 for some related opinions).

These unconscious echoes of Lindblad's remarkable insight from 1951 stemmed largely from a single hint conveyed by diagrams such as Figure 5c: At least in a formal sense, the shorter and shorter trailing waves *C*, *D*, and *E* seem to be the descendants of other, much looser and *superposed* trailing waves such as *A* from the long-wave side of Figure 4. The latter have an outward group velocity, and they seem to *refract* into the short waves only near the "caustic" or turn-around radius *B*, where the group velocity changes sign and the amplitude no doubt tends to be larger than usual. That hint remains very murky, however, owing to the large wavelengths, $\lambda = 0.69$ and 1.28 λ_{crit}, expected already for *B* and *A*. Using the full density of the exponential disk, these lengths correspond to would-be $\alpha = 7$ and 3.5 logarithmic spirals from Figure 3. For wave *B*, such WKBJ estimates might still be crudely trustworthy (after the usual extra care to be taken near such a caustic), but no $\alpha = 3.5$ spiral can be deemed tightly wrapped. Indeed, it is not even clear that the latter can be modeled adequately with any modified local dispersion relations (cf Lynden-Bell & Kalnajs 1972). It is no disaster, of course, that these antecedent waves seem to pass so soon beyond the assured competence of WKBJ analyses. But it certainly is a pity.

It seems only fair to add that such skepticism has not been universal. For instance, Shu, Stachnik & Yost (1971) claimed that even the long waves seem self-evident in M51. Lately, the same kinds of waves have figured anew in the three-wave amplification processes (cf the solid $Q = 1$ curve branches in Figure 5c) near corotation that have been studied extensively by Mark (1974b, 1976a,b), and in the unstable spiral modes announced by Lau, Lin & Mark (1976), using just such "amplifiers."

Mark (1976c) and Lau, Lin & Mark (1976) made one very interesting point in noting that *if* the stability index $Q(r)$ can be arranged to rise steeply (and yet gently!) enough toward the center of a disk, then their asymptotic theory predicts a second *refraction* just outside the inhospitable interior, with the incoming short trailing waves reverting there into the outgoing long ones. In effect, those authors required the "forbidden" strip in Figure 5b to widen rapidly enough toward the left to intercept again a given $v_N(r)$ curve, and yet mildly enough not to violate the assumptions of WKBJ theory. This may not be easy, but if it can reasonably be arranged, such a second refraction at last promises a *simple* feedback cycle for which Lin (1970; see also some apt criticisms by Marochnik & Suchkov 1974 on pp. 94–95 of their review) has long been searching. The damping from the gas would continue, but hardly any from the ILR. Unfortunately, the big question is not whether the WKBJ idea gets stretched out of all proportion by the long waves involved in that scheme—an issue that Lau, Lin, and Mark did not resolve by offering an example in which the active density is everywhere less than one-half of the full surface density implicitly required by their adopted rotation curve. The real dilemma is that one just cannot tell, without embarking upon genuinely full-scale analyses such as we have yet to discuss, whether the special conditions (including $Q \simeq 1$ near the CR, and good dissipation near the OLR) that seem to permit these weakly unstable WKBJ modes might not also admit other, more powerful non-axisymmetric instabilities that would completely swamp the former.

Speaking of wave amplifiers, it should not be forgotten that Goldreich & Lynden-Bell (1965; see also Julian & Toomre 1966) showed long ago, via a different sort of local analysis in which azimuthal forces were retained and played a vital role, that the rapid swinging of open leading waves into trailing ones by the differential rotation has one remarkable consequence: Largely for kinematic reasons arising from temporary near-resonance with κ, this swinging manages to tap the abundant energy of differential rotation—and it amplifies those waves easily a hundred times more per full cycle than the typically two-fold growths obtained by Mark (essentially from a slow and almost axisymmetric shrinkage of already tightly wrapped waves through their most nearly Jeans-unstable wavelengths while drifting with the group velocity.) And that is just one reason why other conceivable but not-entirely-WKBJ feedback cycles—including, for instance, even a slight *reflection* of ingoing short trailing waves into short leading ones that propagate away from our internal "beach"—cannot be dismissed lightly.

Lau & Mark (1976) recently announced "a new physically significant enhancement of the growth rate," itself indebted "to the combined effects of differential rotation and of azimuthal forces." It is easily shown, however, that the key amplifying term labeled T_1 in Equations (4) and (5) of that paper exactly matches the only shear-dependent destabilizing term evident in the important Equation (76) of Goldreich and Lynden-Bell.

All things considered, only cumbersome "global" mode analyses and/or numerical experiments seem to offer any real hope of completing the task of providing the wave idea of Lindblad and Lin with the kind of firm *deductive* basis that one likes to associate with problems of dynamics. Before proceeding to such studies and their own joys and frustrations in Section 7, however, it seems well to reflect that another famous structural problem would not have progressed nearly as far as it has during the past decade or two if geologists had insisted that one should first establish theoretically from basic laws that the Earth's mantle must convect in such and such patterns. The same is true of our subject: As reviewed briefly in the next section, the galactic analogues of magnetic stripes in mid-ocean, chains of fresh volcanoes, and zones of intense crushing of relatively superficial material have lately done just about all the persuading one could possibly desire that some grand driving forces indeed are at work.

5 SHOCKS AND OTHER CONSEQUENCES

The first inkling that the interstellar material in certain galaxies may be repeatedly experiencing shock waves on a very large scale predates even the sketch by Lindblad (1964) reproduced in Figure 1. In the early 1960s, Prendergast often expressed the view that the intense, slightly curving dust lanes seen within the *bars* of such SB galaxies as NGC 1300 and 5383 are probably the result of shocks in their contained gas, which he believed to be circulating in very elongated orbits. In such "geostrophic" flows of presumed interstellar clouds with random motions, Prendergast (1962) wrote that "it is not clear what is to be taken for an equation of state," but he knew that "we should expect a shock wave to intervene before the solution becomes multivalued."

5.1 *Spiral Shock Waves*

As regards the normal spirals, Lin & Shu (1964) stressed from the start that since the gas has relatively little pressure, its density contrast "may therefore be expected to be far larger than that in the stellar components" when exposed to a spiral force field such as they had just postulated. That hint remained largely dormant, however, until Fujimoto (1968, printed tardily and in Russian; see also Prendergast 1967) combined these last two lines of thought. He showed that, when a supposedly isothermal gas layer slides past what Prendergast termed a "gravitational washboard," its steady response includes shock waves already at a modest (although finite) amplitude of that periodic forcing. To be sure, Fujimoto made one fairly serious error of analysis (spelled out by Roberts 1969, p. 140) and perhaps another of judgment—the latter in placing corotation in our Galaxy at a mere 5 kpc, rather as Lindblad had imagined. But Fujimoto's goal was commendable: he suspected that not just "the high-density hydrogen gas in the spiral arms" but also "the dark lanes observed on the concave sides of the bright arms of external galaxies may be due to the present shock wave."

This inspired guess was followed up by Roberts (1969). Together with Lin, he improved the analysis and added the further suggestion that the shock wave somehow also "triggers" the formation of the numerous young star associations that accentuate optical photographs of spiral galaxies. Both these themes are summarized neatly in the two often-copied diagrams reproduced anew in Figure 6. Roberts himself (see also Lin, Yuan & Shu 1969, pp. 734, 737) was very vague as to how the sudden increase of the mean density and pressure of the interstellar material could actually provoke the required gravitational collapse of dense preexisting gas clouds. But this hardly matters. The crucial point is that before the shock idea there had been no defensible explanation at all for the striking geometrical fact, first noticed by Baade in the late 1940s, that the main H II regions in large spirals tend to define considerably crisper and narrower arms than the rest of the

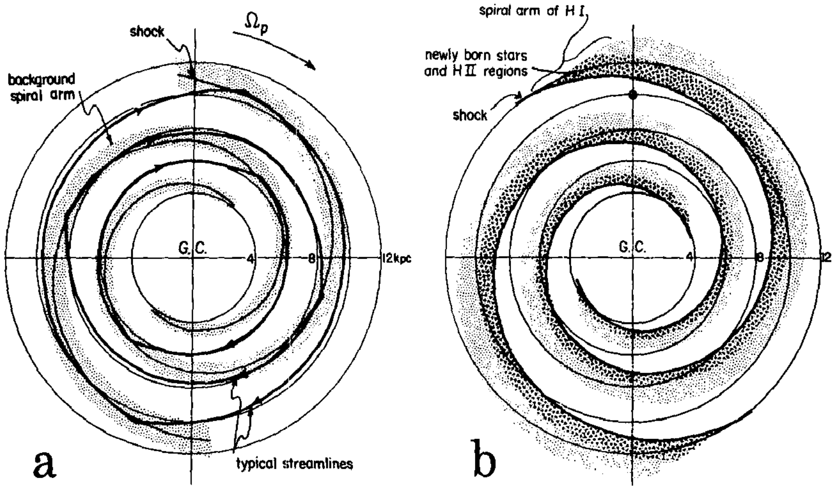

Figure 6 Large-scale spiral shock waves added by Roberts (1969) to a density-wave pattern much like that favored by Lin and Shu for our Galaxy. (*a*) Typical kinked-oval streamlines, viewed from a frame rotating with the presumed pattern speed Ω_p. (*b*) Implied relative locations of the shock fronts, the brightest new stars, and the main gaseous arms.

visible material. Often these emission nebulosities "are strung out like pearls along the arms" (Baade 1963, p. 63; see also Carranza, Crillon & Monnet 1969 and Arp 1976 for two fine modern examples). And like the dust, these highly luminous chains seem biased toward the inner edges of arms, though perhaps not quite as much.

As both Lin (1970) and Roberts (1970) rightly emphasized at the Basel symposium, the occurrence of dust lanes and O and B stars typically (though by no means always) on the *inner* or concave sides of (presumably) *trailing* spiral arms offers strong corroboration not just for the shock waves but indeed also for the theoretical expectation that the basic wave patterns should rotate more *slowly* than most of the disk material.

So much for the "volcanoes" and folded sediments of our subject. Its magnetic stripes are the bright spiral-like ridges in the radio-continuum map of M51 obtained by Mathewson, van der Kruit & Brouw (1972) using the Westerbork telescope. These ridges coincide much better with the major dust lanes of that galaxy than with its optical arms. Here indeed was "first-class observational evidence" for widespread shock-wave compression of a gaseous medium containing both cosmic rays and a weak magnetic field—a circumstance that should lead naturally (cf Pooley 1969) to a spiral pattern of enhanced synchrotron emission. Some difficulties with this picture, resulting mainly from an oversimplified description of the compression by Mathewson et al., have been reviewed already by Kaplan & Pikel'ner (1974) and by van der Kruit & Allen (1976). Hardly more needs to be said about this wave test. It speaks eloquently for itself.

The last of the three principal confirmations to date of the spiral-wave idea came, in this reviewer's opinion, also from Westerbork, this time via 21-cm line studies of M81. As reported by Rots & Shane (1975), the two main optical arms of that galaxy indeed agree well with unmistakable major arms of neutral hydrogen. In the words of perhaps the most vehement critic of the density-wave hypothesis, who himself favored a magnetic explanation, "if so, then it is generally agreed that gravitational forces must be invoked and the density-wave or some similar theory appears inevitable" (Piddington 1973).

These major items of evidence in support of spiral shocks have been accompanied by a host of other observational tests and "applications" of the density-wave hypothesis. Their conclusions have ranged from very encouraging to highly optimistic; none has proved severely contradictory. Much of this (and the above) work has been described at length in several other reviews, including those by Burton (1973), P. O. Lindblad (1974), and Wielen (1974). Rather than repeat here, let us simply close with the reminder that even in such reading one must constantly beware of the old trap of treating often-repeated presumptions as established facts. Perhaps the best warning comes from the confusing situation in our own Galaxy, where the view is far from perfect: it is not at all clear yet whether the famous 3-kpc arm is to be regarded as an ILR terminus of an inward-traveling wave, let alone as the result of violent activity in the nucleus. At least this writer still wonders seriously whether de Vaucouleurs (1964, 1970; see also Lindblad 1951, Kerr 1967) did not come closer to the truth in suggesting that it may instead be "gas streaming along a bar tilted by about 30 to 45° to the sun-center line" in the inner regions of our possibly SAB-type Galaxy.

5.2 *Pendulum Analogy*

In rushing through these important observations, we did not dwell in simple terms upon why one should expect the gas shocks to develop. Roberts (1970, Figure 3) offered an analogy in which a particle gets caught and is dragged some distance radially by the potential well of a given spiral arm. Lin and Shu have often remarked that the density contrast in each subsystem should be roughly proportional to the inverse of its mean-square random velocity. And Kalnajs (1973), somewhat like Prendergast originally, has stressed that the answer lies instead in the attempted crossings of elementary rings or orbits such as shown in the $\alpha = 16.7$ frame of Figure 3. Which, if any, of these helpful hints should one trust?

The first explanation turns out to be essentially the same as what Lindblad wrote about Figure 1—but it does not correspond closely to the sinusoidal gravity forces that Roberts used in his calculations. The second, though fine at large speeds, would clearly be wrong if carried to the zero-pressure extreme. And even the third reason, though most nearly right, does not distinguish plainly enough between the free and forced trajectories of various particles.

A closer analogy to the tightly wrapped waves adopted by Roberts (1969) and Shu, Milione & Roberts (1973) is the (supposedly endless and continuous) row of identical pendula in Figure 7 devised largely by Kalnajs. If we assume that these pendula are long enough to act like harmonic oscillators and that they do not yet bump into each other, then of course their displacements $\xi(r, t)$ from the equilibrium positions r obey the equation

$$\partial^2\xi/\partial t^2+\kappa^2\xi = C\sin\left[k(r+\xi)-\omega t\right], \tag{8}$$

when that system is exposed to the traveling sinusoidal force field given on the right. Equation (8) is exactly equivalent to the zero-pressure limit of the locally-approximated Equations (11) of Shu et al. Obviously, it also mimics the radial dynamics of stars and other collisionless objects in a galaxy subjected to strictly *axisymmetric* forcing of a wavelength short enough that variations of the epicyclic frequency κ can be ignored.

Requiring $\partial(r+\xi)/\partial r > 0$ everywhere, the linearized forced solutions

$$\xi(r,t) = C(\kappa^2-\omega^2)^{-1}\sin(kr-\omega t), \tag{9}$$

obtained after omitting ξ from the right-hand side of Equation (8), imply that one should not expect collisions as long as the forcing amplitude $C < (\kappa^2-\omega^2)/k$. This inference is corroborated by the accurately calculated Figure 7a, where C was set to one-half that nominal critical value, and where $|\omega/\kappa| = 3/4$, about as inferred by Lin and Shu for their presumed wave in the solar vicinity.

By definition, at larger amplitudes where conflicts or "shocks" should first arise, Equation (9) is no longer accurate, since $k\xi \simeq \mathrm{O}(1)$ or the response is already nonlinear. [The same criticism, by the way, applies also to any use of the LSK

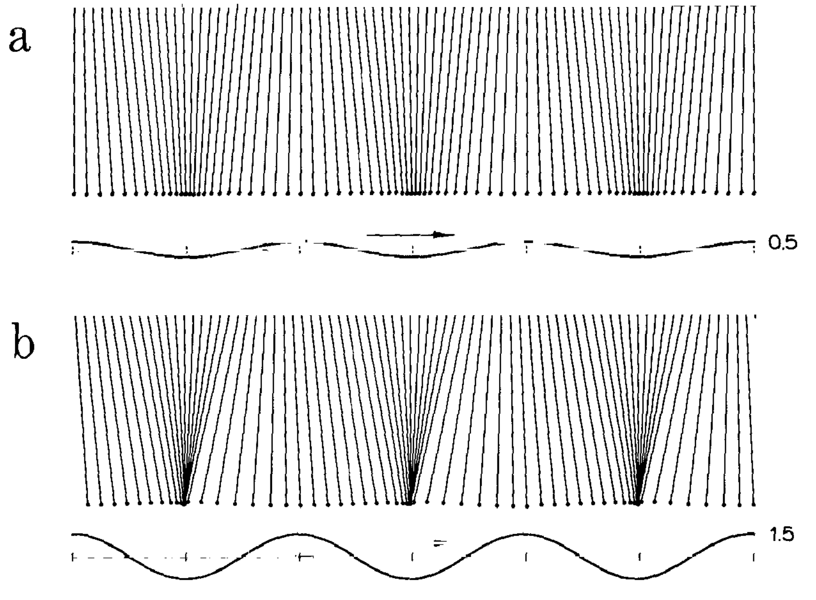

Figure 7 Forced traveling waves in a "curtain" of pendula subject to Equation (8) and inelastic collisions. The sinusoids depict, with 5-fold exaggeration, the respective forcing potentials amounting to (*a*) 0.5 and (*b*) 1.5 times the nominal critical value from the text. The frequency is $\omega = 3\kappa/4$.

dispersion relation whenever *strong* gas shocks are expected. The fact that the perturbation to the overall star density may then still be only of order 5% is misleading. The slow-moving stars expected to contribute the bulk of that extra density would already oscillate in a distinctly nonlinear manner, as illustrated beautifully by Wielen (1975) in his discussion of spiral-forced periodic orbits near the Sun.] Nevertheless, for the above frequency ratio—which lies about as far as possible from either the Lindblad resonance $\omega = \pm\kappa$ or the so-called "ultra-harmonic" resonances $\omega = \pm\kappa/2, \pm\kappa/3, \ldots$ discussed by Shu et al. and also Woodward (1975)—the accurate critical amplitude $C_{crit} = 0.430\ \kappa^2/k$ happens to agree fortuitously well with the linear estimate 0.438.

Beyond such values, of course, there is no substitute for numerical integration. Figure 7b shows one such result, obtained from forcing steadily with an amplitude 1.5 times the latter estimate. It was supposed in this calculation that the pendula (almost like lead balls, or exactly like gas particles in an isothermal shock of infinite Mach number) undergo perfectly *inelastic* collisons. Contrary to a first impression, however, these individual oscillators do not stick to each other in the intense pile-ups seen in Figure 7b. It is simply the ram pressure of the new arrivals from the right that persuades the early comers to linger for a while (19.6% of the full cycle, to be exact) in these peculiar traffic jams near the minima of the imposed potential, until at last they can slip away quietly to the left.

Compared with this naive rederivation, the inclusion of a uniform "effective speed of sound," a = typically 8 km sec^{-1}, in the analyses of Fujimoto, Roberts, and especially Shu, Milione & Roberts (1973) seems to have had two main effects. One is that, understandably, such a finite intrinsic speed softens the density contrast across a shock and also widens the zone of crowding behind it. In particular, if $\omega/ka < 1$—or whenever the forcing appears subsonic in a mean sense—shocks arise only with difficulty (cf Figure 6 of Shu et al.) and tend to yield "broad regions of relatively low gas compression." As Roberts, Roberts & Shu (1975) stressed, this may well account for the thick and "massive" arms seen in smaller spirals like M33.

The other and perhaps more surprising effect arises when the speed $c = \omega/k$ of the imposed force sinusoid appears moderately supersonic: The finite gas pressure then makes it possible to obtain shocks from considerably weaker forcing than in our $a \to 0$ hypersonic limit. It also causes the shocks from a given forcing to be markedly stronger in terms of impact velocity. For example, as Shu et al. reported for $r = 10$ kpc, $\nu = -0.72$, and Mach number $c/a = 1.63$, the first cusp or zero-strength shock occurs in their isothermal gas when the forcing amplitude C is just under a fraction $F = 1.0\%$ of the central force $\Omega^2 r$. By contrast, the pendula in the same setting would not have bumped until $F = 4.0\%$. Moreover, using $F = 5\%$ as again recommended by those authors for the nearby Lin-Shu wave, their $a = 8$ km sec^{-1} gas enters each shock with a relative speed $u_1 = 26$ km sec^{-1} and exits with speed $u_2 = a^2/u_1$, whereas our $a = 0$ pendula would have managed to crash only with speed $u_1 = 12$. With an intermediate sound speed $a = 4$, by the way, one would already find that $u_1 = 22$—which underscores that the extra shock-making declines only slowly with increasing Mach number.

Technically speaking, there is no real mystery about this perverse supersonic behavior. Those stronger shocks seem to occur essentially because the pressure term in Equation (11) of Shu et al. raises the natural frequencies of our pendula, so to speak, markedly closer to the first ultraharmonic resonance—and no doubt it also contributes something akin to the classical Riemann steepening of free acoustic waves (which is a process discussed briefly by Woodward 1975 amidst his nice time-dependent shock calculations). Nevertheless, this reviewer cannot help wondering whether the shock-making might not have been seriously exaggerated by the isothermal model itself. Can we be sure that this "bouncy" inviscid gas of Fujimoto and the later workers mimics well enough the variously damped and heated multi-phase stew that is the real interstellar medium? The question matters because the increased impact speed u_1 seems a very mixed blessing in the context we are about to describe.

5.3 *Damping*

If the forcing of the very inelastic pendula in Figure 7b were to cease abruptly, we would know instinctively that their present strong collisions (as opposed to many lesser "aftershocks") would endure only perhaps another half-cycle. And, if instead of stopping suddenly, this forcing stemmed from the collisionless but not entirely random swinging to-and-fro of other, much more massive pendula (not shown here) that interact gravitationally with ours, then it would be just as clear that the steady drain of mechanical energy via the visible crashes would in the long run cause those unseen collective oscillations to decay as well.

The first of these thought experiments obviously mimics the demise of free axisymmetric shock waves in a purely gaseous disk. Their rapid damping may seem irrelevant, since it is the second case that at first sight imitates much better any tightly wrapped density/shock waves in a real galaxy, where the gas is well known to comprise only a small fraction of the total mass. Kalnajs (1972a) pointed out, however, that a serious fallacy may lurk even here.

The crucial point, in the context of our Galaxy, is that relatively few disk stars manage to take part significantly in waves of such short, $\lambda \cong 4$ kpc length as Lin, Shu and Roberts adopted for the solar vicinity. Using the conventional value of κ, one calculates easily for $\lambda = 4$ kpc that only about 12% (or 40%) of the stars from a Schwarzschild distribution with $\sigma_u = 40$ km sec^{-1} would have random motions small enough to ensure more than the 60% (or 0%) participation discussed at the end of Section 4.1. Thickness corrections would severely reduce even that. This helps explain why Lin (1970) reported that the nearby wave amplitude $\mu = 11\ M_\odot$ pc^{-2} corresponding to a disturbance gravity $F = 5\%$ should consist almost equally of excess gas and star density. Conversely, however, this argument cautions—ignoring the group transport of fresh waves for the moment—that the Roberts shocks might still be damping the local Lin-Shu wave almost as fiercely as one would have guessed for the gas by itself!

Much paraphrased, that is what worried Kalnajs—who himself, like Roberts & Shu (1972) in a not very convincing rebuttal, couched it all in the more expert but also more opaque terms of net torques and negative densities of wave angular momentum. To reassess this vital matter, it is simpler to pretend again that we are dealing with exactly axisymmetric density waves $\mu'(r,t) = \mu_w \cos k(r-ct)$. Let us also suppose, although perhaps more dubiously, that their amplitudes μ_w are small enough for infinitesimal estimates of the density E_w of wave energy per unit area to remain reasonably accurate.

For such self-consistent waves in an extremely thin isothermal gas layer of unperturbed surface density μ, this energy density would equal $E_w = \frac{1}{2}\mu c^2 \cdot (\mu_w/\mu)^2$, quite independent of the (horizontal) sound speed a. The analogous formula for the LSK waves in a razor-thin stellar sheet is more ponderous [cf Toomre 1969, Equation (33)], but it can be recast as just a multiple $H(\lambda/\lambda_{\rm crit};\, Q) > 1$ of that elegant gas result. Some typical values of H are 1.17, 1.55, 2.66, 4.74, 16.86 for the $Q = 1.2$ points A, B, C, D, E in Figure 4; the steep rise of that correction factor toward the smallest wavelengths is helpful, for it means that such short stellar waves tend to contain considerably more kinetic energy (and hidden commotion) than one might have supposed from their phase speed c and net density μ_w alone.

To be compared with the available E_w is the loss ΔE of mechanical energy per unit area and per time interval $T_s = \pi/(\Omega - \Omega_p)$ between successive shocks, here imagined to be steady. As follows strictly from time-averaging over one such cycle, the variable $v(\eta)$ in Equation (11) of Shu, Milione & Roberts (1973), this loss amounts to $\Delta E = \frac{1}{2}\mu_g[u_1^2 - u_2^2 - 2a^2 \ln(u_1/u_2)] \equiv \frac{1}{2}\mu_g u_E^2$, where μ_g is the mean density of the supposedly isothermal gas. At that rate of depletion, the wave energy would last a time

$$T_E = (E_w/\Delta E)T_s = (\mu_w^2/\mu_g\mu)(c/u_E)^2 H T_s. \tag{10}$$

As implied already, $c = \omega/k = 13$ km sec^{-1} and $\mu_w = 11\ M_\odot \text{pc}^{-2}$ for the specific $\lambda = 3.7$ kpc, $\nu = -0.72$, $F = 5\%$ wave at $r = 10$ kpc defined in Table 1 of Roberts and Shu. Assuming $Q = 1$ and ignoring thickness and gas corrections for brevity, those data also imply $H = 5.30$, $\mu = 89\ M_\odot \text{pc}^{-2}$, and $T_s = 2.7 \times 10^8$ yr. Locally, the choice $\mu_g = 5\ M_\odot \text{pc}^{-2}$ seems conservative; together with the above items, and $u_1 = 26$, $a = 8$ km sec^{-1} discussed earlier, it yields a remarkably short $T_E = 2.0 \times 10^8$ yr. This rough appraisal seems, if anything, a little harsher than $T_E = 0.8$ to 1.5×10^8 yr claimed by Kalnajs—who used a perhaps slightly too pessimistic $\mu_g = 10$, and whose low estimate of 0.8 also suffered almost twofold from an erroneous neglect of the $\ln(u_1/u_2)$ term in ΔE mentioned only parenthetically in his paper.

The counterproposal by Roberts and Shu that the relevant decay time works out more like 6 to 9 $\times 10^8$ yr referred in fact to the entire interior annulus between $r = 5$ and 10 kpc, where at least the unperturbed stars seem more abundant. Moreover, those authors adopted a mean gas density $\mu_g =$ only 3 $M_\odot \text{pc}^{-2}$, a value strangely at odds with the 6 $M_\odot \text{pc}^{-2}$ wave amplitude (cf Yuan 1969, p. 897) in the gas alone imagined to provide half of the local $F = 5\%$ forces in the first place. In retrospect, that mean density also seems low in comparison with the dramatically increased interior amounts of molecular hydrogen inferred somewhat tentatively by Gordon & Burton (1976) from CO data. Using instead $\mu_g = 8$ and 12 $M_\odot \text{pc}^{-2}$, as hinted by Gordon and Burton for $r = 8$ and 6 kpc without even including any He, it also appears that $T_E = 1.8$ and 1.9×10^8 yr for the $F = 5.1$ and 5.7% waves considered by Roberts and Shu at those two interior radii.

Like the earlier criticism involving the group velocity, all this may not be quite as devastating as it sounds. For one thing, the increased stellar participation (and faster c_g) in more open spirals should help considerably. Second, as hinted already, the isothermal model may exaggerate the shock losses. And third, one would suppose in any case that those losses should be largely self-correcting upon some reduction of amplitude—given that the shock-making must be a threshold phenomenon that sets in only beyond some minimum level of forcing. The pendula make these last two points very nicely: If only we could be sure that the nearby gas behaves in that naive fashion under the $F = 5\%$ forces, we would reckon that $T_E = 5.1 \times 10^8$ yr, whereas below $F = 4\%$ there would be no damping at all. More annoyingly, however, the idealized $a = 8$ km sec^{-1} isothermal gas under $F = 4\%$ forcing would still imply $T_E = 2.1$, 2.2, and 3.6×10^8 yr at $r = 10$, 8, and 6 kpc. Alas, it is simply not known yet which story is more to be believed.

What does seem broadly clear, nonetheless, from this insight by Kalnajs is that at least the tighter Lin-Shu waves in any fine "normal" spiral should—during the very process of creating their own luminous froth while getting shorter and thus more intense—be almost as capable of dissipating most of their arriving mechanical energy as any foaming water waves incident upon a gently sloping beach. Better still, just as with the ocean waves that start breaking well offshore, it looks as if this *almost insatiable* damping may not even "scour" very badly the vicinity of an ILR itself. Ironically, though, Kalnajs was wrong in his closing remark that "it takes about 10^9 yr for help to arrive" in the solar region via the group velocity. Another look at Figure 5c near its point D should persuade that his estimate is too large, probably at least by a factor 2. In this respect, Roberts and Shu scored a good point in replying that it may be "no accident that the 'damping time' and the 'propagation time' should be comparable." Indeed not. That is exactly what one finds with those breakers off the beach.

Not to be confused with this remarkably potent wave damping is the related fact that any orbiting gas that persists in overtaking a pattern of trailing spiral shocks must on the average drift inward as well. Such a net transfer of angular momentum from the gas to the stars seems to have been noticed first by Pikel'ner (1970; see also p. 172 of Simonson 1970, Kalnajs 1972a, Shu et al. 1972, and Section 6 of the review by Kaplan & Pikel'ner 1974). It meant not only that Roberts (1969) was slightly mistaken in claiming the gas orbits to be closed in the present Figure 6a. Much more important, this inexorable radial evolution implies that even the neatest spiral structures can at best be only quasi-steady. As Pikel'ner noted also, such inward drifting of the gas in very open spirals should not even require many revolutions, given strong shocks.

5.4 *Spiral Shocks without Spiral Forcing*

It would be regrettable if all this well-deserved emphasis upon, first, the tightly wrapped waves of density and force alike, and then the very nonlinear gas dynamics which the latter can soon provoke, were to leave the impression that it is the only conceivable route to making spiral shocks in a galaxy. Such it definitely is not, as Figure 8 now dramatizes.

To produce this diagram, Sanders & Huntley (1976) began their computer simulation with a uniform disk of almost cold "gas" without any self-gravity. It merely revolved differentially in an inverse 5/4-power central force field. Then, rather gradually, some fairly modest (up to $\pm 13\%$ tangential, $\pm 3.25\%$ radial) extra forces were introduced, as if coming from an *oval* distortion to the density of the unseen background stars. These imposed $m = 2$ forces without any spirality of their own were presumed to rotate at a steady rate, placing corotation and the two Lindblad resonances at the radii indicated. The resulting spiral wave stirred up amidst the shearing gas is pictured in Figure 8, approximately one revolution after the forcing had become steady. That response itself, however, remains only quasi-steady—owing essentially to the radial drifting (and its analogue outside CR) known already to Pikel'ner. The shocks as such are none too clear in this reproduction, but at least they can be located by comparing the concave and convex sides of the arms between the ILR and CR.

Artificial though it is, Figure 8 stresses beautifully that the manner in which Lindblad's "characteristic density wave . . . incites the formation of spiral structure by its disturbing action on the internal motions" may not even be so subtle as to require, as an intermediary, the spiral gravity *forces* from any vaguely Lin-Shu type density waves apt to be excited at the same time. Similar warnings have often been uttered by Kalnajs (1970, 1971, 1973)—e.g. "the density wave [in an unstable mode illustrated in his 1970 paper] is essentially a bar-like distortion of the central region

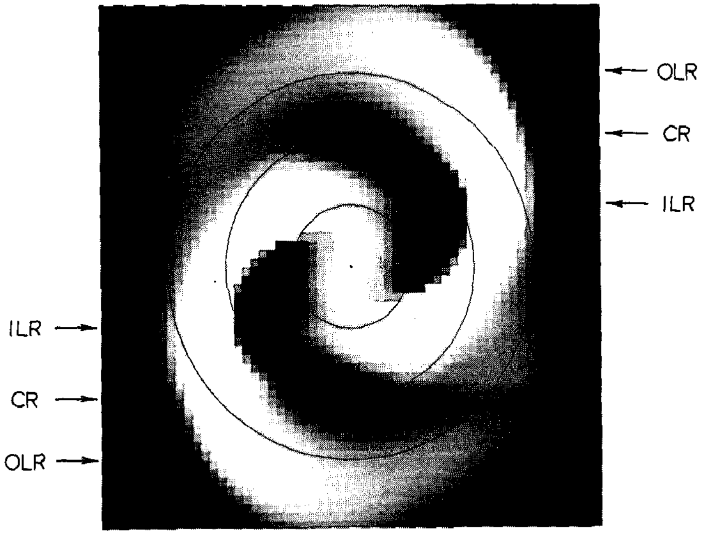

Figure 8 Quasi-steady distribution of density in a shearing gas disk exposed to a mildly oval gravitational potential (Sanders & Huntley 1976). Both that forcing and the gas again rotate clockwise.

of the galaxy which drives the gas." Such alarm signals were also fairly explicit (cf Figure 11) in the N-body experiments of Miller, Prendergast & Quirk (1970) to be discussed below. Lately, Lynden-Bell (1974, 1975; see also 1972, with Kalnajs) has stressed anew that "it is not at all clear to me that the gravity of the 'spiral' is 'important' for the formation of spiral shocks. My advice to observers is to believe those parts of the theory based around propagating shocks and shock-induced star formation, and to be very wary of those parts that rely heavily on the self-gravity of anything but a slight oval distortion or a bar shape."

Faced with such advice, this reviewer remains ambivalent. On the one hand, it seems eminently sensible to presume that the gas forcing might be rather direct in SB/SAB spirals like NGC 1097 (cf Arp 1976 again, though where is corotation?) and perhaps even in such milder SAB types as NGC 4321 = M100. Yet the same warnings seem overdone for the "best" SA spirals, essentially for two reasons. One is that, as Sanders and Huntley explained nicely, their kind of immediate forcing can shift arm longitudes only 90° per resonance radius—and the three familiar resonances are simply too few to meet such hopes in NGC 5364 or M51 or perhaps even M81. The other reason is that Zwicky (1955; see also Sharpless & Franz 1963) really seems to have been onto something with his "composite photography" of M51. As Schweizer (1976) reported from a detailed photometric reexamination of this and several other well-known spirals, the thick "red arms" discovered by Zwicky indeed represent *broad* spiral variations of surface brightness (though not of color!), as if from pronounced waves of similar spirality among the background disk stars as well.

Classic examples like those continue to argue strongly for something much more like the Lin-Shu-Roberts waves than the picture in Figure 8. Yet, as we now know, such waves themselves seem desperately in need of help from larger scales. Hence the main question to be distilled from the last few paragraphs might just be whether the much-desired "crushing" of the gas by forces stemming ultimately from major bars, ovals, long waves, or whatever is best regarded in any given instance as direct, indirect, or some shade in-between. Differences of opinion here among us theorists may merely be akin to those expressed by the blind men who examined the elephant.

It seems right to end this long discussion of galactic shocks almost where it began. A fascinating computer study of (again admittedly impermanent) gas motions in some *strongly* bar-like potentials—with the shocks there falling indeed mostly at the observed locations of the dust lanes that tantalized Prendergast long ago—has recently been published by Sørensen, Matsuda & Fujimoto (1976).

6 A QUICK SYNOPSIS

Lin (1967) has long stressed that he and Shu—with their QSSS hypothesis that a wavelike spiral force field somehow exists and keeps on rotating—jumped deliberately right into the *middle* of a difficult problem, hoping that from this "focal point in the development of a theory" it might be possible to progress faster by working "in both directions." By this technique, much more defensible here than in mountaineering, these and quite a few subsequent explorers landed essentially at point *B* of Figure 9. As discussed in Section 4, the route "downhill" from *B* toward first principles at *A* has since proved surprisingly treacherous; at least, no one has yet managed to descend safely all the way. On the other hand, the going "uphill" from *B* toward the gas arms and spiral shocks at *C* has, as seen in Section 5, turned out to be delightful despite some crevasses. Of course, a nasty segment from *C* to *D* still lies ahead—though even it has lost some of its former terror, chiefly to the helpful notion of a two- or multi-phase interstellar medium favored since the late 1960s (cf Shu et al. 1972, Salpeter 1976). Happily, that task belongs mostly to other specialists. For wave theorists, the big challenge remains much more the logical gap from *A* to *B*. It had better get closed firmly lest those reconnaisance parties far ahead become stranded.

This thumbnail sketch says nothing about the many ragged or patchy or "multi-armed" spiral features that both outnumber and complicate the generally-agreed two-armed "grand designs" among what Hubble called normal spirals. Worse still, our little summary seems even to have forgotten that "barred spirals are not rare oddities of nature. Among 994 bright spirals, about 2/3 show bar structure: 31% SA, 28% SAB, 37% SB, 4% S," to quote Freeman (1970b; see also Sandage 1975) and indirectly de Vaucouleurs.

The first flaw is a real one, to be scarcely remedied in Section 8. As for the bars, however, it is increasingly clear that they, too, belong in Figure 9 as closely related wave phenomena. They are basically to be thought of, it seems now, not as the stiffly revolving Jacobi ellipsoids that one still tends to see quoted, nor as the Dedekind ellipsoids in which the fluid circulates along ellipses but the shape stays put, but more like the in-between Riemann ellipsoids with both rotation and circulation that Chandrasekhar (1969, for a good summary) did much to resurrect and for which Freeman (1966) devised some elegant stellar-dynamical analogues. Of course, the real bars are not isolated entities but tend to be immersed deep within reasonably normal disks, as Freeman (1970b, 1975) stressed in his reviews. An even more vital clue, as he wrote already in 1970, is that unlike the air gap in Hubble's tuning fork, "the transition SA-SAB-SB is so smooth that it seems inconceivable

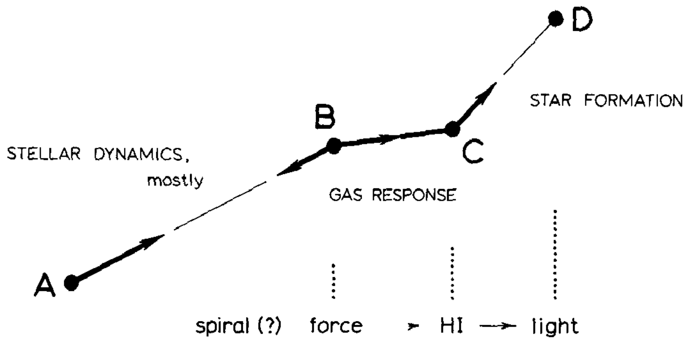

Figure 9 Status of the wave theories. As yet, only the thickened portions of this schematic route seem reasonably secure.

that the structure-producing processes should be essentially different for SA and SB systems." These remarks much deserve to be kept in mind as we now examine briefly in Section 7 the difficult and as yet inconclusive struggle upward from *A* via various large-scale computations.

7 GLOBAL INSTABILITIES

With the sole exception of the completely flattened and uniformly rotating Maclaurin-Freeman disks—for which Bryan (1889), Hunter (1963), and Kalnajs (1972b) together showed how to obtain the entire spectrum of modes, even when "hot," using only pencil, paper, and Legendre polynomials—all existing studies of the very large-scale or "global" behavior of disk-like model galaxies have been heavily numerical. Some of these studies, devoted again to *linear* instabilities and/or modes, have at least made a mild pretense of analysis. Others have amounted frankly to brute-force time-integrations of the *nonlinear* equations of motion of up to about 10^5 gravitationally interacting particles, with their only elegance consisting of various "checkerboard" schemes for rapidly solving the Poisson equation. As might be expected, the mode searches have tended to be much less expensive and, within their limited domain, also more accurate than these large N-body experiments. Yet it is indeed the latter that have taught us far more of what we really needed to know.

7.1 *Tendency toward Bar-Making*

Three of those important lessons are summarized well in Figure 10, which is due to Hohl (1971). One lesson was that disks of stars with random motions barely sufficient to avoid the Jeans instabilities still tend to be grossly unstable in an $m = 2$ sense. Secondly, there is nothing long-lived about the *spiral* structures that ensue among such "stars." And a third point was that the much hotter stable configurations into which these experimental disks rather quickly convert themselves tend often—though not always—to include some sort of *oval* shapes that continue to revolve indefinitely in their interiors. Essentially the same conclusions were reached already in the numerically coarser but physically yet more interesting experiments of Miller, Prendergast & Quirk (1970), which typically included also a subsystem of dissipative "gas" whose own strikingly different behavior is shown, as yet a fourth topic, in Figure 11. Let us consider these four topics in 3-4-2-1 order.

The tendency toward bar-making evident in both figures deserves exceptional emphasis because it is the one corner of our subject where the quest for durable *wavelike* distortions of theoretical galaxies can with some justice be said to have succeeded already. This does not mean that such a finite-amplitude lock-in phenomenon is at all well understood yet. However, it has now arisen in enough situations [including the much-softened 1000-particle integrations by Dzyuba & Yakubov (1970) and—in a milder sense—also the pioneering N-body experiments of P. O. Lindblad (1960)] that even this skeptic feels convinced that it represents the basic recipe for the survival of the bars and oval shapes seen in the interiors of actual SB and SAB galaxies.

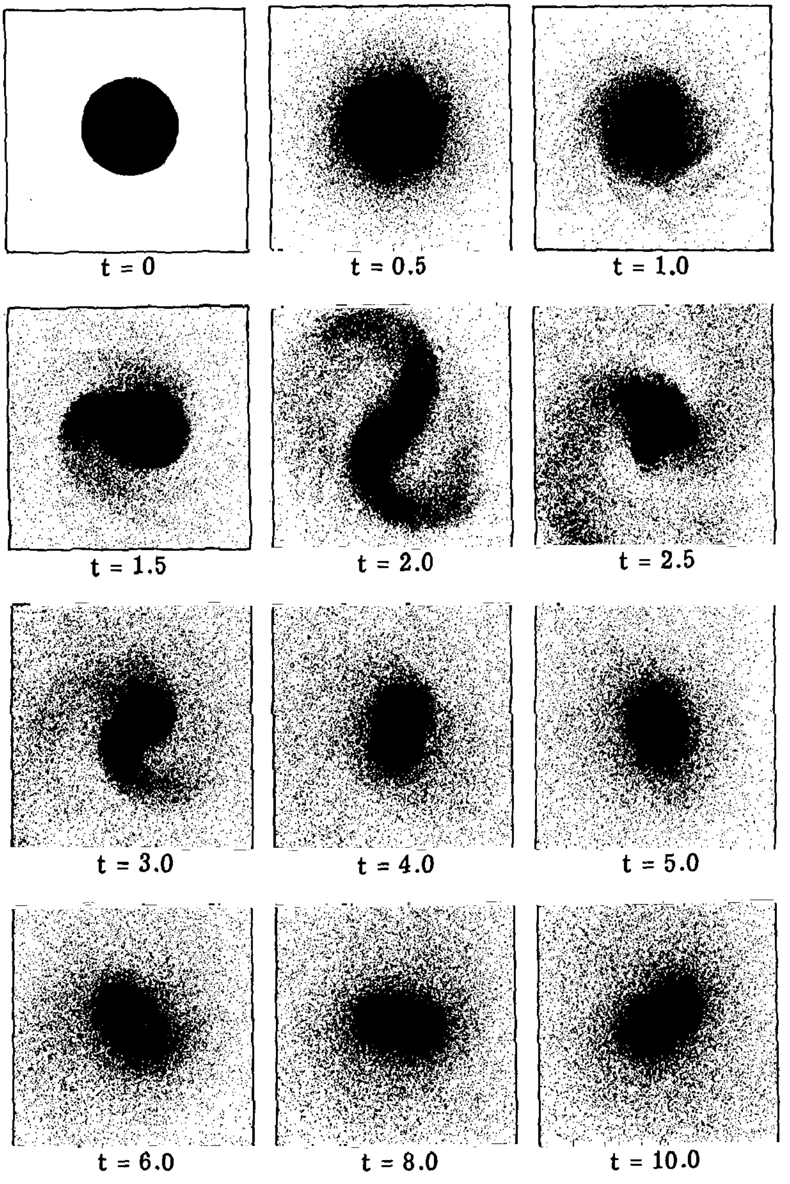

Figure 10 Evolution of a uniformly rotating disk of 10^5 mass points (Hohl 1971). Its surface density began like that of a cold Maclaurin disk; time is reckoned in rotation periods of the latter. Gaussian random velocities were preassigned from Equation (3); hence that model started only roughly in equilibrium, but it was nevertheless essentially stable in an axisymmetric sense, as Hohl also showed.

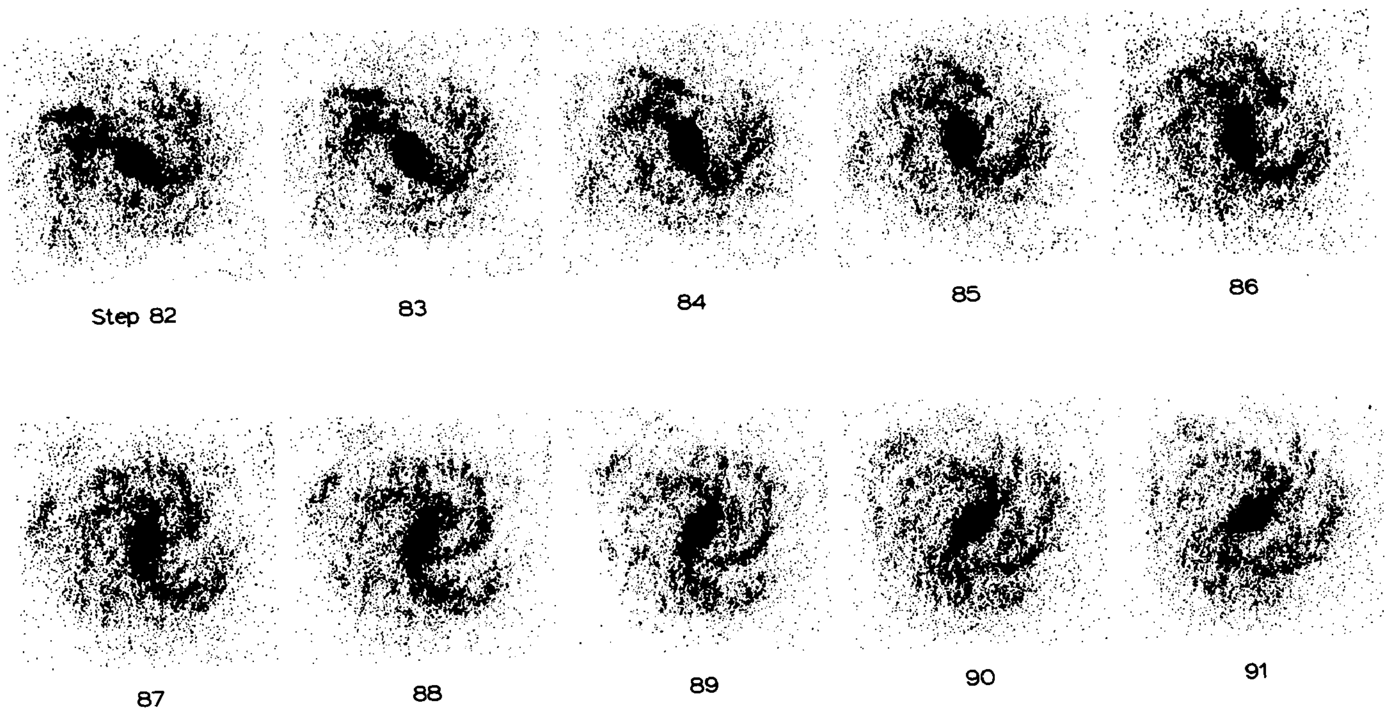

Figure 11 Evolving patterns in the "gas" component of a rotating disk consisting of some 10^5 collisionless "stars." Shown here are the contents of only the central half-squares from Figures 3 and 4 of Miller, Prendergast & Quirk (1970). During this short sequence, the total mass of "gas" decreased from 16.4 to 16.0 per cent of the whole, owing to gradual conversion of such material into additional "stars."

The main thing to notice is that the "bars" in Figures 10 and 11 revolve *in the same direction but more slowly* than the computer particles that undoubtedly circulate (and gyrate) within them. It is, of course, hopeless to try and trace any of the latter in these diagrams, but at least Figure 11 shows clearly enough that some of the material lumps orbiting outside the bar actually manage to overtake it. For Figure 10, Hohl reported and the stroboscopic effect confirms that the rotation period of the oval blob seen in the later frames equals 2.25 times the original spin period—whereas the orbital period of any relevant interior particle probably remains less than 1.5 even then. What we have here, in short, is again something very akin to the slow advance of Lindblad's "dispersion orbit" in Figure 2. To be sure, there are now many players, plus the hot and massive "star" disk that is not even shown in Figure 11. All these interact with each other in an appallingly nonlinear and non-local manner amidst much random motion. Yet, remarkably, the outcome falls short of complete axial symmetry. Instead, it seems that the various precessing orbits even here, helped no doubt by the familiar near-constancy of $\Omega-\kappa/2$, still manage well enough to "trap" each other—for reasons that Contopoulos (1970, 1975), Vandervoort (1973), and especially Lynden-Bell & Kalnajs (1972) made rather plausible—to yield the oval shapes that persist near the center. Characteristically, the angular speeds of even these special kinds of density waves, although slow, again exceed slightly the typical values of $\Omega-\kappa/2$ (cf Quirk 1971, Figure 6).

A second but already considerably vaguer sort of encouragement for wave theorists came from the *spiral* structure in the "gas" outside the obvious bar shape in Figure 11. A lot of this structure is chaotic, and even the nice "third" arm seen shearing rapidly past the 3 o'clock position at step 88 can in essence be only a transient material feature. Besides all that "noise," however, these pictures together with the magnified frames 101–103 and the yet later steps 121–130 shown by Miller et al. convey also a fuzzy but persistent image of a reversed *S*, with the bar serving only as its middle segment. This must indeed be a wave. Similar remarks apply to the barred and vaguely two-armed shapes obtained by Hohl (1971, Figure 17) upon "cooling" one previously featureless hot stellar disk in a gentle but admittedly artificial manner.

Miller, Prendergast and Quirk wrote that "the spiral pattern [or did they chiefly mean just a strong *impression* of spirality?] lasted for about three complete revolutions"—that is, approximately from step 60 till 160, by which time the "pattern . . . finally became too close to be seen in our pictures" and even the bar had mostly disappeared. It would be delightful to infer from this half-success that the bar was somehow essential for preserving that spiral. Regrettably, these authors and especially Quirk (1971) concluded after much thoughtful discussion and Fourier analysis that such an interpretation (among others) would be much too simplistic. They agreed that some sort of a stellar "asymmetry seems to 'drive' the pattern." Yet the mild, warped-oval $m = 2$ shape analyzed in detail in Quirk's Figures 3 and 4 plainly involved *comparable* disturbance densities from the gas and stars alike, and it also extended at least 3 times as far in radius as the obvious bar in our Figure 11. Thus, despite some resemblance to the situation later idealized in Figure 8 by Sanders and Huntley, it was hard to tell here just what had caused what.

The total lack of *spiral* shapes of respectable duration not only in Figure 10 but also in every other purely stellar-dynamical experiment conducted with sizable fractions of "mobile" mass (e.g. some by Hockney & Brownrigg 1974, or others by Hohl using initially Gaussian and exponential disks) is one of those results that almost speaks for itself. It is conceivable, of course, that some milder instabilities, which might themselves have led to more enduring spirals, were thwarted in these experiments by a kind of overheating from the fierce initial behavior. This seems unlikely, however, because of Hohl's extra tests with that artificial cooling: Whenever it was switched off again, the spiral features soon disappeared, no doubt speeded by the reason that follows.

7.2 *Ostriker-Peebles Criterion*

As is well known by now, what remained even after the efforts at cooling, by Miller (1971, Figure 6) as well as by Hohl, were some astonishingly hot disks of make-believe stars, with more random motion than actual rotation. Speaking not only personally, the trickle of such reports from about 1968 onwards came indeed as a rude surprise, especially when augmented by the similar analytical findings of Kalnajs (1972b) as to what it takes to stabilize fully those Maclaurin-Freeman disks. It was of little comfort to recall that "a question which the present discussion [of the short and mostly axisymmetric instabilities] leaves completely unanswered is . . . whether or not a given disk might prefer to develop into a bar-like structure" (Toomre 1964). The trouble with such face-saving was that this writer also believed then (and still does now, though much more timidly!) that temptations like the bar-making *ought* to belong mainly to the inner and more uniformly rotating parts of a disk. Out here in this Galaxy—as hinted by a probably exaggerated near-agreement (cf Toomre 1974a) between the observed and required random motions of the nearby K and M dwarfs, and also by the absence of true nonaxisymmetric instabilities of a *local* sort (cf Julian & Toomre 1966)—it had seemed a safe bet that a stability index like $Q = 1.5$ should curb all lesser misbehavior, and yet not exclude the Lin-Shu waves. Instead, the experimenters kept on reporting $Q \approx 2$, 3, and upwards.

In retrospect, Miller (1971) was clearly too modest in still entertaining the idea that this severe heating arose mostly from procedural flaws in the simulations. Since then, as Miller (1975; see also Hohl 1975) has rediscussed, fears of much-increased relaxation effects in these very thin disks—though quite serious in principle (Rybicki 1971)—were in practice largely laid to rest not only by noting that the near-gravity was much softened in the actual computer codes, but also by checking with further experiments (Hohl 1973, Miller 1974) either known or constrained to remain axisymmetric. In addition, Hohl (1972) confirmed experimentally for the uniformly rotating exact models what Kalnajs already knew from his analyses: they must be hot indeed to avoid the bar-making.

Most convincing of all was the discovery by Ostriker & Peebles (1973; reviewed thoroughly by Bardeen 1975) that the ultimate and/or stable forms of these disks of collisionless particles share the empirical rule or criterion

$$T/W \equiv (\text{organized K.E.})/(\text{negative P.E.}) \lesssim 0.14 \qquad (11)$$

that Ostriker had come to expect for the secular stability of various rotating models of stars. With that, it was suddenly beyond all reasonable doubt that the various investigations were telling essentially the same story—and that it was not a story of numerical errors.

Even today, however, it remains much less certain what this surprising unanimity actually *demands* of seemingly disk-like galaxies like our own. A lot of "heat" is clearly needed, but must it be hidden in faint, massive halos, as Ostriker and Peebles suggested? Or might some very hot inner disks or "spheroidal components" suffice already? Here probably the most important thing to remember is that the Ostriker-Peebles criterion is still not a theorem but just a fine summary of the existing N-body experiments. Reasonable though they seem, those experiments have in fact not been very diverse. One strong bias has been toward disks initially in fairly uniform rotation. Another and possibly even more serious limitation may have been the fact that nearly all constituents of those experimental disks were either prearranged or "cooled" again to rotate in the *same* direction. This concern may seem bizarre until we recall (*a*) that no one would set out to construct a stable spherical galaxy from stars with only one sense of angular momentum, and (*b*) that the average disk galaxy indeed contains a spherical component. Almost none of the existing N-body simulations, it seems, has either included or managed to develop any sizable fraction of stars orbiting backwards near the center. As Kalnajs (1977) has just explained, such mobile retrograde stars ought to interfere considerably with any bar-making tendency.

In short, the *need* for proper halos has still not been established firmly in our subject—though this is neither to imply that they would be unwelcome, nor to dispute the growing evidence from flat rotation curves (cf Roberts 1975, Krumm & Salpeter 1977) in favor of much extra mass of low luminosity. In fact, halos of interest to disk theorists would not even have to be inordinately massive. Roughly speaking, it seems (cf Kalnajs 1972b, Bardeen 1975, Hohl 1976) that already the adoption of a *rigid* halo—or any other frozen background distribution—of total mass comparable to that of the mobile disk within it cures at least the violent bar-making. Of course, immobilizing all but 10 or 20% of the mass does even better. As Hohl (1970) first illustrated and Hockney & Brownrigg (1974) confirmed, experiments starting with cold disks then yield not exactly "a spiral density wave, unchanged for at least 10 rotations" (as the latter authors claimed extravagantly) but something else almost as interesting: What one tends to find are multi-armed structures that survive only in a (gradually evolving) *statistical* sense—though indeed for quite a few revolutions—while their detailed features wind up and yet somehow keep on reappearing!

It is doubtful, however, whether any postulated halos, especially if they rotate to some extent, can be relegated safely to an altogether passive role. For one thing, as several authors including Mark (1976c) have suggested in the logical footsteps of Sweet (1963), some "two-stream" instabilities might well arise between such systems and the cooler disks embedded within them. And more important, even those very hot collections of stars might, like the one shown in Figure 10 or hoped for in Figure 8, still prefer oval distortions of their own.

7.3 *Unstable Spiral Modes*

The main alternative to these fascinating but expensive N-body experiments has been to pretend that a galaxy consists not of a finite number but a *continuum* of stars and/or gas particles, and to try and determine the fates and shapes of various large-scale but *small-amplitude* perturbations to disk-like equilibrium configurations of such material. Serious nonaxisymmetric efforts to explore the overall stability and assorted global modes of shearing disk galaxies in this manner go back at least to Hunter (1965) and Kalnajs (1965). Since then, as summarized in part by Hunter (1972) and Bardeen (1975), a lot of additional cleverness and computing money has been expended—indeed far more than is immediately evident from the literature. Yet the overall progress in this area has been painfully slow.

That situation should improve when several major ongoing studies, particularly the ones by Kalnajs (1971, 1976) and Bardeen (1975), reach print in final form. These promise many interesting comparisons with WKBJ estimates, stability criteria, resonant particle effects, and ideas about angular momentum transfer. Here it seems foolish to try and anticipate such conclusions. Instead, it may be more instructive to look backward for a moment and to ask why all this laudable global-mode calculating has taken so long.

The reasons involve both principles and complexity. For one thing, it has gradually dawned upon most workers that stable self-gravitating disks (especially ones composed of collisionless stars) may not even possess many *discrete* normal modes of the sort that one associates with church bells and Cepheids. The vibrational spectra of our systems seem, by and large, to be *continuous*. The associated modes must look disagreeable indeed near any resonance radii—somewhat like the van Kampen (1955) modes from plasma physics—as Hunter (1969) demonstrated for certain $m \neq 0$ waves near their corotation radii, and as Erickson (1974) showed even for some seemingly innocuous $m = 0$ vibrations. Hunter also cautioned of similar complications near edges of low density.

A second difficulty has been that, even if strictly oscillatory discrete $m = 2$ modes do exist, it remains unproven that they include any of *spiral* form. Indeed, such truly permanent spiral waves of infinitesimal amplitude now seem hardly possible. This personal opinion stems much more from the formidable resonance effects and gravitational torques discussed, for instance, by Lynden-Bell & Kalnajs (1972) than from the weak "anti-spiral theorem" of Lynden-Bell & Ostriker (1967)—which in essence states only that the existence of a steady trailing mode implies a twin of leading shape, and vice versa, because of time-reversibility.

The above frustrations have not referred especially to *unstable* spiral modes. (Quite reasonably, it is the milder versions of just such modes from nondissipative linear theory that are widely thought now to offer the best hope for some sort of conversion into *quasi-steady* spiral patterns of finite amplitude—and this is where threshold effects such as the nonlinear shock damping stressed in Section 5.3 not only assume a special importance but may harbor a second vital reason as to why good spirals seem to need gas.) The problem here has been much more that one has tended to find such unstable modes, even discrete ones, far too fiercely and far too often! Just as in the N-body studies, much of this nuisance has doubtless had good physical causes, but also as with those experiments, some may even have stemmed from ingenious tricks that backfired.

One such example comes from a labor-saving idea that occurred to Miller (1971, 1974) among others. He noticed that the use of the "softened" gravitational potential $\phi = -GM(b^2+d^2)^{-1/2}$ rather than the precise $\phi = -GM/d$ for the interaction of two particles separated by a distance d would yield the dispersion relation

$$\omega^2 = \kappa^2(r) - 2\pi G\mu(r)\,|k|\,e^{-|k|b} \tag{12}$$

instead of Equation (1) for short axisymmetric vibrations of a cold thin disk. The exponential mimics the "reduction factor" of Lin & Shu (1966), and so this simple relation indeed resembles the one in Figure 4. Toward the short wavelengths in particular it seems much more representative of the $\omega \to \kappa$ behavior of waves in a stellar disk than would be implied by the analogous formula

$$\omega^2 = \kappa^2(r) - 2\pi G\mu(r)\,|k| + a^2k^2 \tag{13}$$

(cf Safronov 1960, Hunter 1972) for a thin gas layer with "honest" gravity but an imagined sound speed a.

The irony here is that Erickson (1974), using just this presumably superior gravity shortcut (to avoid all the stellar-dynamical bother with the Boltzmann equation in his global-mode calculations for some Gaussian disks) indeed obtained numerous spiral instabilities resembling the $m = 2$ mode in Figure 12. Unfortunately, he also found himself unable to turn off all such growing modes, no matter how large he chose that cutoff length b. By contrast, Bardeen (1975) adopted the seemingly weaker gas idealization and managed to stabilize such disks altogether, at least for energy ratios $T/W \lesssim 0.26$ corresponding to dynamical (rather than secular) stability of Maclaurin spheroids. With these experiences in mind—and also noting that Kalnajs (1970 onwards) has not yet succeeded in obtaining fully stable stellar disks, apart from his *tour de force* with the Maclaurin-Freeman models—it is no wonder that mode calculators remain a little unsure of just where they stand.

To conclude this litany of surprises, some signs of a counterexample to the Ostriker-Peebles criterion have surfaced recently in the work of Zang (1976) on the unstable modes of a class of exact stellar disks with random motions. These disks themselves, which were noticed also by Bisnovatyi-Kogan (1975), epitomize "flat" rotation curves in assuming $V(r) = \text{const} = V_0$ at all radii; this means that the surface density $\mu(r) = V_0^2/2\pi Gr$, or that the total mass grows infinite linearly with the radius r. The latter property, and also the central singularity of μ itself, need not concern us too much, since both can doubtless be remedied with plausible cutoffs. One big advantage of this model is that it is everywhere self-similar. Another is that its velocity distribution function

$$f(u,v,r) = F(E,J) = \text{const}\cdot J^q e^{-E/\sigma_u^2}, \tag{14}$$

with $E = \frac{1}{2}(u^2+v^2) + V_0^2 \ln r$, $J = rv$, and $q+1 = V_0^2/\sigma_u^2$, can be written so compactly for $J > 0$ (with $f = 0$ otherwise).

As far as axisymmetric (or $m = 0$) modes are concerned, Zang found this class of models to become globally stable when $\sigma_u \gtreqless 0.378V_0$, in good agreement with the "local" estimate given by Equation (3). Already for that critical value of the velocity dispersion, however, Zang was unable to locate numerically, despite much systematic searching, any two- or multi-armed (i.e. $m \gtreqless 2$) instabilities. This important result needs to be qualified in two ways. One is that Zang literally examined not the full models given by Equation (14), but ones where he had immobilized the central regions—as if the central "bulge" were particularly hot and therefore unresponsive—by multiplying that f by the factor $[1+(J_0/J)^n]^{-1}$; he did this mostly to avoid immense angular speeds as $r \rightarrow 0$, but also to provide a length scale J_0/V_0. The other proviso is that Zang managed to interrogate his difficult integral equation, patterned closely after the one by Kalnajs (1971), only for *discrete* unstable modes, which he assumed to grow like e^{st} while rotating with an arbitrary pattern speed Ω_p. An unstable continuum seems unlikely, but Zang could provide no proof.

Thus, to be precise, Zang's inability to locate any $m = 2$ instabilities referred only to discrete modes of modified disks with $n = 1$ or 2—or ones whose centers had been "carved out" fairly gently. As that truncation was made more abrupt, some unstable modes actually emerged; one such mode, for index $n = 4$, is pictured in Figure 12. Modes like that were definitely indebted, however, to the artificial *sharpness* of the inner boundary. Although superficially similar to the modes inferred by Lau, Lin & Mark (1976) from WKBJ theory, the attractive but unrealistic $m = 2$ spiral wave in Figure 12 grew an order of magnitude faster than then expected by those authors.

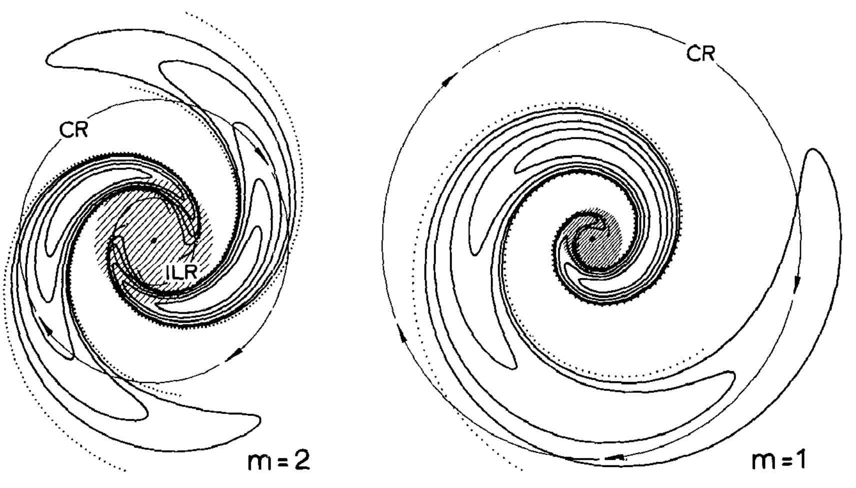

Figure 12 Two most unstable spiral modes obtained by Zang (1976) for centrally cut-out versions of the $V(r) = V_0$ disks. In both cases, only the positive parts of the disturbance density are shown here, using contour lines set at 80, 60, 40, 20, and 10% of peak values; dots mark the nodes. The shaded areas have radius $r = J_0/V_0$. The $m = 2$ mode (referring to the model with velocity dispersion $\sigma_u = 0.378V_0$ and cut-out index $n = 4$) has pattern speed $\Omega_p = 0.439$ and growth rate $s = 0.127$, both expressed as multiples of V_0^2/J_0. The analogous data for the $m = 1$ mode (obtained for $n = 2$) are $\Omega_p = 0.141$ and $s = 0.066$.

Even this story, however, proved to contain a nasty surprise. Regardless of whether he tried cut-out indices $k = 1, 2, \ldots$, Zang remained plagued with numerous *one*-armed unstable modes! One of them is also shown in Figure 12. Unlike the few two-armed modes, this $m = 1$ instability could not be attributed very plausibly either to the frozen center or even to the fact that such a core—in a deliberate rerun of a rare old error by Maxwell—was here kept mostly fixed. Zang concluded ruefully that with this curious swap of difficulties "we may merely have jumped from the frying pan into the fire."

8 SHEARING BITS AND PIECES

As our last topic, it needs to be said briefly but aloud that there seems nothing fundamentally wrong with the simple idea that much of spiral structure—which, after all, is often very ragged and confusing, as in most of M101—consists more nearly of *things* rather than waves, to use Prendergast's (1967) apt word. The difficulties with this notion are mainly just quantitative, once one excludes the "grand designs." Obviously all shearing "things" in a galaxy will briefly exhibit some spirality, like cream poured into a newly-stirred cup of coffee. Hence the real task remains instead to decide how and whether the material in question can keep on making suitable fresh bits and pieces to replace those older fragments that wrap up and disappear.

The only promising site for such *recurrent* instabilities is, of course, the gas layer of a galaxy. Already von Weizsäcker (1951) thought so, with his vague theory of "turbulence . . . produced by nonuniform rotation." Jeans (1929, p. 379) did almost the same when he wrote of the "chaos of moving clusters" in the outer parts of spiral nebulae, and remarked that "the theory of gravitational instability makes it easy to understand how these large clusters come to exist."

In fact, the regeneration of large new fragments in a gas disk is not quite so easy. Von Weizsäcker in effect forgot that our kind of shear flow is supported much more by gravity forces than by pressure gradients, and Jeans did not realize that rotating disks also have a *maximum* critical length for gravitational instability, such as λ_{crit} from Equation (2), which need not be at all comparable to overall dimensions. Indeed that length can be distressingly short if the "active" surface density μ_g of the subsystem of interest is only a small fraction of the total, and if the stellar disk cooperates hardly at all. For example, even if we adopt $\mu_g = 10\ M_\odot \text{pc}^{-2}$ for the nearby gas in this Galaxy, its λ_{crit} works out as only 1.7 kpc—and then an effective sound speed $a \geqq 4\ \text{km sec}^{-1}$ should already curb all Jeans instabilities [cf Equation (13)].

All this was well appreciated by Goldreich & Lynden-Bell (1965) in what remains the only extensive modern exploration of the theme that Sc galaxies consist largely of "a swirling hotch-potch of pieces of spiral arms." To overcome the above difficulties, these authors suggested that "spiral arm formation should not be regarded as an instability in [just] the gas but rather as an instability of the whole star-gas mixture which is triggered by an increase in gas density" in a volumetric sense—or, what amounts to the same thing, by any cooling or damping that lowers the sound speed a in the gas. Even those plans, however, had two serious flaws. One was the fact that the fiercely amplified shearing waves studied by Goldreich and Lynden-Bell were, strictly speaking, no true instabilities. The other flaw was the presumption that strong help from the undamped stars would continue despite all their repeated heating by the postulated clumpings of the gas; this idea rested mostly upon wishful thinking. Nevertheless, in that severe amplification, Goldreich and Lynden-Bell offered one real nugget of a discovery that greatly softens the last criticism. It is shown in a more refined form in Figure 13.

This old diagram reemphasizes that a shearing stellar (or gas) disk that is locally quite stable can still *respond* with remarkable vigor to the nonaxisymmetric gravity forces from any fortuitously bound, orbiting gas lumps. Notice that the excess mass that pauses in the trailing "wakes" illustrated here exceeds the imposed mass by about an order of magnitude, even though $Q = 1.4$ plus a modest softening due to thickness were assumed in this example. Notice especially that this arm-like shape does not undergo any shearing. It is itself a steady density *wave*, albeit a forced one. In short, Figure 13 renews hope that here may be a reason why some of the observed spiral irregularities extend, as in M101, over quite sizable fractions of the radius despite the fact that the gas content tends to be hardly of order 10 per cent. It also cautions that even the "hotch-potch" is apt to be a little more wavelike in character than one might perhaps have supposed.

Any reader who (rightly) finds these remarks unpersuasive should look again at Figure 11 and especially at the experiments of Hohl (1970) and Hockney & Brownrigg (1974), all conducted with active masses totaling only 10 or 20%. He

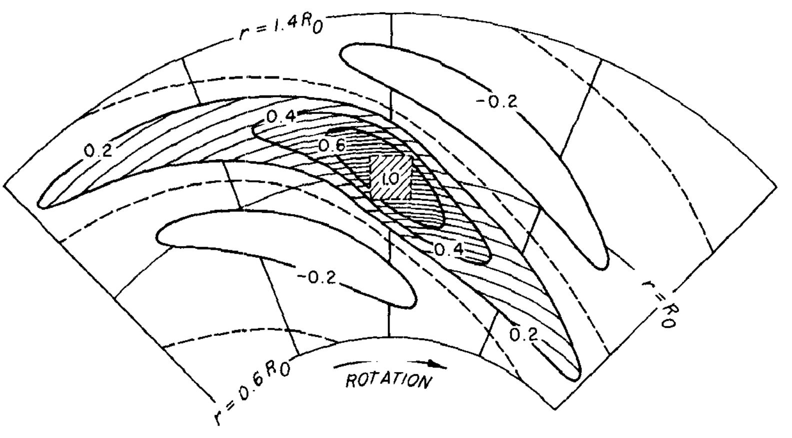

Figure 13 Wavelike ridge of excess density in a $Q = 1.4$, $V(r) = \text{const}$ disk of stars shearing past a small orbiting mass point (Julian & Toomre 1966). By comparison, that mass itself would yield unit density if spread uniformly over the little square. This diagram stems from small-amplitude theory with several "local" approximations; all subsequent overtakings of, and by, stars have been neglected here. The self-gravity of stars was included, however, and the assumption that $\lambda_{crit} = R_0$, made only in plotting, possibly underestimates the true extent of this quasi-steady forced wave.

should ask how he would explain those examples of "material clumping [that] is periodically destroyed by differential rotation and regenerated by gravitational instability" (Lin & Shu 1964), if not by something akin to the amplification process of Goldreich and Lynden-Bell. Of course, the bigger question remains: Why should enough disk material *still* be "cool" enough in its random motion to act even as cooperatively as shown in Figure 13? Are we, for instance, underestimating the amounts of gas present, or else the recent history either of infall of gas from outside or its return from dying stars, including those in any halos?

All in all, this topic of various "secondary" spiral features seems much overdue for renewed attention. Not even von Weizsäcker's idea can yet be dismissed entirely: some instabilities might indeed be largely hydrodynamic. It still needs to be explored, for instance, whether the well-known lumpiness of the major arms themselves on about a 1 or 2 kpc scale might not stem mainly from some sort of Kelvin-Helmholtz instability (rather than Jeans instability, or even the Parker = magnetic Rayleigh-Taylor kind advocated by Shu 1974) of the strong shear flows that are inevitable when gas from a whole range of radii crowds together behind a spiral shock.

Finally, what about spiral structures provoked from outside? A quick answer is that this old tidal idea indeed seems to work fine—in a small handful of obscure cases like Arp 295 (Stockton 1974) and NGC 2535/36 or 3808. In those instances just as in some simple computations (cf Toomre 1974b), the two "material arms" are rather open and the one toward the companion often looks grotesquely straighter and longer than the other. It remains far less certain, however—despite much speculation going back at least to Chamberlin (1901) in the case of M51, and again to Lindblad (1941) with M81/82—that the familiar "grand designs" of M51, M81, and NGC 5364 are indebted in any serious way to the (admittedly worrisome) proximity of major companions.

For reasons of tightness and extent of spiral structure, plus the Westerbork detection of ample H I also *between* the main arms of M81, those famous examples cannot possibly be tidal in the naive kinematic sense implied above. Here the tidal process, if indeed any, must have been a lot subtler. The first step would have had to be some major distortion of the outer disk or halo. This in turn may or may not have induced a relatively transient spiral wave in the proper disk, traveling inward with the group velocity. If only some decently cool but stable models of galaxies existed already, it would be fairly easy to check whether such a scheme makes any sense dynamically. Until then, however, claims such as by Tully (1974) that "a density wave has been generated [in M51] by the innermost material clumping arising from the encounter" with NGC 5195 strike even this sympathetic reviewer as very premature. In fact, judging from the *broad* outer contours of M51 toward the south—especially evident in one very deep photograph and in various H I data that Lynds and Shane have kindly shown—it seems that Toomre & Toomre (1972, Figure 20a) erred in reckoning the present epoch of that encounter to be as late as "$t = 2.4$." A better guess appears to be $t = 1.6$ or 1.8, in the same units. Given that those outer distortions become pronounced only from $t = 1$ onward, this shorter elapsed time gravely aggravates a difficulty noted already by Lin (1975), namely that "the time required for the propagation of such an influence into the central regions is too long." Hence it looks more and more as if even in M51 the main *spiral* structure is in essence just a badly-dented version of what would have been there anyway.

To end on a more positive note, it has emerged recently that density/shock waves of roughly circular shape—now indeed caused by a second galaxy—seem to be present in the rare but very striking oddities known as ring galaxies. As Fosbury & Hawarden (1977) have just illustrated for the fine example of the "Cartwheel" discovered by Zwicky (1941), it seems there again that simple mechanical crushing has evoked intense star formation in the knots of H II regions which mark that ring itself—and also that the *residue* of this indelicate processing includes the usual "feathers" or "interarm branches" or secondary spiral arms!

ACKNOWLEDGMENTS

Among several kind friends, I am most indebted here to Agris Kalnajs—and to editors Burbidge, Goldberg, and Layzer for much patient nudging. I am also very grateful for all the support of the NSF.

Literature Cited

Arp, H. C. 1976. *Ap. J.* 207:L147
Baade, W. 1963. *Evolution of Stars and Galaxies*. Cambridge: Harvard Univ. Press
Bardeen, J. M. 1975. *Proc. IAU Symp.* 69:297
Bisnovatyi-Kogan, G. S. 1975. *Pis'ma Astron. Zh.* 1(9):3. Transl. 1976. *Sov. Astron. Lett.* 1:177 (from Russian)
Bryan, G. H. 1889. *Philos. Trans. R. Soc. London Ser. A* 180:187
Burton, W. B. 1973. *Publ. Astron. Soc. Pac.* 85:679
Carranza, G., Crillon, R., Monnet, G. 1969. *Astron. Astrophys.* 1:479
Chamberlin, T. C. 1901. *Ap. J.* 14:17
Chandrasekhar, S. 1969. *Ellipsoidal Figures of Equilibrium*. New Haven: Yale Univ. Press
Contopoulos, G. 1970. *Ap. J.* 160:113
Contopoulos, G. 1975. In *La Dynamique des Galaxies Spirales*, ed. L. Weliachew, p. 21. Paris: CNRS
Curtis, H. D. 1919. *J. Wash. Acad. Sci.* 9:217
de Vaucouleurs, G. 1964. *Proc. IAU Symp.* 20:195
de Vaucouleurs, G. 1970. *Proc. IAU Symp.* 38:18
Dewar, R. L. 1972. *Ap. J.* 174:301
Dzyuba, B. M., Yakubov, V. B. 1970. *Astron. Zh.* 47:3. Transl. 1970. *Sov. Astron.* 14:1 (from Russian)
Erickson, S. A. 1974. *Vibrations and Instabilities of a Disk Galaxy with Modified Gravity*. PhD thesis. Mass. Inst. Technol., Cambridge, Mass. 179 pp.
Feldman, S. I., Lin, C. C. 1973. *Stud. Appl. Math.* 52:1
Feynman, R. P. 1963. *Lectures on Physics* 1: p. 7-7. Reading, Mass: Addison-Wesley
Fosbury, R. A. E., Hawarden, T. G. 1977. *MNRAS* 178:473
Freeman, K. C. 1966. *MNRAS* 134:1; 134:15
Freeman, K. C. 1970a. *Ap. J.* 160:811
Freeman, K. C. 1970b. *Proc. IAU Symp.* 38:351
Freeman, K. C. 1975. In *Galaxies and the Universe*, ed. A. Sandage, M. Sandage, J. Kristian, p. 409. Chicago: Univ. Chicago Press
Fujimoto, M. 1968. *Proc. IAU Symp.* 29:453
Goldreich, P., Lynden-Bell, D. 1965. *MNRAS* 130:125
Gordon, M. A., Burton, W. B. 1976. *Ap. J.* 208:346
Hockney, R. W., Brownrigg, D. R. K. 1974. *MNRAS* 167:351
Hohl, F. 1970. *Proc. IAU Symp.* 38:368
Hohl, F. 1971. *Ap. J.* 168:343
Hohl, F. 1972. *J. Comp. Phys.* 9:10
Hohl, F. 1973. *Ap. J.* 184:353
Hohl, F. 1975. *Proc. IAU Symp.* 69:349
Hohl, F. 1976. *Astron. J.* 81:30
Hunter, C. 1963. *MNRAS* 126:299
Hunter, C. 1965. *MNRAS* 129:321
Hunter, C. 1969. *Stud. Appl. Math.* 48:55
Hunter, C. 1972. *Ann. Rev. Fluid Mech.* 4:219

Jeans, J. H. 1929. *Astronomy and Cosmogony*. Cambridge: Univ. Press; 1961. New York: Dover
Julian, W. H. 1967. *Ap. J.* 148:175
Julian, W. H., Toomre, A. 1966. *Ap. J.* 146: 810
Kalnajs, A. J. 1965. *The Stability of Highly Flattened Galaxies*. PhD thesis. Harvard Univ., Cambridge, Mass. 129 pp.
Kalnajs, A. J. 1970. *Proc. IAU Symp.* 38:318
Kalnajs, A. J. 1971. *Ap. J.* 166:275
Kalnajs, A. J. 1972a. *Astrophys. Lett.* 11:41
Kalnajs, A. J. 1972b. *Ap. J.* 175:63
Kalnajs, A. J. 1973. *Proc. Astron. Soc. Aust.* 2:174
Kalnajs, A. J. 1976. *Ap. J.* 205:745; 205:751
Kalnajs, A. J. 1977. *Ap. J.* 212:637
Kaplan, S. A., Pikel'ner, S. B. 1974. *Ann. Rev. Astron. Astrophys.* 12:113
Kerr, F. J. 1967. *Proc. IAU Symp.* 31:239
Krumm, N., Salpeter, E. E. 1977. *Astron. Astrophys.* 56:465
Lau, Y. Y., Lin, C. C., Mark, J. W.-K. 1976. *Proc. Natl. Acad. Sci. USA* 73:1379
Lau, Y. Y., Mark, J. W.-K. 1976. *Proc. Natl. Acad. Sci. USA* 73:3785
Lin, C. C. 1967. *Ann. Rev. Astron. Astrophys.* 5:453
Lin, C. C. 1970. *Proc. IAU Symp.* 38:377
Lin, C. C. 1971. In *Highlights of Astronomy*, ed. C. de Jager, 2:88. Dordrecht: Reidel
Lin, C. C. 1975. In *Structure and Evolution of Galaxies*, ed. G. Setti, p. 119. Dordrecht: Reidel
Lin, C. C., Shu, F. H. 1964. *Ap. J.* 140:646
Lin, C. C., Shu, F. H. 1966. *Proc. Natl. Acad. Sci. USA* 55:229
Lin, C. C., Shu, F. H. 1967. *Proc. IAU Symp.* 31:313
Lin, C. C., Yuan, C., Shu, F. H. 1969. *Ap. J.* 155:721
Lindblad, B. 1941. *Stockholm Obs. Ann.* 13: No. 10
Lindblad, B. 1948. *MNRAS* 108:214
Lindblad, B. 1951. *Publ. Obs. Univ. Mich.* 10:59
Lindblad, B. 1958. *Stockholm Obs. Ann.* 20: No. 6
Lindblad, B. 1961. *Stockholm Obs. Ann.* 21: No. 8
Lindblad, B. 1962. *Proc. IAU Symp.* 15:146
Lindblad, B. 1963. *Stockholm Obs. Ann.* 22: No. 5
Lindblad, B. 1964. *Astrophys. Norv.* 9:103
Lindblad, P. O. 1960. *Stockholm Obs. Ann.* 21: No. 4
Lindblad, P. O. 1974. *Proc. IAU Symp.* 58: 399
Lynden-Bell, D. 1974. In *Proc. Eur. Astron. Meeting, 1st, Athens*, 3:114. Berlin: Springer
Lynden-Bell, D. 1975. See Contopoulos 1975, p. 91
Lynden-Bell, D., Kalnajs, A. J. 1972. *MNRAS* 157:1
Lynden-Bell, D., Ostriker, J. P. 1967. *MNRAS* 136:293
Mark, J. W.-K. 1971. *Proc. Natl. Acad. Sci. USA* 68:2095
Mark, J. W.-K. 1974a. *Ap. J.* 193:539
Mark, J. W.-K. 1974b. *Proc. IAU Symp.* 58: 417
Mark, J. W.-K. 1976a. *Ap. J.* 203:81
Mark, J. W.-K. 1976b. *Ap. J.* 205:363
Mark, J. W.-K. 1976c. *Ap. J.* 206:418
Marochnik, L. S., Mishurov, Yu. N., Suchkov, A. A. 1972. *Astrophys. Space Sci.* 19:285
Marochnik, L. S., Suchkov, A. A. 1974. *Usp. Fiz. Nauk* 112:275. Transl. 1974. *Sov. Phys. Usp.* 17:85 (from Russian)
Mathewson, D. S., van der Kruit, P. C., Brouw, W. N. 1972. *Astron. Astrophys.* 17:468
Maxwell, J. C. 1859. In 1890. *Scientific Papers*, 1:288. Cambridge: Univ. Press
Miller, R. H. 1971. *Astrophys. Space Sci.* 14: 73
Miller, R. H. 1974. *Ap. J.* 190:539
Miller, R. H. 1975. See Contopoulos 1975, p. 45
Miller, R. H., Prendergast, K. H., Quirk, W. J. 1970. *Ap. J.* 161:903
Oort, J. H. 1956. *Sci. Am.* 195(3):100
Oort, J. H. 1962. In *Interstellar Matter in Galaxies*, ed. L. Woltjer, p. 234. New York: Benjamin
Oort, J. H. 1966. *Q. J. R. Astron. Soc.* 7:329
Ostriker, J. P., Peebles, P. J. E. 1973. *Ap. J.* 186:467
Piddington, J. H. 1973. *Ap. J.* 179:755
Pikel'ner, S. B. 1970. *Astron. Zh.* 47:752. Transl. 1971. *Sov. Astron.* 14:602 (from Russian)
Pooley, G. G. 1969. *MNRAS* 144:101
Prendergast, K. H. 1962. See Oort 1962, p. 217
Prendergast, K. H. 1967. *Proc. IAU Symp.* 31:303
Quirk, W. J. 1971. *Ap. J.* 167:7
Roberts, M. S. 1975. *Proc. IAU Symp.* 69:331
Roberts, W. W. 1969. *Ap. J.* 158:123
Roberts, W. W. 1970. *Proc. IAU Symp.* 38: 415
Roberts, W. W., Roberts, M. S., Shu, F. H. 1975. *Ap. J.* 196:381
Roberts, W. W., Shu, F. H. 1972. *Astrophys. Lett.* 12:49
Rogstad, D. H., Wright, M. C. H., Lockhart, I. A. 1976. *Ap. J.* 204:703
Rosse, Earl of. 1850. *Philos. Trans. R. Soc. London*, p. 503

Rots, A. H. 1975. *Astron. Astrophys.* 45:43
Rots, A. H., Shane, W. W. 1975. *Astron. Astrophys.* 45:25
Rybicki, G. B. 1971. *Astrophys. Space Sci.* 14:15
Safronov, V. S. 1960. *Dokl. Akad. Nauk* 130:53. Transl. 1960. *Sov. Phys. Dokl.* 5:13 (from Russian)
Salpeter, E. E. 1976. *Ap. J.* 206:673
Sandage, A. 1975. See Freeman 1975, p. 1
Sanders, R. H., Huntley, J. M. 1976. *Ap. J.* 209:53
Schweizer, F. 1976. *Ap. J. Suppl.* 31:313
Sharpless, S., Franz, O. G. 1963. *Publ. Astron. Soc. Pac.* 75:219
Shu, F. H. 1970. *Ap. J.* 160:99
Shu, F. H. 1973. *Am. Sci.* 61:524
Shu, F. H. 1974. *Astron. Astrophys.* 33:55
Shu, F. H., Milione, V., Gebel, W., Yuan, C., Goldsmith, D. W., Roberts, W. W. 1972. *Ap. J.* 173:557
Shu, F. H., Milione, V., Roberts, W. W. 1973. *Ap. J.* 183:819
Shu, F. H., Stachnik, R. V., Yost, J. C. 1971. *Ap. J.* 166:465
Simonson, S. C. 1970. *Astron. Astrophys.* 9:163
Sørensen, S.-A., Matsuda, T., Fujimoto, M. 1976. *Astrophys. Space Sci.* 43:491
Spitzer, L., Schwarzschild, M. 1953. *Ap. J.* 118:106
Stockton, A. N. 1974. *Ap. J.* 190:L47
Sweet, P. A. 1963. *MNRAS* 125:285
Toomre, A. 1964. *Ap. J.* 139:1217
Toomre, A. 1969. *Ap. J.* 158:899
Toomre, A. 1974a. In *Highlights of Astronomy*, ed. G. Contopoulos, p. 457. Dordrecht: Reidel
Toomre, A. 1974b. *Proc. IAU Symp.* 58:347
Toomre, A., Toomre, J. 1972. *Ap. J.* 178:623
Tully, R. B. 1974. *Ap. J. Suppl.* 27:449
van der Kruit, P. C., Allen, R. J. 1976. *Ann. Rev. Astron. Astrophys.* 14:417
Vandervoort, P. O. 1973. *Ap. J.* 180:739
van Kampen, N. G. 1955. *Physica* 21:949
von Weizsäcker, C. F. 1951. *Ap. J.* 114:165
Wielen, R. 1974. *Publ. Astron. Soc. Pac.* 86:341
Wielen, R. 1975. See Contopoulos 1975, p. 357
Woodward, P. R. 1975. *Ap. J.* 195:61
Woolley, R., Candy, M. P. 1968. *MNRAS* 139:231; 141:277
Yuan, C. 1969. *Ap. J.* 158:889
Zang, T. A. 1976. *The Stability of a Model Galaxy.* PhD thesis. Mass. Inst. Technol., Cambridge, Mass. 204 pp.
Zwicky, F. 1941. *Applied Mechanics* von Kármán Anniv.: 137
Zwicky, F. 1955. *Publ. Astron. Soc. Pac.* 67:232

CONTENTS

INDEXES