

## DYNAMICS OF DISKS AND WARPS

J. A. SELLWOOD

Rutgers University, Department of Physics & Astronomy,  
136 Frelinghuysen Road, Piscataway, NJ 08854-8019  
*sellwood@physics.rutgers.edu*

*Chapter for Planets, Stars and Stellar Systems v5. Printed on July 30, 2010*

### ABSTRACT

This chapter reviews theoretical work on the stellar dynamics of galaxy disks. All the known collective global instabilities are identified, and their mechanisms described in terms of local wave mechanics. A detailed discussion of warps and other bending waves is also given. The structure of bars in galaxies, and their effect on galaxy evolution, is now reasonably well understood, but there is still no convincing explanation for their origin and frequency. Spiral patterns have long presented a special challenge, and ideas and recent developments are reviewed. Other topics include scattering of disk stars and the survival of thin disks.

*Subject headings:* galaxies: evolution – galaxies: haloes – galaxies: kinematics and dynamics – galaxies: spiral

### 1. INTRODUCTION

A significant fraction of the stars in the universe resides in the rotationally supported disks of galaxies. Disks are mostly thin and flat, but the disk is often warped away from its principal plane in the outer parts. Disk galaxies usually manifest spiral patterns, and rather more than half host bars. Most, but not all, disk galaxies have a central bulge, perhaps also a thick stellar disk, and generally a small fraction of the stars resides in a quasi-spherical stellar halo, while the central attraction at large distances from the center is dominated by a dark halo. The material in most disks overwhelmingly orbits in a single sense, although a small fraction of galaxies have been found to host substantial counter-rotating components.

This chapter is primarily concerned with the dynamics of rotationally supported disks of stars. Stellar disks are believed to have formed over time from gas that had previously settled into centrifugal balance in the gravitational well of the galaxy, and the process of star formation continues to the present day in most disk galaxies. While stars are the dominant dynamical component today, the small gas fraction (usually  $\lesssim 10\%$  by mass) can still play an important dynamical role in some contexts.

Disk dynamics is a rich topic for two principal reasons: (a) the organized orbital motion facilitates gravitationally-driven collective behavior and (b) outward transfer of angular momentum extracts energy from the potential well. Space limitations preclude a detailed development and this review will mostly be confined to a summary of the principal results and open issues. The derivations of the principal formulae can be found in the excellent textbook by Binney & Tremaine (2008, hereafter BT08). Furthermore, no attempt is made to cite every paper that relates to a topic.

Closely related topics are described in other chapters in this volume: the structure of the disks of the Milky Way (Freeman), the formation and structure of bulges (Kormendy), the atomic and molecular gaseous components of the Milky Way (Dickey & Bania respectively), and our current understanding of the processes that lead to the formation of galaxies (Steinmetz). The distributions of

light and mass within galaxies are described in volume 6 (Gallagher).

### 2. PRELIMINARIES

#### 2.1. *Relaxation Rate*

Because stellar disks contain many stars, the attraction from individual nearby stars is negligible in comparison with the aggregated gravitational field of distant matter. The Appendix explains how the usual rough calculation to support this assertion, which was derived with quasi-spherical systems supported by random motion in mind (*e.g.* BT08 pp. 35-38), must be revised for disks. Three factors all conspire to reduce the relaxation time in disks by several orders of magnitude below the traditional estimate (eq. A4), although it remains many dynamical times.

Fig. 1 shows the non-smooth distribution of stellar velocities of  $> 14\,000$  F & G dwarf stars in the vicinity of the Sun, as found in the Geneva-Copenhagen Survey (Nordström *et al.* 2004; Holmberg *et al.* 2009, hereafter GCS). The determination of the radial velocities has confirmed the substructure that was first identified by Dehnen (1998) from a clever analysis of the HIPPARCOS data without the radial velocities (The implications of this Figure are discussed more fully in §10.) As collisional relaxation would erase substructure, this distribution provides a direct illustration of the collisionless nature of the solar neighborhood (unless the substructure is being recreated rapidly, *e.g.* De Simone *et al.* 2004).

The usual first approximation that stars move in a smooth gravitational potential well therefore seems adequate.

#### 2.2. *Mathematical Formulation*

This approximation immediately removes the need to distinguish stars by their masses. A stellar system can therefore be described by a **distribution function** (DF),  $f(\mathbf{x}, \mathbf{v}, t)$  that specifies the stellar density in a 6D phase space of position  $\mathbf{x}$  and velocity  $\mathbf{v}$  at a particular time  $t$ . Since masses are unimportant, it is simplest conceptually to think of the stars being broken into infinitesimal fragments so that discreteness is never an issue.

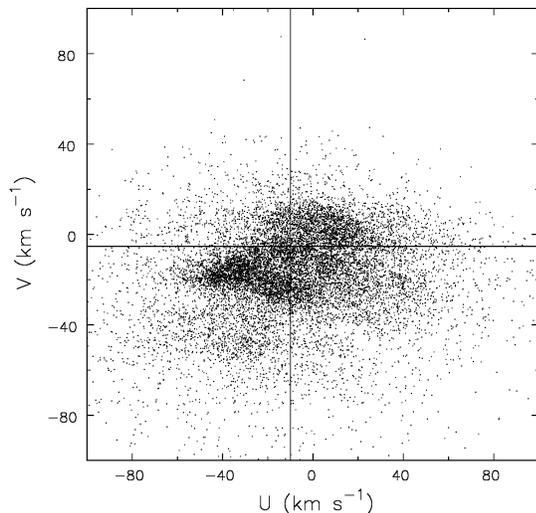


FIG. 1.— The velocity distribution, in Galactic coordinates, of stars near the Sun in the GCS.  $U$  is the velocity of the star toward the Galactic center and  $V$  is the component in the direction of Galactic rotation, both measured relative to the Sun. The intersection of the vertical and horizontal lines shows the local standard of rest estimated by Aumer & Binney (2009).

With this definition, the mass density at any point is

$$\rho(\mathbf{x}, t) = \int f d^3\mathbf{v}, \quad (1)$$

which in turn is related to the gravitational potential,  $\Phi$ , through **Poisson's equation**

$$\nabla^2\Phi(\mathbf{x}, t) = 4\pi G\rho(\mathbf{x}, t). \quad (2)$$

This is, of course, just the potential from the stellar component described in  $f$ ; the total potential includes contributions from dark matter, gas, external perturbations, *etc.* Finally, the evolution of the DF is governed by the **collisionless Boltzmann equation** (BT08 eq. 4.11):

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{x}} + \dot{\mathbf{v}} \cdot \frac{\partial f}{\partial \mathbf{v}} = 0, \quad (3)$$

where the acceleration is the negative gradient of the smooth total potential:  $\dot{\mathbf{v}} = -\nabla\Phi_{\text{tot}}$ . The time evolution of a stellar system is completely described by the solution to these three coupled equations. Note that collisionless systems have no equation of state that relates the density to quantities such as pressure.

The most successful way to obtain global solutions to these coupled equations is through  **$N$ -body simulation**. The particles in a simulation are a representative sample of points in phase space whose motion is advanced in time in the gravitational field. At each step, the field is determined from a smoothed estimate of the density distribution derived from the instantaneous positions of the particles themselves.<sup>1</sup> This rough and ready approach is powerful, but simulations have limitations caused by **noise** from the finite number of particles, and **bias** caused by the smoothed density and approximate solution for the field, and other possible artifacts.

Understanding the results from simulations, or even knowing when they can be trusted, requires dynamical insight that can be obtained only from analytic treatments.

<sup>1</sup> *i.e.* a simulation solves eqs. (1)–(3) by the method of characteristics.

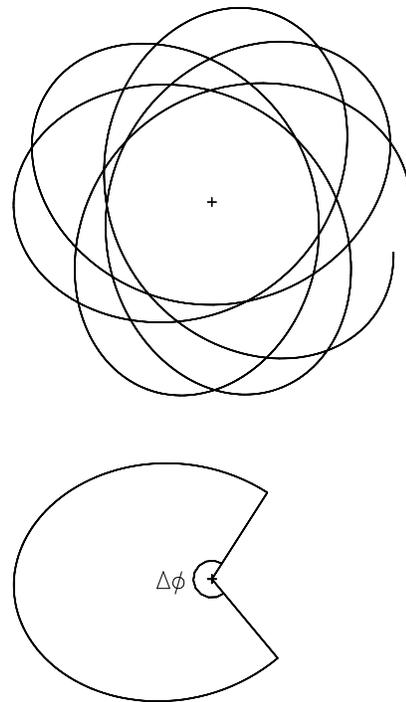


FIG. 2.— An orbit of a star is generally a non-closing rosette (upper panel). The lower panel shows one radial oscillation, from pericenter to pericenter say, during which time the star moves in azimuth through the angle  $\Delta\phi$ .

This chapter therefore stresses how the basic theory of stellar disks inter-relates with well designed, idealized simulations to advance our understanding of these complex systems.

### 2.3. Orbits

Since orbital deflections by mass clumps can be neglected (§2.1), the stars in a disk move in a smooth gravitational potential. It is simplest to first discuss motion of stars in the disk mid-plane, before considering full 3D motion.

The orbit of a star of **specific energy**  $E$  and **specific angular momentum**  $L_z$  in the mid-plane of an axisymmetric potential is, in general, a non-closing rosette, as shown in Fig. 2. The motion can be viewed as a **retrograde epicycle** about a **guiding center** that itself moves at a constant rate around a circle of radius  $R_g$ , which is the radius of a circular orbit having the same  $L_z$ . In one complete radial period,  $\tau_R$  the star advances in azimuth through an angle  $\Delta\phi$ , as drawn in the lower panel. In fact,  $\pi \leq \Delta\phi \leq 2\pi$  in most reasonable gravitational potentials; specifically  $\Delta\phi = \sqrt{2}\pi$  for small epicycles in a flat rotation curve.

These periods can be used to define two angular frequencies for the orbit:  $\Omega_\phi = \Delta\phi/\tau_R$ , which is the angular rate of motion of the guiding center, and  $\Omega_R = 2\pi/\tau_R$ . In the limit of the radial oscillation amplitude,  $a \rightarrow 0$ , these frequencies tend to  $\Omega_\phi \rightarrow \Omega$ , the angular frequency of a circular orbit at  $R = R_g$ , and  $\Omega_R \rightarrow \kappa$ , the **epicyclic frequency**, defined through

$$\kappa^2(R_g) = \left( R \frac{d\Omega^2}{dR} + 4\Omega^2 \right)_{R_g}. \quad (4)$$

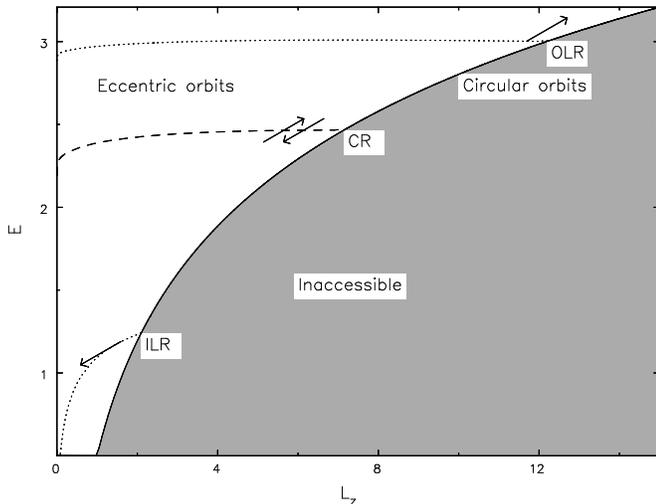


FIG. 3.— The Lindblad diagram for a disk galaxy model. Circular orbits lie along the full-drawn curve and eccentric orbits fill the region above it. Angular momentum and energy exchanges between a steadily rotating disturbance and particles move them along lines of slope  $\Omega_p$  as shown. The dotted and dashed lines are the loci of resonances for an  $m = 2$  perturbation of arbitrary pattern speed.

#### 2.4. Resonances

If the potential includes an infinitesimal non-axisymmetric perturbation having  $m$ -fold rotational symmetry and which turns at the angular rate  $\Omega_p$ , the **pattern speed**, stars in the disk encounter wave crests at the Doppler-shifted frequency  $m|\Omega_p - \Omega_\phi|$ . Resonances arise when

$$m(\Omega_p - \Omega_\phi) = l\Omega_R, \quad (5)$$

where simple orbit perturbation theory breaks down for steady potential perturbations. At **corotation** (CR), where  $l = 0$ , the guiding center of the star's orbit has the same angular frequency as the wave. At the **inner Lindblad resonance** (ILR), where  $l = -1$ , or at the **outer Lindblad resonance** (OLR), where  $l = +1$ , the guiding center respectively overtakes, or is overtaken by, the wave at the star's unforced radial frequency. Other resonances arise for larger  $|l|$ , but these three are the most important. Note that resonant orbits close after  $m$  radial oscillations and  $l$  turns about the galactic center in a frame that rotates at angular rate  $\Omega_p$ .

The solid curve in Fig. 3, the so-called Lindblad diagram, marks the locus of circular orbits for stars in an axisymmetric disk having a flat rotation curve. Stars having more energy for their angular momentum pursue non-circular orbits. The broken curves show the loci of the principal resonances for some bi-symmetric wave having an arbitrarily chosen pattern speed; thus resonant orbits, that close in the rotating frame, persist even to high eccentricities.

#### 2.5. Wave-Particle Scattering

The arrows in Figure 3 indicate how these classical integrals are changed for stars that are scattered by a steadily rotating mild potential perturbation. Since **Jacobi's invariant**,

$$I_J \equiv E - \Omega_p L_z \quad (6)$$

(BT08 eq. 3.112), is conserved in axes that rotate with the perturbation, the slope of all scattering vectors in Fig. 3 is  $\Delta E/\Delta L_z = \Omega_p$ .

Lynden-Bell & Kalnajs (1972) showed that a steadily rotating potential perturbation causes lasting changes to the integrals only at resonances. Since  $\Omega_c = \Omega_p$  at CR, scattering vectors are parallel to the circular orbit curve at this resonance, where angular momentum changes do not alter the energy of random motion. Outward transfer of angular momentum involving exchanges at the Lindblad resonances, on the other hand, does move stars away from the circular orbit curve, and enables energy to be extracted from the potential well and converted to increased random motion.

Notice also from Fig. 3 that the direction of the scattering vector closely follows the resonant locus (dotted curve) at the ILR only. Thus, when stars are scattered at this resonance, they stay on resonance as they gain random energy, allowing strong scatterings to occur. The opposite case arises at the OLR, where a small gain of angular momentum moves a star off resonance.

#### 2.6. Vertical Motion

Stars also oscillate vertically about the disk mid-plane. If the vertical excursions are not large, and the in-plane motion is nearly circular, the vertical motion is harmonic with the characteristic frequency  $\nu = (\partial^2 \Phi / \partial z^2)^{1/2}$ , which principally depends on the density of matter in the disk mid-plane at  $R_g$ . As  $\nu$  is significantly larger than  $\kappa$ , the frequency of radial oscillation, the two motions can be considered to be decoupled for small amplitude excursions in both coordinates. In practice, the vertical excursions of most stars are large enough to take them to distances from the mid-plane where the potential departs significantly from harmonic (*i.e.* for  $|z| \gtrsim 200$  pc in the solar neighborhood, Holmberg & Flynn 2004).

The principal **vertical resonances** with a rotating density wave occur where  $m(\Omega_p - \Omega_\phi) = \pm\nu$ , which are at radii farther from corotation than are the in-plane Lindblad resonances. Since spiral density waves are believed to have substantial amplitude only between the Lindblad resonances, the vertical resonances do not affect density waves, and anyway occur where the forcing amplitude will be small.

The situation is more complicated when the star's radial amplitude is large enough that  $\nu$  varies significantly between peri- and apo-galacticon. Also, the vertical periods of stars lengthen for larger vertical excursions, allowing vertical resonances to become important (see §8.1).

### 3. LOCAL THEORY OF DENSITY WAVES

#### 3.1. Plane Waves in a Thin Mass Sheet

Since solutions of Poisson's equation can be obtained analytically only for the simplest mass distributions, theoretical treatments necessarily make quite drastic simplifying assumptions. Local theory, summarized in this section, is built around a WKB-type potential solution for density variations in the form of short wavelength plane waves in a locally uniform, razor thin disk. In this case, a plane-wave disturbance  $\Sigma_1 e^{ikx}$  to the surface density in the  $(x, y)$ -plane at some instant gives rise to the potential

$$\Phi_1(\mathbf{x}) = -\frac{2\pi G}{|k|} \Sigma_1 e^{ikx} e^{-|kz|}, \quad (7)$$

(cf. eq. 6.31 of BT08). This relation strictly applies only to straight waves of infinite extent. However, the sinusoidal density variations cause the field of distant waves to cancel quickly, and the formula is reasonably accurate near the center of a short wave packet. Note that eq. (7) does not depend on the inclination of the wavefronts to the radial direction, but the curvature of the wavefronts must also be negligible, which requires  $|kR| \gg 1$ , with  $R$  being the distance from the disk center.

### 3.2. Dispersion Relations

A dispersion relation gives the relationship between the wavenumber  $k$  and frequency  $\omega$  of self-consistent waves. Assuming (a) the above relation between density and potential for a razor-thin disk and (b) the wave is of small amplitude, so that second order terms are negligible, the local dispersion relation for short axisymmetric density waves may be written

$$\omega^2 = \kappa^2 - 2\pi G \Sigma |k| \mathcal{F}(s, \chi) \quad (8)$$

(Kalnajs 1965), where  $\Sigma$  is the local undisturbed surface density and the complicated **reduction factor**  $\mathcal{F}$ , with two dimensionless arguments, is explained below. Lin & Shu (1966) independently derived a generalized relation for *tightly-wrapped* spiral waves, which required two additional assumptions: (c) that the wave vector  $\mathbf{k}$  is approximately radial, and (d) that the disturbance is not close to any of the principal resonances. Their better-known relation simply replaces the frequency of the wave with the Doppler-shifted forcing frequency experienced by the orbit guiding center:

$$(\omega - m\Omega)^2 = \kappa^2 - 2\pi G \Sigma |k| \mathcal{F}(s, \chi) \quad (9)$$

(see also BT08 §6.2.2), where now  $\omega \equiv m\Omega_p$ .

The reduction factor,  $\mathcal{F}$ , is always positive and is unity when the disk has no random motion. Since all the factors in the self-gravity term on the RHS are intrinsically positive, eqs. (8) & (9) say that self-gravity enables a supporting response from the stars at a frequency that is lower than their free epicycle frequency  $\kappa$ .

When the disk has random motion, the vigor of the stellar response depends on two factors: the ratio of the forcing frequency experienced by a star to its natural frequency,  $s \equiv |\omega - m\Omega|/\kappa$ , and the ratio of the typical sizes of the stellar epicycles ( $\propto \langle v_R^2 \rangle^{1/2}/\kappa$ ) to the wavelength of the wave ( $\propto |k|^{-1}$ ). Thus for a Gaussian distribution of radial velocities,  $\mathcal{F}$  is a function of both  $s$  and  $\chi \equiv \sigma_R^2 k^2 / \kappa^2$ . Clearly, when  $\chi$  is large, most stars will oscillate over many radial waves, and the weakened supporting response arises mainly from the small fraction of stars near the center of the velocity distribution.

The dispersion relation for barotropic fluid (gaseous) disks<sup>2</sup> is similar (BT08 eq. 6.55)

$$(\omega - m\Omega)^2 = \kappa^2 - 2\pi G \Sigma |k| + v_s^2 k^2, \quad (10)$$

where  $v_s$  is the sound speed in the gas. Both dispersion relations are shown graphically in Fig. 4, below.

<sup>2</sup> Goldreich & Lynden-Bell (1965a) derived the vertically-integrated equations for “2D pressure” in a gas disk.

### 3.3. Axisymmetric Stability

Toomre (1964) determined the condition for marginal stability of short axisymmetric waves by solving for  $\omega^2 = 0$  in eq. (8). Since all the factors in the self-gravity term are intrinsically positive,  $\omega^2$  could be negative for large  $|k|$ , implying instability. Without random motion,  $\mathcal{F} = 1$  and short waves with  $k > k_{\text{crit}}$  or  $\lambda < \lambda_{\text{crit}} = 4\pi^2 G \Sigma / \kappa^2$ , will be unstable. Unlike for the Jeans instability in a stationary medium, longer waves are stabilized by rotation, embodied in the  $\kappa^2$  term.

For disks with random motion, the reduction factor  $\mathcal{F} \rightarrow 0$  for large  $|k|$ , irrespective of the frequency, for the reason given above. Thus random motion stabilizes short waves, as for the Jeans instability. Toomre (1964) showed for a Gaussian distribution of radial velocities that the RHS of eq. (8) is  $\geq 0$  for all  $|k|$  when  $\sigma_R \geq \sigma_{R,\text{crit}}$ , where

$$\sigma_{R,\text{crit}} \simeq \frac{3.36 G \Sigma}{\kappa}. \quad (11)$$

Thus **Toomre’s local axisymmetric stability criterion** is  $Q \equiv \sigma_R / \sigma_{R,\text{crit}} \geq 1$ . When reasonably constant with radius, the locally estimated  $Q$  value is a good indicator of global axisymmetric stability. For example, the local values in the equilibrium models proposed by Kalnajs (1976) are in reasonable agreement with those he derived from global axisymmetric modes (his Fig. 2). It should be noted that an axisymmetrically stable disk, *i.e.* for which  $Q \geq 1$  everywhere, could still be unstable to non-axisymmetric modes.

The local axisymmetric stability criterion for rotating gas disks is similar. The longest unstable wavelength,  $\lambda_{\text{crit}}$  in a cold ( $v_s = 0$ ) disk is the same as for a stellar disk. The minimum of the quadratic expression in  $k$  on the RHS of eq. (10) occurs for  $k = \pi G \Sigma / v_s^2$ , and the condition of marginal stability ( $\omega^2 = 0$ ) at this minimum is readily solved, yielding  $v_{s,\text{crit}} = \pi G \Sigma / \kappa$ .

Vandervoort (1970), Romeo (1992), and others have extended the discussion of axisymmetric stability to disks of finite thickness. The principal difference is the dilution of the self-gravity term caused by spreading the disk matter in the vertical direction, resulting in a somewhat more stable disk. The precise correction depends on the assumed vertical structure of both the disk and the wave, but is minor when the characteristic disk thickness  $z_0 \ll \lambda_{\text{crit}}$ , which is usually the case.

More realistic composite disks consist of multiple stellar populations having a range of velocity dispersions, as well as a cool gas component. Jog & Solomon (1992) considered the stability of two gravitationally coupled gas disks having different sound speeds, but Rafikov (2001) improved their analysis to include multiple stellar components. These analyses show that while the gas component may contain a small fraction (typically  $\lesssim 10\%$ ) of the total mass, its low sound speed causes it to have a disproportionate effect on the overall stability.

### 3.4. Self-consistent Density Waves

Fig. 4 gives, for four different values of  $Q$ , a graphical representation of the stellar (solid curves),<sup>3</sup> and gaseous

<sup>3</sup> The reduction factor used assumes a Gaussian distribution of radial velocities

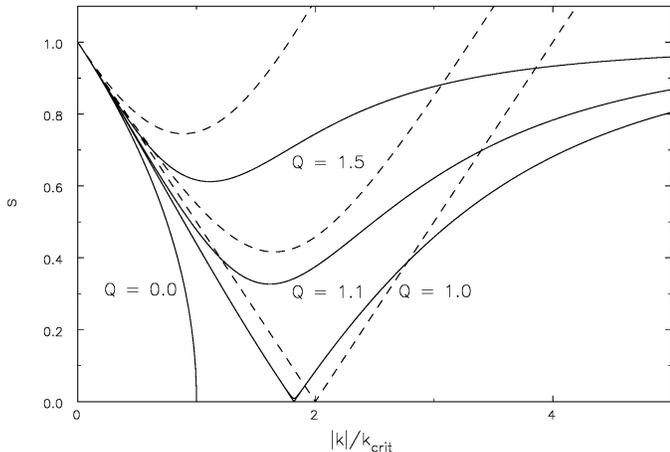


FIG. 4.— The solid lines show the dispersion relation (9), while the dashed lines are from (10), each for several values of  $Q$ . The ordinate  $s = |\omega - m\Omega|/\kappa$ . Since the signs of both the abscissa and the ordinate are suppressed, the graph is reflection symmetric about both axes. Note that  $s = 0$  at CR and  $s = 1$  at the Lindblad resonances.

(dashed curves) dispersion relations eqs. (9) & (10) respectively. The curves do not differ for  $Q = 0$  for which  $s$  is imaginary when  $k > k_{\text{crit}}$ . The dashed and solid curves have similar forms for moderate  $k$  but differ substantially for large  $k$ , where frequencies in the stellar sheet cannot exceed  $|\omega - m\Omega| = \kappa$ . Thus waves in a stellar sheet extend only from CR, where  $s = 0$ , to the Lindblad resonance on either side, where  $s = 1$ .<sup>4</sup>

A gas layer, on the other hand, can support gravitationally modified sound waves of arbitrarily high frequency. Note that this important difference from the stellar case can lead to artificial modes when hydrodynamic equations are used as a simplifying approximation for a stellar disk; while meaningful results can be obtained using this approximation, it is important to check that the derived modes do not cross Lindblad resonances.

Since the dispersion relation (9) does not depend on the sign of  $k$ , curves for negative  $k$ , or leading waves, are simple reflections about the  $k = 0$  axis. Furthermore, the relation gives the square of  $(\omega - m\Omega)$ , implying that solutions when this frequency is negative, which arise inside CR, are simple reflections about the  $s = 0$  axis. Thus were the signs not suppressed, Fig. 4 shows only the panel for trailing waves outside CR, while the other quadrants would be reflected images.

In the quadrant shown, there are either two values of  $k$  for each value of  $s$ , known as the **short-** and **long-wave branches**, or there are none. Only for the marginally stable case of  $Q = 1$  do solutions exist for all frequencies  $0 \leq s \leq 1$  and all  $|k|$ . When  $Q > 1$ , a **forbidden region** opens up in the vicinity of CR where waves are evanescent.

Recall that the relation (9) was derived assuming both  $|kR| \gg 1$  and that the waves are tightly wrapped – *i.e.* that the wave vector is directed radially. These approximations may not be too bad on the short-wave branch (large  $k$ ) but long-wave branch solutions may not exist at all, except in a very low-mass disks where  $k_{\text{crit}}$  is large, and *must* fail as  $|k| \rightarrow 0$ . Lynden-Bell & Kalnajs (1972) gave an im-

proved dispersion relation (their eq. A11) that relaxed the requirement that the wave vector be radial, but still used the plane wave potential (7) that requires  $|kR| \gg 1$ . Their relation is no different for tightly-wrapped waves, but has no long-wave branch or forbidden region for open waves when  $Q \gtrsim 1$ .

### 3.5. Group Velocity

The waves in a disk described by the local dispersion relations (9) & (10) have a phase speed, equal to the pattern speed, in the azimuthal direction. However, Toomre (1969) pointed out that these dispersive waves also have a group velocity that is directed radially. Since  $v_g = \partial\omega/\partial k$ , the group velocity is proportional to the slope of the lines in Fig. 4. Each portion of a wavetrain, or wave packet, is carried radially at the group velocity, retaining its pattern speed  $\omega$  whilst gradually adjusting its wavelength, and also transporting **wave action** or angular momentum.

For trailing waves outside CR, the situation illustrated in Fig. 4, the group velocity is positive on the short-wave branch, and the waves carry angular momentum outwards towards the OLR. The curves have the opposite slope inside CR, where density waves are disturbances of negative angular momentum (stars there lose angular momentum as the wave is set up, Lynden-Bell & Kalnajs 1972), and therefore the inward group velocity on the short-wave branch leads once more to outward angular momentum transport. Outward angular momentum transport is in the sense expected from the gravitational stresses of a trailing spiral (Lynden-Bell & Kalnajs 1972).

However, the sign of the slope  $\partial\omega/\partial k$  on the dubious long-wave branch is opposite to that of short waves, indicating that angular momentum in that regime is transported in the direction opposite to the gravity torque! This apparent contradiction is resolved by the phenomenon of **lorry transport**, a term coined by Lynden-Bell & Kalnajs (1972) to describe an advective Reynolds stress (see also Appendix J of BT08). Stars in their epicyclic oscillations gain angular momentum from the wave near apo-center and lose it back to the wave near peri-center; no star suffers a net change, but angular momentum is still carried inward at a sufficient rate to overwhelm the gravity torque.

Thus where eq. (9) holds in a disk with approximately uniform  $Q \gtrsim 1$ , a *tightly wrapped* trailing wave packet originating on the long-wave branch, will travel towards CR until it reaches the edge of the forbidden zone where it “refracts” into a short wave that carries it away towards the Lindblad resonance. The details of the turning point at the forbidden zone, which requires an evanescent wave propagating into the forbidden region, were described by Mark (1976) and by Goldreich & Tremaine (1978).

The fate of the short-wave propagating towards the Lindblad resonance also requires more sophisticated analysis because the dispersion relation (9) does not hold near resonances. Mark (1974) carried the analysis to second order to show that the wave is absorbed there through the wave-particle coupling described by Lynden-Bell & Kalnajs (1972).

### 3.6. Swing Amplification

Goldreich & Lynden-Bell (1965b), Julian & Toomre (1966) and Toomre (1981) extended the above analysis for tightly-

<sup>4</sup> Additional Bernstein-type waves exist near integer values of  $s > 1$ , but such solutions seem to be of little dynamical importance.

wrapped waves to waves of arbitrary inclination. They again adopted the approximate potential of a short-wavelength plane wave (eq. 7), and focused on a local patch of the disk whose dimensions were assumed small compared with  $R_0$ , the radial distance of the patch center, which orbits the disk center at the local circular speed  $V = R_0\Omega_0$ .

Following Hill (1878), they then introduced Cartesian-like coordinates,  $x \equiv R - R_0$  and  $y \equiv R_0(\phi - \phi_0)$ , where  $\phi_0$  is the azimuth of the orbiting center of the patch. The vertical coordinate,  $z$ , is unchanged from the usual cylindrical polar coordinate. Since the angular rotation rate decreases outwards in galaxies, the motion of stars in the patch is that of a shear flow, with a speed that varies as  $\dot{y} = -2Ax$ , together with Coriolis forces arising from the rotation of the patch at angular rate  $\Omega_0$ . The Oort “constant”  $A = -\frac{1}{2}Rd\Omega/dR$ . The undisturbed surface density of matter,  $\Sigma$ , is assumed constant over the patch. This set of approximations is described as the **sheared sheet** (see BT08 pp. 678-681).

If  $\alpha$  is the inclination angle of a plane density wave to the azimuthal direction, with  $180^\circ > \alpha > 90^\circ$  for leading waves and  $90^\circ > \alpha > 0^\circ$  for trailing waves, then the shear causes the pitch angle to change with time as  $\cot \alpha = 2At$ , with  $t = 0$  when  $\alpha = 90^\circ$ , *i.e.* when the wave is purely radial. Both Goldreich & Lynden-Bell (1965b) for a gaseous sheet and Julian & Toomre (1966) for a stellar sheet found waves amplify strongly, but transiently, as they shear from leading to trailing.

The unremitting shear flow ultimately tears apart any transient disturbance, and indeed Julian & Toomre (1966) asserted that the stellar sheet is locally stable at small-amplitudes when  $Q \geq 1$ , but they did not give the proof.

Fig. 5, from Toomre (1981), illustrates the dramatic transient trailing spiral that results from a small input leading disturbance. This calculation illustrates that even a global, small-amplitude disturbance does not persist, but decays after its transient flourish. In the late stages, the “wave action” propagates at the group velocity (§3.5) away from corotation, where it is absorbed at the Lindblad resonances.

Notice also that the wave in this  $Q = 1.5$  disk extends across the forbidden region near CR. The absence of solutions near  $s = 0$  when  $Q > 1$  in Fig. 4 is a consequence of assuming a steady, tightly-wrapped wave; shearing waves do not respect this restriction.

The responsiveness of a disk to input leading signal depends both on the degree of random motion – it decreases rapidly over the range  $1 < Q < 2$  – and upon the azimuthal wavelength, which is  $\lambda_y = 2\pi R/m$  for an  $m$ -armed pattern. Julian & Toomre (1966) defined the parameter

$$X \equiv \frac{\lambda_y}{\lambda_{\text{crit}}} = \frac{Rk_{\text{crit}}}{m}. \quad (12)$$

In a disk having a flat rotation curve, Toomre (1981) showed that the strongest amplification occurs when  $1 \lesssim X \lesssim 2.5$ , or about *twice* the scale of the longest unstable axisymmetric wave. The peak amplification shifts to larger  $X$  values for declining rotation curves and *vice versa* for rising ones. Naturally, the phenomenon of swing amplification disappears entirely in uniformly rotating disks.

By making the assumption that spiral patterns in galaxies are amplified solely by this mechanism, Athanassoula *et al.* (1987) argued that the multiplicity of the observed

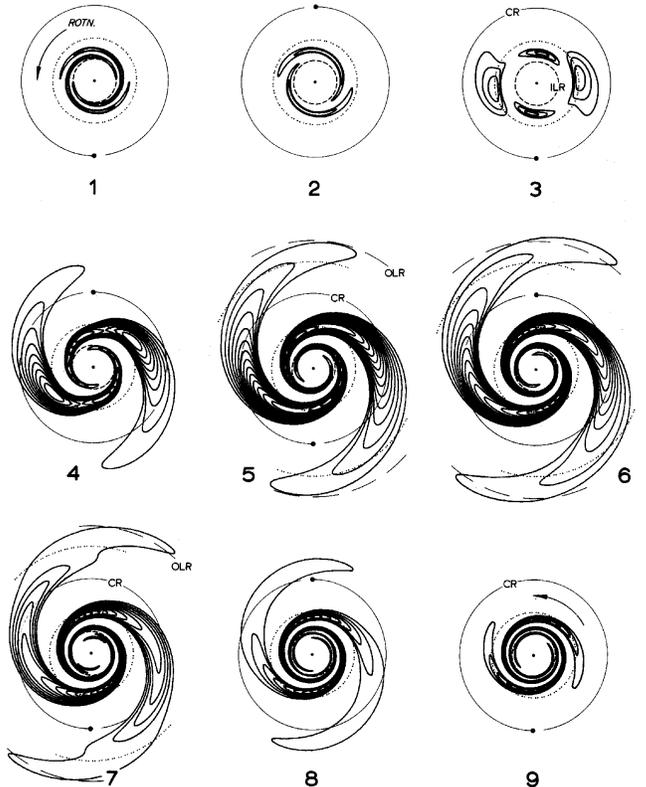


FIG. 5.— The time evolution of an input leading wave packet in the half-mass Mestel disk. The unit of time is half a circular orbit at the radius marked CR. Reproduced from Toomre (1981).

spiral arm pattern can be used to place bounds on the mass density in the disk. The circular orbit speed in a nearly axisymmetric galaxy affords a direct estimate of the central attraction, but it is hard to determine how much of the total attraction is contributed by the disk, and how much by other components, such as the dark matter halo. Because  $\lambda_{\text{crit}}$  is proportional to the disk surface density, for a given rotation curve, the most vigorous amplification shifts to higher sectoral harmonics in lower mass disks. Generally, disks that contribute most of the central attraction, *i.e.* close to **maximum disks**, would prefer bi-symmetric spirals, while higher arm multiplicities would indicate significantly sub-maximal disks (see also Sellwood & Carlberg 1984).

#### 4. BAR INSTABILITY

The puzzle of how galaxy disks could be stable presented a major obstacle to the development of our understanding of disk dynamics for many years. Superficially reasonable models of disk galaxies were found repeatedly, both in simulations (*e.g.* Miller *et al.* 1970; Hohl 1971; Zang & Hohl 1978; Combes & Sanders 1981; Sellwood 1981; Athanassoula & Sellwood 1986; Khoperskov *et al.* 2007; Dubinski *et al.* 2009) and in linear stability analyses (Kalnajs 1978; Aoki *et al.* 1979; Toomre 1981; Sawamura 1988; Vauterin & Dejonghe 1996; Pichon & Cannon 1997; Korchagin *et al.* 2005; Polyachenko 2005; Jalali 2007) to possess vigorous, global and disruptive bar-forming instabilities. Fig. 6

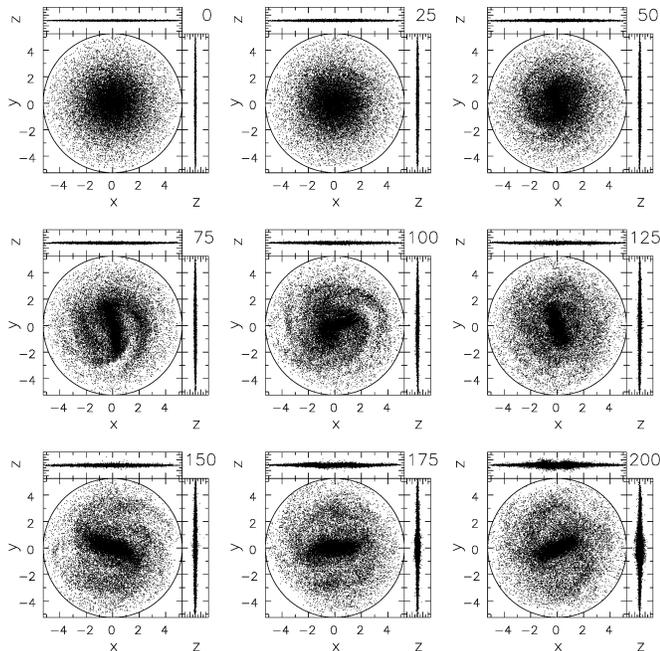


FIG. 6.— The formation of a bar in an (unpublished)  $N$ -body simulation. The three orthogonal projections show only particles in the initially exponential disk, halo particles are omitted. The disk, which started with  $Q = 1.2$ , has unit mass and unit scale length and  $G = 1$ . One orbit at  $R = 2$  takes 15 time units, where the central attractions of the disk and halo are nearly equal.

illustrates the global, disruptive nature of this instability. While it is premature to claim that this problem has been completely solved, it now seems that the stability of a massive disk galaxy requires only a dense bulge-like mass component near the center, and owes little to the inner density of a dark matter halo. Galaxies lacking a central mass concentration, however, are still believed to require significant inner halo mass for global stability.

#### 4.1. Mechanism for the Bar Mode

Toomre (1981) provided the most important step forward by elucidating the mechanism of the bar instability (see also BT08 §6.3). Linear bar-forming modes are standing waves in a cavity, akin to the familiar modes of organ pipes and guitar strings. Reflections in galaxies take place at the center and at the corotation radius, except that outgoing leading spiral waves incident on the corotation circle are super-reflected into amplified ingoing trailing waves (*i.e.* swing amplification), while also exciting an outgoing transmitted trailing wave. The feedback loop is closed by the ingoing trailing wave reflecting off the disk center into a leading wave, which propagates outwards because the group velocity of leading waves has the opposite sign to that of the corresponding trailing wave. The amplitude of the continuous wave-train at any point in the loop rises exponentially, because the circuit includes positive feedback.

Toomre supported this explanation with linear stability studies of two disks. The Gaussian disk, which has a low central density, manifests a set of modes in which the more slowly-growing, higher “overtones” display the kind of standing wave pattern to be expected from the superposition of ingoing trailing and outgoing leading waves.

(The eigenfrequencies of these modes are shown below in Fig. 9.)

The other linear stability study he presented was for the inappropriately-named “Mestel” disk,<sup>5</sup> whose unusual stability properties were first described by Toomre’s student Zang (1976) in an unpublished thesis, and later by Evans & Read (1998). This disk has the scale-free surface density profile  $\Sigma = V_c^2/(2\pi GR)$ , with  $V_c$  being the circular orbital speed that is independent of  $R$ . Zang had carried through a global, linear stability analysis of this disk, with random motions given by a smooth distribution function. In order to break the scale-free nature of the disk, Zang introduced a central cutout, and later an outer taper, in the active mass density, replacing the removed mass by rigid components in order that the central attraction remained unchanged at all radii. The dominant linear instabilities he derived for the tapered disk were confirmed in  $N$ -body simulations by Sellwood & Evans (2001).

Zang (1976) showed that a full-mass Mestel disk is stable to bi-symmetric modes, even if  $Q < 1$ , provided the tapers are gentle enough. Evans & Read (1998) extended this important result to other power-law disks, finding they have similar properties. Such disks are not globally stable, however, because they suffer from lop-sided instabilities (see §5). By halving the active disk mass, with rigid mass preserving the overall central attraction, Toomre (1981) was well able to eliminate the lop-sided mode from the Mestel disk. In fact, he claimed that when  $Q = 1.5$ , the half-mass Mestel disk was globally stable, making it the only known model of a non-uniformly rotating disk that is stable to all small amplitude perturbations.

Cutting the feedback loop really does stabilize a disk (see next section), which seems to confirm Toomre’s cavity mode mechanism. Despite this, Polyachenko (2004) pointed out that the strong emphasis on phenomena at corotation places the blame for the instability on a radius well outside where the resulting bar has its peak amplitude. He therefore proposed an alternative mechanism for the bar instability based upon orbit alignment (Lynden-Bell 1979, see also §9.1). Even though the mechanism was originally envisaged as a slow trapping process, Polyachenko (2004) argued it may also operate on a dynamical timescale, and he devised (Polyachenko 2005) an approximate technique to compute global modes that embodied this idea. While his method should yield the same mode spectrum as other techniques, his alternative characterization of the eigenvalue problem may shed further light on the bar-forming mechanism.

#### 4.2. Predicted Stability

The fact that small-amplitude, bi-symmetric instabilities are so easily avoided in the Mestel disk, together with his understanding of bar-forming modes in other models, led Toomre (1981) to propose that stability to bar-formation merely required the feedback cycle through the center to be cut. He clearly hoped that a dense bulge-like mass component, which would cause an ILR to exist for most reasonable pattern speeds, might alone be enough to stabilize a cool, massive disk.

Unfortunately, this prediction appeared to be contra-

<sup>5</sup> Mestel (1963) solved the far greater challenge of a disk of finite radius that has an exactly flat rotation curve.

dicted almost immediately by the findings of Efstathiou *et al.* (1982), whose  $N$ -body simulations formed similar bars on short timescales irrespective of the density of the central bulge component! They seemed to confirm previous conclusions (Ostriker & Peebles 1973) that only significantly sub-maximal disks can avoid disruptive bar-forming instabilities.

However, Sellwood (1989a) found that Toomre’s prediction does not apply to noisy simulations because of non-linear orbit trapping. Simulations in which shot noise was suppressed by quiet start techniques did indeed manifest the tendency towards global stability as the bulge was made more dense, as Toomre’s linear theory predicted. The collective response to shot noise from randomly distributed particles was of large enough amplitude for non-linearities to be important, and a noisy simulation of a linearly stable disk quickly formed a strong bar, consistent with the results reported by Efstathiou *et al.* (1982).

Since density variations in the distribution of randomly distributed particles are responsible for bar formation in this regime, the reduced shot noise level from larger numbers of particles must result in lower amplitude responses that ultimately should avoid non-linear trapping. The precise particle number required depends on the responsiveness of the disk, which is weakened by random motion, lower surface density, and by increased disk thickness or gravity softening. Efstathiou *et al.* (1982) employed merely 20 000 particles, which was clearly inadequate to capture the linear behavior. However, robustly stable, massive disks have been simulated by Sellwood & Moore (1999) and Sellwood & Evans (2001) that employed only slightly larger particle numbers. Note that these latter models also benefitted from more careful set-up procedures to create the initial equilibrium.

Thus the stabilizing mechanism proposed by Toomre (1981) does indeed work in simulations of high enough quality, and presumably therefore also in real galaxy disks. Indeed, Barazza *et al.* (2008) found a decreased incidence of bars in galaxies having luminous bulges, and argued that their result supports Toomre’s stabilizing mechanism.

Thus the absence of bars in a significant fraction of high-mass disk galaxies does not imply that the disk is sub-maximal. The old stability criteria proposed by Ostriker & Peebles (1973) and Efstathiou *et al.* (1982) apply only to disks that lack dense centers; indeed Evans & Read (1998) showed explicitly that the power-law disks are clear counter-examples to the simple bi-symmetric stability criterion proposed by Ostriker & Peebles (1973).

#### 4.3. Residual Concerns

While all this represents real progress, a few puzzles remain. The most insistent is the absence of large, strong bars in galaxies like M33, which has a gently rising rotation curve. Although many spiral arms can be counted in blue images (Sandage & Humphreys 1980), the near IR view (Block *et al.* 2004) manifests an open two-arm spiral, suggesting that the disk cannot be far from maximal, and also reveals a mild bar near the center of the galaxy. Corbelli & Walterbos (2008) also found kinematic evidence for a weak bar. Perhaps the stability of this galaxy can be explained by some slightly larger halo fraction, or perhaps the weak bar has some unexpected effect, but there is

no convincing study to demonstrate that this galaxy, and others like it, can support a 2-arm spiral without being disruptively unstable.

The second concern is that lop-sided instabilities appear in extended full-mass disks with flat or declining rotation curves, which is discussed next. A third concern, which is discussed in §9.2, is that the mechanism is unable to predict the presence or absence of a bar in a real galaxy.

## 5. LOP-SIDED MODES

Many galaxies have apparently lop-sided disks. The treatment here will not go into detail, since Jog & Combes (2009) have recently reviewed both the observational data and theory.

Both theoretical and simulation work on  $m = 1$  distortions to an axisymmetric disk require special care, since the absence of rotational symmetry can lead to artifacts unless special attention is paid to linear momentum conservation. Rigid mass components present particular difficulties, since they should not be held fixed, and extensive mass components are unlikely to respond as rigid objects.

As noted above, Zang (1976) found that the dominant instability of the centrally cut-out Mestel disk was not the usual bar instability, but a lop-sided mode, which persists in a full mass disk no matter how large a degree of random motion or gentle the cutouts. This surprising finding was confirmed and extended to general power-law disks by Evans & Read (1998). A lop-sided instability dominated simulations (Sellwood 1985) of a model having some resemblance to Zang’s, in that it had a dense massive bulge and no extended halo, while Saha *et al.* (2007) reported similar behavior in simulations of a bare exponential disk. Lovelace *et al.* (1999) found pervasive lop-sided instabilities near the disk center in a study of the collective modes of a set of mass rings.

Various mechanisms have been proposed to account for this instability. Baldwin *et al.* (1980) and Earn & Lynden-Bell (1996) explored the idea that long-lived lop-sidedness could be constructed from cooperative orbital responses of the disk stars, along the lines discussed for bars by Lynden-Bell (1979). Tremaine (2005) discussed a self-gravitating secular instability in near-Keplerian potentials.<sup>6</sup> A more promising mechanism is a cavity mode, similar to that for the bar-forming instability (Dury *et al.* 2008): the mechanism again supposes feedback to the swing-amplifier, which is still vigorous for  $m = 1$  in a full-mass disk.

Feedback through the center cannot be prevented by an ILR for  $m = 1$  waves, since the resonance condition  $\Omega_p = \Omega_\phi - \Omega_R$  (eq. 5) is satisfied only for retrograde waves. But the lopsided mode can be stabilized by reducing the disk mass, which reduces the  $X$  parameter (eq. 12) until amplification dies for  $m = 1$  (Toomre 1981). Sellwood & Evans (2001) showed that, together with a moderate bulge, the dark matter required for a globally stable disk need not be much more than a constant density core to the minimum halo needed for a flat outer rotation curve.

A qualitatively different lop-sided instability is driven by counter-rotation. This second kind of  $m = 1$  instability was first reported by Zang & Hohl (1978) in a series of

<sup>6</sup> The “sling amplification” mechanism proposed by Shu *et al.* (1990) applies only to gaseous accretion disks, since it relies on sound waves propagating outside the OLR.

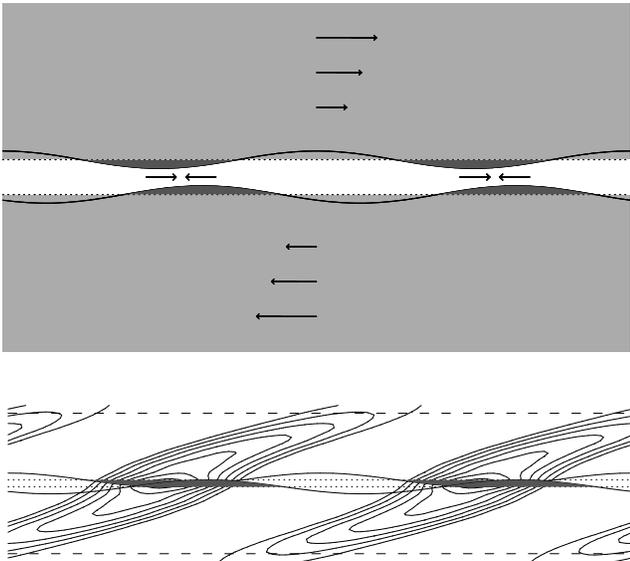


FIG. 7.— The top panel illustrates a groove in the sheared sheet (§3.6) model. The light shaded region has the full undisturbed disk surface density,  $\Sigma$ , which in this frame has a linearly varying shear flow. The white region is the groove, which has lower surface density, and the density excesses brought into the groove by small wave-like disturbances on either side of the groove are highlighted by the dark shading. The width of the groove is exaggerated for clarity. The lower panel, which has the correct aspect ratio, shows the supporting responses to the density variations on the groove edges. The dashed lines mark the locations of the Lindblad resonances.

$N$ -body simulations designed to explore the suppression of the bar instability by reversing the angular momenta of a fraction of the stars; they found that a lop-sided instability was aggravated as more retrograde stars were included in their attempts to subdue the bar mode. Analyses of various disk models with retrograde stars (Araki 1987; Sawamura 1988; Dury *et al.* 2008) have revealed that the growth rates of lop-sided instabilities increase as the fraction of retrograde stars increases. Merritt & Stiavelli (1990) and Sellwood & Valluri (1997) found lop-sided instabilities in simulations of oblate spheroids with no net rotation. Their flatter models had velocity ellipsoids with a strong tangential bias, whereas Sellwood & Merritt (1994) found that disks with half the stars retrograde, together with moderate radial motion were surprisingly stable.

Weinberg (1994) pointed out that lop-sided distortions to near spherical systems can decay very slowly, leading to a protracted period of “sloshing”. This **seiche mode** in a halo, which decays particularly slowly in mildly concentrated spherical systems, could provoke lop-sidedness in an embedded disk (Kornreich *et al.* 2002; Ideta 2002).

If lop-sidedness is due to instability, then the limited work so far suggests that it would imply a near-maximum disk. But lop-sidedness in the outer parts could also be caused by tidal interactions, or simply by asymmetric disk growth, with the effects of differential rotation being mitigated perhaps by the cooperative orbital responses discussed by Earn & Lynden-Bell (1996).

## 6. GROOVE AND EDGE MODES

Not all true global instabilities in disks require a feedback loop. Another class of modes is simply driven from

corotation, with trailing waves propagating both inwards and outwards to be absorbed, in stellar disks, at the Lindblad resonances on either side.

Lovelace & Hohlfield (1978) showed that disks are destabilized by a local extremum in the radial variation of the ratio of surface density to vorticity,  $\Sigma/|\nabla \times \mathbf{v}| \equiv \Sigma\Omega/(2\kappa^2)$ , or the reciprocal of potential vorticity, and proposed that the instability created flat rotation curves for which the potential vorticity is also flat. Sellwood & Kahn (1991) demonstrated that the instability caused by a rather insignificant, narrow, decrease in surface density, *i.e.* a “groove”, is a global spiral mode.

The mechanism is easiest to visualize in a disk without random motion, where small-scale surface density variations are not blurred by epicyclic motions. In this case, a deficiency of stars over a small range of angular momentum creates a groove in the surface density, as shown in the sheared sheet model in the top panel of Fig. 7. The groove itself is unstable because of the gravitational coupling between disturbances on either side. The dark shaded areas in the Figure illustrate regions where small sinusoidal radial displacements of material on each edge have created high density regions where the density was previously low. Disturbing gravitational forces arise from the density excesses, as illustrated, which are directed along the groove if the wavelength is long compared with the groove width. Material that is pulled back loses angular momentum and sinks toward the center of the galaxy, while that which is urged forward gains and rises outwards. Thus each density excess pulls on the other across the groove in such a way as to cause it to grow exponentially, *i.e.* the combined disturbance on the two sides is unstable. The groove edges need be only steep gradients, not discontinuities. Furthermore, the mechanism is the same, but harder to visualize, in a disk with random motion where the density of stars is depleted over a narrow range of angular momentum.

The growing disturbance in the groove creates wave-like mass variations along the groove that are effectively growing co-orbiting mass clumps of the type envisaged by Julian & Toomre (1966). Generalizing their apparatus to allow for exponentially growing masses, Sellwood & Kahn (1991) estimated the expected disk response, as shown in the lower panel of Fig. 7 for the parameters  $Q = 1.8$  and  $X = 2$ , and low growth rate. The disk supporting response transforms the quite trivial disturbance in the groove into an extensive spiral instability!

Unlike the sheared sheet, the azimuthal wavelength in a full disk can take on only discrete values,  $\lambda_y = 2\pi R/m$ , which are all unstable, but the one which grows most rapidly is that for which swing-amplification (§3.6) is the most vigorous, *i.e.*  $1 \lesssim X \lesssim 2.5$  (eq. 12) in a flat rotation curve. The simulation shown in Fig. 8 illustrates the scale and vigor of the mode, which saturates at  $\sim 20\%$  overdensity because of the onset of horseshoe orbits at corotation (Sellwood & Binney 2002).

Lovelace & Hohlfield (1978) and Sellwood & Kahn (1991) found that almost any narrow feature in the angular momentum density is destabilizing, although the modes of a simple ridge, for example, come in pairs with CR some distance from the ridge center.

**Edge modes** (Toomre 1981; Papaloizou & Lin 1989) are close cousins of groove modes, and the mechanism can

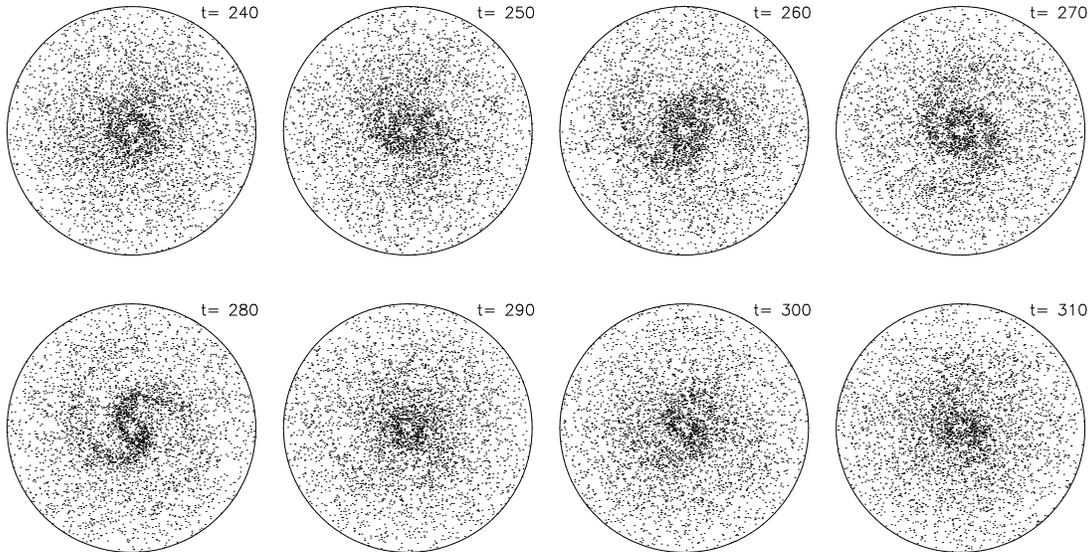


FIG. 8.— The later part of the growth and subsequent decay of an isolated spiral mode in a disc that was seeded with a groove. Disk rotation is counter-clockwise and disturbance forces in this simulation, taken from Sellwood & Binney (2002), were restricted to  $m = 2$  only.

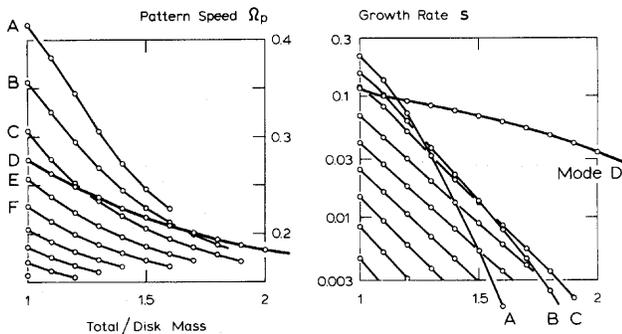


FIG. 9.— The variation of pattern speeds and growth rates for the first few  $m = 2$  modes of the Gaussian disk as the active disk mass is decreased. From Toomre (1981).

be understood in the same fashion. Density variations on a single edge without a supporting disk response are neutrally stable, but the necessarily one-sided wakes in the massive interior disk add angular momentum to the density enhancements on the edge, causing them to grow.

The eigenfrequencies of the low order bisymmetric modes of the Gaussian disk vary with decreasing disk mass as shown in Fig. 9, taken from Toomre (1981). Most are cavity modes, but the edge-mode, labelled “mode D” stands out because its growth rate in particular declines more slowly with decreasing disk mass. If the mechanism for both types of mode involves swing amplification, one might expect their vigor to be similarly affected by the rise in the  $X$  parameter as the surface density decreases. But the decreased surface density also slows the group velocity while the rapidly declining pattern speed of each cavity mode moves CR farther out in the disk. Thus the growth rates of cavity modes drop more quickly because of the increased travel time for the wave packet to complete the feed-back loop, a factor that is absent for the edge mode.

The edge instability requires only a steep gradient in

the surface density, which need not drop to zero. Toomre (1989) gave the condition for instability as “the radial distance over which the disc density undergoes most of its rapid change should be no larger than about one quarter of the axisymmetric stability length  $\lambda_{\text{crit}}$ ”, assuming “the disc is massive and cool enough for vigorous swing-amplification.”

Curiously, global bar-forming modes were originally thought to be related the rotational instabilities of uniform-density Maclaurin spheroids of incompressible fluid (Ostriker & Peebles 1973), an idea reinforced by the vigor of the bisymmetric mode of the sharp-edged Maclaurin disk (Kalnajs 1972). Toomre (1981) noted that the instability of such unrealistic galaxy models may be more closely related to the edge mode than to the cavity mode described in §4.1.

As the consequence of an edge instability in a realistic galaxy model is to blur the edge, it seems unlikely that galaxy disks can retain unstable density gradients for interestingly long periods. Similarly, the large-amplitude evolution of groove or ridge modes quickly erases the feature that gave rise to them. Nevertheless, the modes have other possible consequences (see §7.6).

## 7. SPIRAL STRUCTURE

Despite many decades of effort, no theory to account for the graceful spiral patterns in disk galaxies is widely accepted. Most workers in this field agree that spiral patterns are gravitationally driven variations in surface density in the old stellar disk, as supported by photometric data (*e.g.* Grosbøl *et al.* 2004; Zibetti *et al.* 2009) and streaming motions in high spatial resolution velocity maps (*e.g.* Shetty *et al.* 2007).

There seems little doubt that some spiral patterns are tidally driven (*e.g.* Salo & Laurikainen 1993; Dobbs *et al.* 2010), while others could be the driven responses to bars (Buta *et al.* 2009). Although these two ideas may account

for a large fraction of the cases (Kormendy & Norman 1979), especially if orbiting dark matter clumps can excite patterns (Dubinski *et al.* 2008), spirals can still develop in the absence of either trigger, as revealed in simulations.

The idea that spirals could be self-excited oscillations of the stellar disk represents the greatest theoretical challenge. While there is general agreement that gas seems to be essential, no picture seems complete, and current theories disagree on even the lifetimes of the patterns. One idea (*e.g.* Bertin & Lin 1996), is that spiral features are manifestations of quasi-steady global modes of the underlying disk. Alternatively, they could be short-lived, recurrent transient patterns that originate either from forcing by density fluctuations (*e.g.* Toomre 1990), or else from recurrent vigorous instabilities (*e.g.* Sellwood 2000).

A serious obstacle to progress in this area has been the absence of observational discriminants that would favor one of these radically differing viewpoints over the other. The predictions for density variations or gas responses at a single instant are essentially independent of the generating mechanism and do not depend strongly on the lifetime of the pattern. Meidt *et al.* (2009) attempted to measure radial variations of pattern speeds using a generalization of the method devised by Tremaine & Weinberg (1984a) (see §9.5). They reported lower pattern speeds at larger radii, but their measurements still tell us little about spiral lifetimes. However, the velocity distribution of stars in the solar neighborhood (Nordström *et al.* 2004, and Fig. 1) is most naturally accounted for in the transient picture (see §7.6).

### 7.1. Spirals as Global Modes of Smooth Disks

Simple models of disk galaxies possess many linear instabilities (*e.g.* Jalali 2007). The bar-forming mode (§4) is usually the fastest growing, but it has almost no spirality. These studies are therefore important to understand stability, but even the higher modes are not particularly promising for spiral generation.<sup>7</sup>

The “density wave” theory for spiral modes, described in detail by Bertin & Lin (1996), invokes a more specific galaxy model with a sub-maximal disk that is dynamically cool in the outer parts and hot in the inner disk. The local stability parameter is postulated to be  $1.0 \lesssim Q \lesssim 1.2$  over most of the disk and to rise steeply to  $Q > 2$  near the center. Bertin, Lin, and their co-workers perform a global analysis using the hydrodynamic approximation (BT08 §5.1), which reveals slowly growing spiral modes under these specific conditions.

The mechanism (Mark 1977) is a cavity mode, having qualitative similarities to that for the bar mode (§4.1), but the tightly-wrapped waves are trailing around the entire cycle. The inner turning point is at a  $Q$ -barrier: a steeply increasing  $Q$  value causes the forbidden zone (§3.4) to broaden to the point that in-going short waves get “refracted” into outgoing long waves, which prevents the wave

train from reaching the ILR where it would be damped. The long waves then propagate out to near the CR in the  $Q \sim 1$  part of the disk, where they switch back to the shortwave branch with a small degree of amplification. The WASER mechanism (Mark 1976) at this turning point involves a third, transmitted wave that is “radiated” outwards on the short-wave branch, carrying away the angular momentum to excite the mode in the inner disk. Thus the amplifier involves a small “swing” from the long- to the short-wave branch, whereas the bar instability uses a full swing from leading to trailing. This difference, together with their assumption that the disk is sub-maximal (*i.e.*  $X > 3$  for  $m = 2$ , see eq. 12), allows the mode to be slowly growing. Lowe *et al.* (1994) present a model of this kind to account for the spiral structure of M81.

In order to justify the “basic state” of the disk they require, Bertin & Lin (1996) argued heuristically that rapidly evolving features would have disappeared long ago and that low-growth-rate instabilities in a cool disk, created by gas dissipative processes and star formation, will dominate at later times. They invoked shocks in the gas to limit the amplitude of the slowly growing mode, leading to a quasi-steady global spiral pattern.

The main objection to their picture is that it is likely that such a lively outer disk will suffer from other, more vigorous, collective responses with  $m > 2$  that will quickly heat the disk, as described in §7.5, and destroy their postulated background state.

### 7.2. Recurrent Transients

From the early work by Miller *et al.* (1970), Hohl (1971), Hockney & Brownrigg (1974), James & Sellwood (1978), and Sellwood & Carlberg (1984),  $N$ -body simulations of cool, sub-maximal disks have exhibited recurrent transient spiral activity. This basic result has not changed for several decades as numerical quality has improved.

Claims of long-lived spiral waves (*e.g.* Thomasson *et al.* 1990) have mostly been based on simulations of short duration. For example, Elmegreen & Thomasson (1993) presented a simulation that displayed spiral patterns for  $\sim 10$  rotations, but the existence of some underlying long-lived wave is unclear because the pattern changed from snapshot to snapshot. Donner & Thomasson (1994) ran their simulations for fewer than 7 disk rotations and the bi-symmetric spiral they observed appeared to be an incipient bar instability. Zhang (1996,1998) adopted a similar mass distribution and also reported long-lived patterns in her simulations. The author has attempted to reproduce her results, and indeed obtained similar bi-symmetric features, but they appear to be the super-position of several waves having differing pattern speeds.

Sellwood & Carlberg (1984) stressed that patterns fade in simple simulations that do not include the effects of gas dissipation; the reason is the disk becomes less responsive as random motion rises due to particle scattering by the spiral activity (§10), which is therefore self-limiting. They also demonstrated that mimicking the effects of dissipative infall of gas, such as by adding fresh particles on circular orbits, allowed patterns to recur “indefinitely.” Later work (*e.g.* Carlberg & Freedman 1985; Toomre 1990) has shown that almost any method of dynamical cooling can maintain spiral activity, as also happens in modern galaxy formation

<sup>7</sup> Korchagin *et al.* (2005) calculated essentially gas-dynamical modes for models of specific galaxies, and argued that the shapes of one, or more, of the lower-order modes could be matched to the observed spiral pattern. However, it is unclear that rapidly growing, linear modes can be seen for long at finite amplitude before non-linear effects will change their appearance, and it seems even less likely that two modes with different growth rates should have similar large amplitudes at the time a galaxy is observed.

simulations (*e.g.* Agertz *et al.* 2010).

Thus the transient spiral picture offers a natural explanation for the absence of spiral patterns in S0 disk galaxies that have little or no gas; maintenance of spiral activity requires a constant supply of new stars on near-circular orbits. Other pieces of indirect evidence that also favor the transient spiral picture are reviewed in §10.

### 7.3. Spirals as Responses to Density Fluctuations

Goldreich & Lynden-Bell (1965b) and Toomre (1990) suggested that a large part of the spiral activity observed in disk galaxies is the collective response of the disk to clumps in the density distribution. As a spiral wake is the collective response of a disk to an individual co-orbiting perturber (Julian & Toomre 1966), multiple perturbers will create multiple responses that all orbit at different rates. The behavior of this polarized disk reveals a changing pattern of trailing spirals, which can equivalently be regarded as swing-amplified (§3.6) noise.<sup>8</sup>

Toomre & Kalnajs (1991) studied the amplified noise that arose in their  $N$ -body simulations of the sheared sheet. The massive particles themselves provoke spiral responses with an amplitude proportional to the input level of shot noise, caused by density variations in the distribution of randomly distributed particles. Comparison of their expectations with linear theory predictions revealed that the simulations were livelier than they expected, by a factor  $\lesssim 2$ , apparently from a gradual build-up of correlations between the mean orbital radii of the particles.

This could be a mechanism for chaotic spirals in very gas rich discs, where a high rate of dissipation may be able to maintain the responsiveness of the disc (Toomre 1990) while the clumpiness of the gas distribution may make the seed noise amplitude unusually high. However, it seems likely that spiral amplitudes (*e.g.* Zibetti *et al.* 2009) are too large to be accounted for by this mechanism in most galaxies. Also, the spiral structure should be chaotic, with little in the way of clear symmetry expected.

### 7.4. Non-linear Spiral Dynamics

Tagger and his co-workers (Tagger *et al.* 1987; Sygnet *et al.* 1987; Masset & Tagger 1997) suggested that global modes in stellar disks could be coupled through non-linear interactions. They proposed that wave 1 excites wave 2 through second-order coupling terms that are large when CR of wave 1 lies at approximately the same radius as the ILR of wave 2. Conservation rules require a third wave such that  $m_3 = m_1 \pm m_2$ ; *i.e.* most likely an axisymmetric wave ( $m_3 = 0$ ) if  $m_1 = m_2$ . Many examples of multiple waves in  $N$ -body simulations have been reported with remarkable coincidences for the radii of the main resonances.

Fuchs *et al.* (2005) developed a similar argument for waves in the sheared sheet. They found that amplified trailing waves could excite fresh leading waves in their second-order theory. They proposed this mechanism as an alternative source of the amplitude excess noted above in the simulations by Toomre & Kalnajs (1991), and they speculated that the same mechanism may also account for the larger than expected amplitudes of spirals in global  $N$ -body simulations (see §7.5). As both this mechanism, and

that discussed in the previous paragraph, employ terms that are second-order in the perturbation amplitude, they will become important only when features are strong.

Patsis *et al.* (1991) attempted to construct, by orbit superposition, self-consistent steady spiral waves of finite amplitude to match the observed non-axisymmetric patterns in specific galaxies. They experienced great difficulty in finding solutions near CR, and suggested that either this resonance or the 4:1 resonance<sup>9</sup> marks the outer radius of the spiral. Their finding is not unexpected for two reasons: (a) the dispersion relation for steady, tightly-wrapped, small-amplitude waves (Fig. 4) predicts a forbidden region around CR for all  $Q > 1$ , and (b) the non-linear dynamics of orbits in barred potentials (see §9.3) finds only chaotic orbits near CR, which are unfavorable to self-consistency as Patsis *et al.* (1991) found. It should be noted that the difficulty of finding a self-consistent solution near CR is a direct consequence of their assumption of a steady wave pattern; transient waves do not suffer from this problem (Fig. 5) and rapidly growing groove modes (§6) have *peak* overdensities at CR.

Tsoutsis *et al.* (2009) and Athanassoula *et al.* (2009) suggested that spirals in barred galaxies could be created by the slow migration of stars along Lyapunov manifold tube orbits emanating from the unstable Lagrange points at the ends of the bar.<sup>10</sup> The ambitious hope of these preliminary papers is for an ultimate unified picture for the co-existence of bars, spirals and rings, all having the same pattern speed.

### 7.5. Spirals in Global $N$ -body Simulations

Sellwood & Carlberg (1984) and Sellwood (1989b) reported that their global simulations manifested more vigorous spiral activity than was consistent with amplified shot noise. A brief summary of some further results to support this claim is given here, and will be described more fully elsewhere.

As noted in §4.1, Toomre (1981) predicted the half-mass Mestel disk to be globally stable to small-amplitude disturbances. Thus  $N$ -body simulations of this disk should exhibit no activity in excess of the inevitable swing-amplified shot noise. Fig. 10 reveals that this is not the case. The ordinate shows the largest value of  $\delta = \Delta\Sigma/\Sigma$  from  $m = 2$  disturbances in a sequence of simulations with increasing numbers of particles. The unit of time is  $R_i/V_c$ , where  $V_c R_i$  is the center of the inner angular momentum cut out. Thus the orbit period at this small radius is  $2\pi$ .

At  $t = 0$ ,  $\delta \propto N^{-1/2}$ , as appropriate for shot noise, and swing-amplification causes an almost immediate jump by a factor of a few for all  $N$ . When  $N = 5 \times 10^4$ , amplified noise causes  $\sim 20\%$  overdensities almost immediately. For larger  $N$  the amplitude eventually rises to similar values in later evolution, once the inner disk has developed a pronounced bar. But for the largest two values of  $N$

<sup>9</sup> The resonance condition (eq. 5) for a pure  $\cos(m\theta)$  potential variation also implies frequency commensurabilities at multiples of  $m$ . The small denominators that characterize the principal resonances (BT08, §3.3.3) arise at these “ultraharmonic resonances” only for non-circular orbits. A new family of orbits appears at the 4:1 resonance that closes after 4 radial oscillations (see §9.3 and Sellwood & Wilkinson 1993).

<sup>10</sup> Interestingly, this idea harks back to the old “garden sprinkler” model for spirals proposed by Jeans (1923) (see also Jeans 1929, fig. 55 & pp 357–360).

<sup>8</sup> Cuzzi *et al.* (2010) found evidence for similar behavior within Saturn’s A ring.

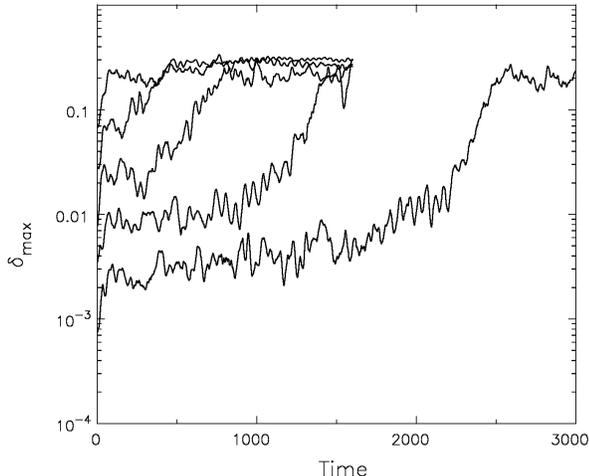


FIG. 10.— The time evolution of the peak overdensity in a series of simulations of the half-mass Mestel disk with different numbers of particles. The model was predicted by Toomre (1981) to be globally stable. The ordinate shows the maximum values of  $\delta$  on grid rings. The number of particles in each simulation rises by a factor 10 from  $N = 5 \times 10^4$  to  $N = 5 \times 10^8$ , and the initial amplitude is  $\propto N^{-1/2}$ . Linear theory predicts the later amplitudes should have the same scaling.

shown, a period of slow growth occurs after the initial swing-amplified surge, offering tentative support for the linear theory prediction of global stability, with the slow growth perhaps arising from the gradual development of particle correlations as described by Toomre & Kalnajs (1991). Even in these cases, however, the amplitude rises more rapidly once  $\delta \gtrsim 2\%$ .

Thus even in these highly restricted simulations, spiral activity always exhibits runaway growth – albeit more and more delayed as  $N$  is increased – behavior that is quite clearly not in accord with linear theory predictions. The rapid growth once  $\delta \gtrsim 2\%$  suggests that the behavior has already become non-linear in some respect at this modest amplitude. Note that the largest number of particles,  $N = 5 \times 10^8$ , is within a factor of 100 of the number of stars in a real galaxy disk, where in reality the mass distribution is far less smooth, owing to the existence of star clusters and giant gas clouds.

### 7.6. A Recurrent Instability Cycle?

Sellwood & Lin (1989) reported evidence for a recurrent instability cycle in their simulations of a very low-mass disk. They observed that each spiral pattern created substantial changes to the distribution of particles at the Lindblad resonances, which they suspected created conditions for a new global instability of the groove mode kind (see §6). They explicitly demonstrated that the later features were indeed true instabilities, since when they continued a parallel simulation after randomly shuffling the particles in azimuth at some moment, the pattern speed of the next mode to appear was the same as that in the original simulation that had not been interrupted.<sup>11</sup> Thus, each coherent wave leaves behind an altered DF that apparently provokes a new instability.

<sup>11</sup> This behavior is inconsistent with the non-linear mode coupling ideas discussed in §7.4.

The runaway growth in the larger  $N$  models shown in Fig. 10 is not caused by a single unstable mode, but appears to be a succession of separate coherent waves, each having a lower rotation rate and reaching a higher amplitude than the last. Sellwood (2000) reported similar behavior in lower- $N$  simulations, and demonstrated that one of the waves did indeed cause strong scattering at the ILR. It should be noted that resonance scattering is a second-order effect, but the evidence shown in Fig. 10 suggests that it becomes important at a relative overdensity of just  $\Delta\Sigma/\Sigma \sim 2\%$ . Exactly how the demise of one mode creates the conditions for the next instability remains unclear, however.

Since the only evidence for this behavior came from a (well-tested)  $N$ -body code, it seemed best not to pursue the idea further until supporting evidence could be found. Sellwood (1994) therefore expressed the hope that evidence of resonance scattering could be found in the HIPPARCOS measurements of the local stellar kinematics. The publication of the GCS with distances and full phase-space motions of  $\sim 14000$  F & G dwarf stars (see Fig. 1) enabled Sellwood (2010) to show that the so-called Hyades stream is in fact caused by scattering at a recent ILR.

This very recent evidence supports the idea that spirals in the local Milky Way, and presumably elsewhere, do in fact behave as the simulations had indicated. Further work is required to expose the details of the recurrence mechanism. However, it now seems misguided to search for a spiral instability as some devious sort of cavity mode in a smooth disk; *i.e.* the assumption of a featureless DF may have thrown the spiral baby out with the bathwater!

## 8. BUCKLING INSTABILITIES AND WARPS

The optically visible parts of galactic disks are usually remarkably thin and flat, whereas the more extended H I disks of many edge-on galaxies appear noticeably warped with an integral sign shape. Stellar warps (Reshetnikov *et al.* 2002; Saha *et al.* 2009) are much less pronounced than the warps in the extended gaseous disks. The long-known warp in the H I layer of the Milky Way (Oort *et al.* 1958) has been most recently analyzed by Levine *et al.* (2006), while the dust and stars of the outer disk appear to be distorted in a similar, though less extensive, shape (Reylé *et al.* 2009). Both gaseous and stellar warps are frequently asymmetric, as appears to be the case for the Milky Way. The fact that stellar warps usually follow the same warped surface as do the gaseous ones (see also Cox *et al.* 1996), is strong evidence that warps are principally a gravitational phenomenon.

Warps are extremely common. H I observations of edge-on galaxies (Bosma 1991; García-Ruiz *et al.* 2002b) find very high fractions of warps, and the true fraction must be even higher, since warps directed close to our line of sight may be missed. Warps can also be detected kinematically even when the system is not edge-on. Their ubiquity suggests that warps are either repeatedly regenerated or long-lived.

Briggs (1990) studied a sample of 12 warped galaxies with high-quality 21-cm data, and found that galactic warps obey three fairly simple rules:

1. The H I layer typically is coplanar inside radius  $R_{25}$ , the radius where the B-band surface brightness is  $25 \text{ mag arcsec}^{-2} = 25\mu_B$ , and the warp develops between  $R_{25}$  and  $R_{26.5}$  (the Holmberg radius).
2. The **line of nodes** (LoN) is roughly straight inside  $R_{26.5}$ .
3. The LoN takes the form of a loosely-wound *leading spiral* outside  $R_{26.5}$ .

Kahn & Woltjer (1959) first drew attention to the winding problem presented by warps (see also BT08 §6.6.1), and argued that while self-gravity would slow the differential precession, it could not be strong enough to create a long-lived warp. Lynden-Bell (1965), on the other hand, suggested that warps could result from a persisting misalignment between the spin axis and the disk normal, *i.e.* a long-lived mode.

This section describes the theory of bending waves in general before addressing warps more directly. Unstable bending modes will be denoted **buckling modes**, although other names that come from plasma physics, such as hose, firehose, or hosepipe instabilities, are commonly used.

### 8.1. Local Theory of Bending Waves

Toomre (1966) considered the bending stability of an infinite, thin slab of stars having a velocity dispersion  $\sigma_x$  in the  $x$ -direction and some characteristic thickness  $z_0$ . The self-gravity of the slab causes it to bend coherently provided the vertical oscillations of stars are adiabatically invariant as they move over the bend, which requires the slab to be thin compared with the wavelength of the bend. Furthermore, if the bending amplitude is small, its effect on the horizontal motion is negligible.

Toomre derived the dispersion relation for small-amplitude, long-wave ( $kz_0 \ll 1$ ) distortions of the form  $h(x, t) = He^{i(kx - \omega t)}$ :

$$\omega^2 = 2\pi G\Sigma|k| - \sigma_x^2 k^2, \quad (13)$$

where  $\Sigma$  is the vertically-integrated surface density of the slab. The first term on the RHS is the restoring force from the perturbed gravity caused by the bend, while the second is the inertia term due to the vertical acceleration needed for stars to follow the corrugations. The inertia term is destabilizing and outweighs the gravitational restoring force when  $\lambda < \lambda_J = \sigma_x^2/G\Sigma$ , causing the distortion to grow exponentially. The unstable range of the buckling instability is precisely complementary to that of the Jeans instability in the plane in the absence of rotation, which is unstable for wavelengths  $\lambda > \lambda_J$  (Toomre 1966).

The dispersion relation (13) predicts a buckling instability, at sufficiently short wavelengths, for any razor-thin system. However, it does not hold for wavelengths shorter than or comparable to the actual vertical thickness of the slab. Since  $z_0 \sim \sigma_z^2/(G\Sigma)$ , where  $\sigma_z$  is the vertical velocity dispersion, one expects that bending at all wavelengths will be suppressed when  $\sigma_z/\sigma_x$  exceeds some critical value, which Toomre (1966) estimated to be 0.30.<sup>12</sup>

<sup>12</sup> Kulsrud *et al.* (1971) and Fridman & Polyachenko (1984) considered the bending instability in a constant density slab of stars with sharp edges.

Araki (1985) carried through a linear normal mode analysis of the infinite, isothermal slab (Spitzer 1942; Camm 1950), which has the properties  $\rho = \rho_0 \text{sech}^2(z/2z_0)$ ,  $z_0 = \sigma_z^2/(2\pi G\Sigma)$ , and  $\Sigma = 4\rho_0 z_0$ . He assumed a Gaussian distribution of  $x$ -velocities, with  $\sigma_x \neq \sigma_z$ , and determined the range of unstable wavelengths as the slab was made thicker. He showed that the buckling instability could be avoided at all wavelengths when  $\sigma_z > 0.293\sigma_x$ , in good agreement with Toomre's earlier estimate. At the marginal stability threshold, the last unstable mode has a wavelength  $\simeq 1.2\lambda_J$ .

Galaxy disks are not, of course, infinite slabs subject to plane-wave distortions, but the radial velocity dispersion,  $\sigma_R$ , which is larger than the azimuthal dispersion, could perhaps be great enough to drive a buckling instability. As the observed ratio of velocity dispersions for stars in the Solar neighborhood is  $\sigma_z/\sigma_R \sim 0.6$  (Nordström *et al.* 2004), Toomre (1966) concluded that at least this region of our Galaxy is "apparently well clear of this stability boundary." It should be noted that the approximate solution for the potential, which requires  $kR \gg 1$ , is stretched in this case, since  $\lambda_J \approx 7 \text{ kpc}$  when  $\sigma_R = 40 \text{ km s}^{-1}$  and  $\Sigma = 50 M_\odot \text{ pc}^{-2}$ .

When the disk is embedded in some external potential, arising from a halo or the distant bulge of the galaxy say, the dispersion relation for short-wavelength waves in a thin slab (13) acquires an additional stabilizing term

$$\omega^2 = \nu_{\text{ext}}^2 + 2\pi G\Sigma|k| - \sigma_x^2 k^2, \quad (14)$$

where  $\nu_{\text{ext}}^2 = |\partial^2 \Phi_{\text{ext}}/\partial z^2|_{z=0}$  (BT08 eq. 6.114). Taking this additional factor into account further reinforces local stability and global, axisymmetric simulations (Sellwood 1996a) confirmed that Toomre's conclusion holds everywhere in an axisymmetric, but otherwise plausible, model of the Milky Way disk.

As for WKB spiral waves (§3.2), the dispersion relation (14) can be generalized to *tightly-wrapped* non-axisymmetric bending waves simply by replacing  $\omega$  with  $\omega - m\Omega$ , with the angular rate of precession of the bending wave that has  $m$ -fold rotational symmetry being  $\Omega_p = \omega/m$ . It should be borne in mind that since observed warps in galaxies are very far from being tightly wound, analyses that make this approximation yield results that are at best only indicative of the dynamical behavior.

Vertical resonances between the bending wave and the stars arise where  $m(\Omega_p - \Omega) = \pm\nu$  (§2.6), although the meaning of  $\nu$  here depends on the context. BT08 (§6.6.1) considered only razor thin disks, for which the internal oscillation frequency  $\nu_{\text{int}}$  is infinite and the resonances occur where  $m(\Omega_p - \Omega) = \pm\nu_{\text{ext}}$ . In disks of finite thickness, the stars have a natural internal vertical frequency  $\nu_{\text{int}}^2 \approx 4\pi G\rho_0$  (exact in the mid-plane of an infinite slab), and for vertical resonances the appropriate value of  $\nu = (\nu_{\text{int}}^2 + \nu_{\text{ext}}^2)^{1/2}$ . Henceforth,  $\nu$  will be used to mean either this total frequency in a thickened layer or  $\nu_{\text{ext}}$  for a razor-thin sheet.

Eq. (14) is satisfied for both  $\pm m(\Omega_p - \Omega)$ , implying two possible pattern speeds known as **fast** and **slow waves**. Because the gravity term raises the Doppler-shifted frequency above  $\nu$ , waves in a cold disk ( $\sigma_R = 0$ ) extend away from the vertical resonances, and do not exist in the region between them that includes CR. The slow wave,

which is a retrograde pattern for  $m = 1$ , is of most interest because the unforced precession rate,  $\Omega_p = \Omega - \nu$ , has much the milder radial variation.

Much like density waves, neutrally-stable bending waves propagate in a disk in the radial direction at the **group velocity** (Toomre 1983)

$$v_g = \frac{\partial \omega}{\partial k} = \frac{\text{sgn}(k)\pi G \Sigma - k \sigma_R^2}{m(\Omega_p - \Omega)}. \quad (15)$$

Since the denominator is negative for slow bending waves, trailing waves ( $k > 0$ ) in a cold disk ( $\sigma_R = 0$ ) propagate inwards, while leading waves ( $k < 0$ ) propagate outwards. As for waves in non-uniform whips, the bending amplitude rises when the wave propagates into a region of lower surface density, and *vice versa*.

Unfortunately, a full description of wave propagation in a sheet of finite thickness requires a solution of the linearized Boltzmann and Poisson equations (Toomre 1966; Araki 1985; Weinberg 1991; Toomre 1995). The results are not analytic and surprisingly more complex than equation (14) for the razor-thin case. Because the vertical potential of the disk is anharmonic, stars whose vertical oscillation takes them far from the mid-plane have lower vertical frequencies. Thus, any Doppler-shifted frequency  $m(\Omega_p - \Omega)$  of the bending wave will be in resonance with some stars that will damp the wave by converting wave energy into increased random motion. In general, short-wavelength modes  $kz_0 \gtrsim 0.5$  damp in less than one wavelength (Weinberg 1991; Toomre 1995), while long-wavelength modes  $kz_0 \ll 1$  can propagate large distances.

Sellwood *et al.* (1998) studied the propagation and damping of axisymmetric bending waves in disks having both finite thickness and non-zero random velocities. Waves launched from the center of the disk at a fixed frequency propagated outwards and were damped as they approached the vertical resonance. As a result the disk thickened over a small radial range, with the peak occurring just interior to the resonance.

### 8.2. Global Bending Modes

Because of the complication caused by finite thickness, analytic work almost always adopts the razor-thin approximation, which also requires  $\sigma_R = 0$ , since thin disks with velocity dispersion are buckling unstable.<sup>13</sup> With these assumptions, the disk can be approximated by a finite number of gravitationally coupled circular rings of matter, each having the appropriate angular momentum, which always admits a discrete spectrum of bending modes. The real modes of the continuum disk can be distinguished by showing they are independent of the number of rings employed.

The gravitational restoring forces are correctly captured in this approach, but the lack of random motion omits the additional coupling between adjacent mass elements caused by the epicyclic motions of the stars. This extra mechanism further stiffens the disk's resistance to bending (Debattista & Sellwood 1999), especially in the high density inner regions where radial epicycles are larger.

Hunter & Toomre (1969) developed the coupled ring approach to study the bending dynamics of rotationally-supported, razor-thin disks with no random motion and

no halo. They were able to prove that general disks of this kind have no axisymmetric ( $m = 0$ ) or warping ( $m = 1$ ) instabilities, but they could not extend their proofs to higher sectoral harmonics. In fact they noted that were the disk composed of two equal counter-rotating populations of stars (on circular orbits) it would be buckling unstable for all  $m \geq 2$ .

Hunter & Toomre (1969) also studied the particular case of the sharp-edged Maclaurin disk in which all stars orbit at the same angular rate. When the sense of rotation is the same for all stars, the disk is stable to all bending waves, and there is a simple set of discrete neutral modes for all  $m$ . Polyachenko (1977) extended their analysis to Maclaurin disks with random motion (the Kalnajs disks) and was able to solve for the complete spectrum of bending modes. Needless to say, the introduction of random velocities in this razor-thin system caused buckling instabilities to appear for all  $m$ .

Since the Maclaurin disk bears little resemblance to real galaxies, Hunter & Toomre (1969) modified the disk to blur the sharp edge. They demonstrated that discrete warp modes in a cold, razor-thin disk can exist only when the edge is unrealistically sharp. Note that all isolated disks admit two trivial zero-frequency modes: a vertical **displacement** of the entire disk and a **tilt** of the disk plane to its original direction. More interesting standing wave solutions (modes) require traveling waves to reflect off the disk edge, but a realistic disk with a fuzzy edge does not reflect bending waves, because the group velocity decreases with the disk surface density (eq. 15) and a wave packet in a cold disk will never reach the edge (Toomre 1983).

### 8.3. Simulations of Buckling Modes

Sellwood & Merritt (1994) used  $N$ -body simulations to study the global instabilities of hot disks with no net rotation, *i.e.* with half the particles counter-rotating. (See §9.7 for buckling modes of more normal disks with large net rotation.) The form and vigor of the principal instabilities in any one of their models varied with the balance between radial and azimuthal pressure and with disk thickness: an in-plane lop-sided instability was the most disruptive for cool models (§5). The radially hotter thin disks were disrupted by axisymmetric bending instabilities (bell modes), as illustrated in Fig. 11. The instability creates a thick inner disk resembling a pseudo-bulge (see chapter by Kormendy). Very thin disks also buckled in an  $m = 2$  “saddle” mode, and an  $m = 1$  warp instability was detectable in some models, but never dominant.

Remarkably, instabilities in a counter-streaming model having intermediate radial pressure caused rather mild changes and led to an apparently stable, moderately thin, and almost axisymmetric disk. The in-plane velocities in this model resemble those reported by (Rubin *et al.* 1992) for the S0 galaxy NGC 4550, and indicate this galaxy could be stable even without large quantities of dark matter. The stability of this end product also demonstrated that thin axisymmetric systems with modest radial pressure are more stable to bending modes than those having isotropic or radially biased DFs.

Merritt & Sellwood (1994) proposed a criterion for the stability of a stellar system to buckling modes that can

<sup>13</sup> A razor thin disk is not destabilized by orbital motions with no velocity spread.

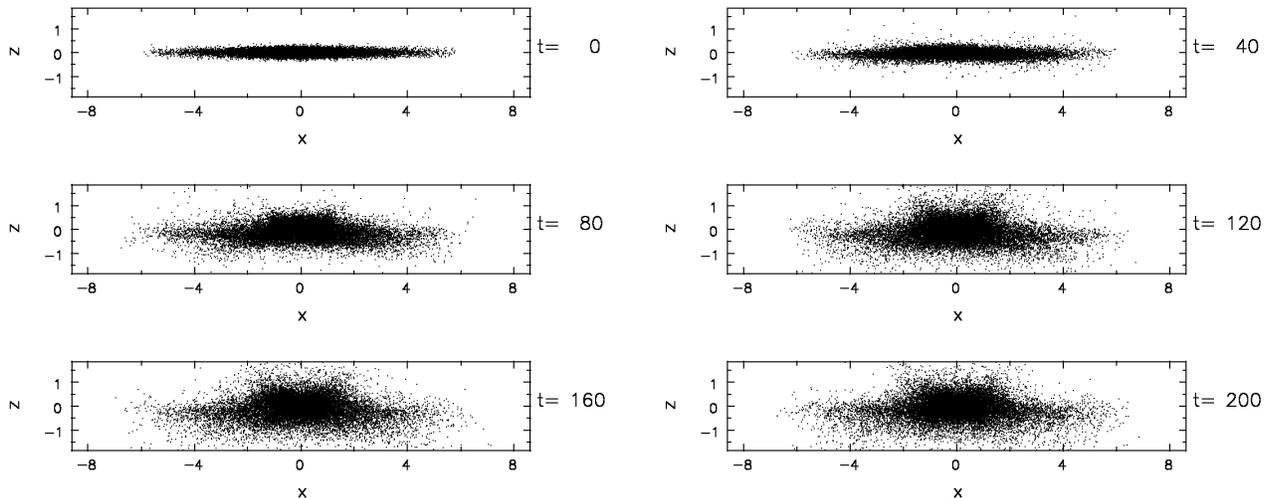


FIG. 11.— The development of a buckling instability in a simple model of an isolated stellar disk with  $Q \sim 2$ . The disk is KT/4 model described by Sellwood & Merritt (1994), with equal numbers of particles orbiting in each direction, but here the simulation uses a code of much higher spatial resolution. The buckling mode is axisymmetric and, while non-axisymmetric features were permitted by the code, none developed. The disk mass and radial scale are unity, and  $G = 1$ ; the orbit period at  $R = 2$  is 16 in these units.

be applied globally. A particle moving at speed  $\dot{x}$  in a mildly bent sheet with characteristic wavenumber  $k$  experiences a periodic vertical forcing at frequency  $k\dot{x}$  and, like any harmonic oscillator, the phase of its response depends upon whether the forcing frequency is greater or less than its natural vertical frequency. If  $k\dot{x} < \nu$ , the density response of the system to an imposed perturbation will be supportive, and the disturbance can be sustained or even grow. However, if  $k\dot{x} > \nu$  for most particles, the overall density response to the perturbation will produce a potential opposite to that imposed and the disturbance will be damped. They showed that their proposal successfully accounted for their  $N$ -body results, for the behavior of the infinite slab at short wavelengths, and for the apparent absence of elliptical galaxies more flattened than E7.

Sellwood (1996b) found a long-lived bending oscillation in an  $N$ -body simulation of a warm disk that was constrained to be axisymmetric. Apparently the system was able to support a standing wave, at most mildly damped, between the center and the edge of the disk, even though the stars had random motion in a disk of finite thickness. The frequency of the bending or flapping mode was low enough to avoid vertical resonances with almost all disk particles, in agreement with the requirement stated by Mathur (1990). While this result remains an isolated curiosity, since the flapping mode would probably be quickly damped in a halo, it is a clear counter-example to the argument by Hunter & Toomre (1969) that realistic disks do not possess global bending modes, which apparently holds only for disks without random motion.

#### 8.4. Disks in Halos

Since the paper by Hunter & Toomre (1969), ideas of warp formation have relied in some way on the interaction between the disk and its dark matter halo. Dekel & Shlosman (1983) and Toomre (1983) suggested that a flattened halo misaligned with the disk could form a long-lasting warp. Sparke & Casertano (1988) and Kuijken (1991) obtained long-lived warps (dubbed **modified tilt modes**)

of disks in rigid, misaligned halos, which were insensitive to the details of the disk edge. Lovelace (1998) studied the tilting dynamics of a set of rings also in a rigid halo, but assumed that the inner disk lay in the symmetry plane of the spheroidal halo.

Dubinski & Chakrabarty (2009) noted that the dark matter halos that result from cosmic structure formation simulations are usually aspherical, with frequent misalignments between the principal axes of the inner and outer halo. The disks in their simulations warped nicely when forced with slowly rotating, but otherwise rigid, perturbing fields representative of such halos.

However, dark matter halos are not rigid, and a responsive halo alters the dynamics in several ways. Nelson & Tremaine (1995) showed that were the inner disk misaligned with the principal plane of the flattened halo, as supposed by Sparke & Casertano (1988), its precession would be damped through dynamical friction, bringing the disk into alignment with the halo on time scales much shorter than a Hubble time. But a more compelling objection to the modified tilt mode emerged from  $N$ -body simulations with live halos: Dubinski & Kuijken (1995) found that the warp did not survive while Binney *et al.* (1998) showed that the inner halo quickly aligns itself with the disk, not *vice versa*. The large store of angular momentum in the disk maintains its spin axis, but the pressure supported inner halo can readily adjust its shape slightly to align itself with the disk.

#### 8.5. Misaligned Infall

The idea that galaxy warps are manifestations of *eternal* warp modes seems doomed by the damping effect of a live halo. But slowly evolving warps remain viable, provided that suitable external perturbations occur in enough cases.

In hierarchical galaxy formation scenarios, late infalling material probably has an angular momentum axis misaligned with the disk spin axis. Ostriker & Binney (1989) therefore proposed that warps arise due to the slewing of the galactic potential as material with misaligned angular

momentum is accreted. Structure formation simulations by Quinn & Binney (1992) confirmed that the mean spin axis of a galaxy must slew as late arriving material rains down on the early disk. The less-than-critical matter density in modern  $\Lambda$ CDM universe models implies that infall is less pervasive at later times, but it manifestly continues to the present day in gravitationally bound environments (Sancisi *et al.* 2008, and chapter by van Woerden & Bakker).

Jiang & Binney (1999) and Shen & Sellwood (2006) presented results of experiments in which a disk was subjected to the torque from a misaligned, massive torus at a large radius. This well-defined perturbation is a very crude model of an outer halo that is rotationally flattened, and having its spin axis misaligned with that of the disk. It is misaligned and farther out because, in hierarchical scenarios, the mean angular momentum of the later arriving outer halo is probably greater and misaligned from that of the original inner halo and disk. The accretion axis is, in reality, likely to slew continuously over time, so a model with a constant inclination is somewhat unrealistic.

Rather than striving for realism, Shen & Sellwood (2006) used this simple forcing to reach an understanding of how the warp develops and why the LoN usually forms a loosely-wound, leading spiral. The inner disk maintained a coherent plane because of both self-gravity and random motion. The torque arising from the misaligned outer torus caused the inner disk, and the aligned inner halo, to precess rigidly even though the torque increased with radius, but the outer disk beyond  $\sim 4R_d$  started to warp.

As soon as the outer disk became misaligned with the inner disk, the strongest torque on the outer parts of the disk arose from the inner disk. The torque from the interior mass was responsible for the leading spiral of the line of nodes, even though the adopted external field would have produced a trailing spiral. The fact that the LoN of most warps forms a leading spiral over an extended radial range seems to imply massive disks.

Even though the disk precessed due to the external torque, its motion was barely damped over many Gyr, in contrast to the expectations from Nelson & Tremaine (1995). Damping was weak because the slow precession rate allowed the inner halo to remain closely aligned with the disk, which therefore caused little drag. The weak damping seemed to be caused more by the relative precession of the inner and outer parts of the halo. Also the warp evolved slowly as the layer settled to the main plane at gradually increasing radii, in apparent agreement with the decreasing outward group velocity.

Fig. 12 shows the H I observation of NGC 4013 by Bottema (1996), together with the warp obtained in the simulation by Shen & Sellwood (2006). Their simulations revealed that the warp persisted for cosmologically interesting times, even when the external forcing field was removed. Thus the persistence of warps is not nearly as perplexing as previous studies had suggested. Furthermore, the model had a flat inner disk and the warp in the outer disk matched all of Briggs’s rules quite well.

A fixed outer torus is clearly unrealistic and the halo axis probably shifts continuously or episodically, as argued by Quinn & Binney (1992), making warp lifetimes a side issue. Warps formed this way can be repeatedly regenerated

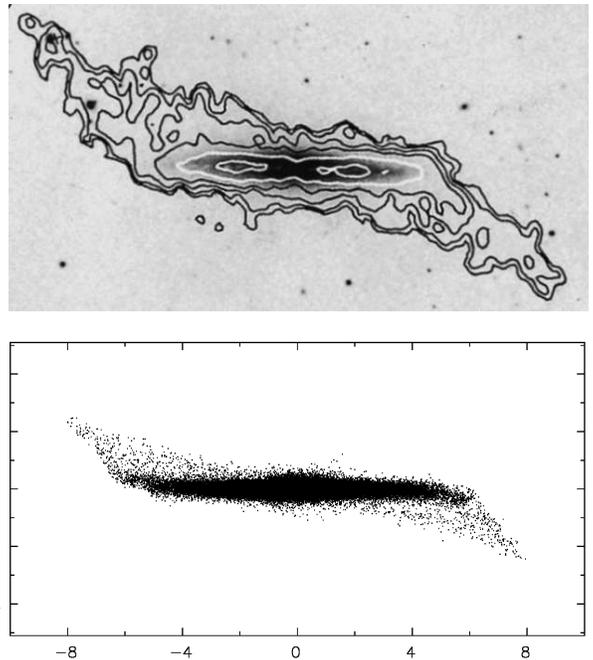


FIG. 12.— The upper panel shows the observed H I warp of NGC 4013 (Bottema 1996), and the lower panel the warp in the simulation by Shen & Sellwood (2006) that closely resembles it. The length unit on the axes is the scale length  $R_d$  of the exponential disc.

when a new infall event happens. Since cosmic infall and mergers are more likely to happen in a denser environment, warps can be induced more frequently in such an environment, which is consistent with the statistics (García-Ruiz *et al.* 2002b).

### 8.6. Warps Driven by Tides

Tidal interactions between galaxies could be an additional cause of warps in disks, and one that is quite likely to produce asymmetrical warps. This idea has been explored most fully to explain the warp of the Milky Way’s (MW) disk that results from the proximity of the Magellanic Clouds, especially the Large Cloud (LMC). The orientation of the principal axes of the warp (Levine *et al.* 2006) at least seems favorable to this hypothesis. Bailin (2003) suggested that the Sagittarius dwarf galaxy is another possible culprit.

Hunter & Toomre (1969), in models that did not include dark matter, concluded that the LMC could be responsible for the warp of the MW, but only if it was a lot more massive than was then suspected and had recently passed close to the edge of the disk. García-Ruiz *et al.* (2002a) tested the hypothesis in fully self-consistent simulations that included live halos and used updated information about the distance and motion of the LMC. They concluded that neither the amplitude, nor the orientation, of the warp in the disk of the MW was consistent with the tidal hypothesis. Weinberg & Blitz (2006), on the other hand, suggested that a co-operative response from the halo to the passage of the LMC could have generated the observed warp in the disk. Thus there is no consensus yet on the origin of the warp in the Milky Way.

## 9. BARS

Bars are common features in disk galaxies. An earlier review (Sellwood & Wilkinson 1993) of the vast topic of bars is now somewhat dated but, as it provides a still useful summary of the basic dynamics, the present article will update a few main points and the reader is referred to the earlier review for a more detailed discussion.

### 9.1. *Origin of Bars*

For a long time, dynamicists were struggling to understand the absence of bars in some disk galaxies (Ostriker & Peebles 1973), but since reasonable models of luminous galaxies that include a dense bulge of moderate mass are now known to be stable (§4), the problem has become almost the opposite! Strong bars are seen in many galaxies whose mass distributions now appear unfavorable to the dynamical bar instability, as evidenced by a nuclear gas ring (see §9.4), which can form only if the center is dense enough to have inhibited bar formation by Toomre’s mechanism.

However, the fact that the center is dense today does not require that it was dense when the bar formed; secular inflow of gas (§9.4) can build up the central density after the bar has formed. Alternatively, the bar could have grown in size and slowed through disk evolution (Sellwood 1981; Berentzen *et al.* 2007) or halo friction (Athanasoula 2002, see §9.8), or have been triggered by large density fluctuations in the disk (Sellwood 1989a), by tidal interactions (Noguchi 1987; Berentzen *et al.* 2004; Curir *et al.* 2006), halo substructure (Romano-Díaz *et al.* 2008b) or a non-axisymmetric halo (Dubinski & Chakrabarty 2009). Any of these considerations, or that described in the next paragraph, could plausibly reconcile the existence of bars in these galaxies with Toomre’s stabilizing mechanism.

In an elegant piece of dynamics, Lynden-Bell (1979) demonstrated that the inner parts of galaxies are regions where eccentric orbits have a tendency to align themselves, which allows a bar to grow slowly through orbit trapping. The region where a cooperative response to a mild perturbation occurs is where the overall radial density profile of the galaxy flattens into a more uniform core. In this region, the radial variation of  $\Omega - \kappa/2$  has a maximum at some non-zero radius and an infinitesimal bar pattern can have a pattern speed that allows two ILRs. Lynden-Bell’s aligning mechanism, which operates best on the more eccentric orbits, requires  $\Omega_p \simeq \Omega_\phi - \Omega_R/2$  of orbits in the aligning region. As this pattern speed is much lower than that expected from the global bar instability (§4.1), the aligning mechanism offers an additional route to bar formation in otherwise globally stable disks. Although the cooperative region has a small radial extent, Lynden-Bell (1979) suspected the bar could be much larger.

Erwin (2005) presented a useful study of bar properties and pointed out that bars in late-type galaxies are often much smaller relative to the disk size than are those formed in simulations.

### 9.2. *Frequency of Bars*

Strong bars are clearly visible in 25% – 30% of disk galaxies (*e.g.* Masters *et al.* 2010), and the fraction rises to  $\gtrsim 50\%$  when more objective criteria are applied to red or near-IR images (Eskridge *et al.* 2000; Marinova & Jogee

2007; Reese *et al.* 2007, and further references cited below). Barazza *et al.* (2008) found a higher bar fraction in later Hubble types, while Méndez-Abreu *et al.* (2010) found no bars in either very luminous, or very faint, galaxies.

Whatever may be the mechanism responsible for the formation of bars in real galaxies, none of the above suggestions makes a clear prediction for the frequency of bars. Bosma (1996), Courteau *et al.* (2003), and others have pointed out that barred galaxies seem little different from their unbarred cousins in most respects, *e.g.* they lie on the same Tully-Fisher relation. Minor systematic differences do exist: *e.g.* Davoust & Contini (2004) note that barred galaxies seem to have smaller mass fractions of neutral H I gas, but this seems more likely to be the result, rather than the cause, of the bar. The anti-correlation of bar frequency with the bulge half-light (Barazza *et al.* 2008) possibly results from Toomre’s stabilizing mechanism, but this cannot be the whole story because some near-bulgeless disks are unbarred while other barred disks have massive bulges.

If no dynamical property, other than their eponymous one, can be identified that cleanly separates barred from unbarred galaxies, then the existence of a bar in a galaxy may possibly be determined by external factors, such as a chance encounter. Elmegreen *et al.* (1990) reported an increased fraction of bars in groups and clusters, but more recent work (Barazza *et al.* 2009; Li *et al.* 2009; Aguerri *et al.* 2009) has found little or no variation of bar fraction with environment. It is also possible that the bar fraction could be changing with time; different groups disagree (Jogee *et al.* 2004; Sheth *et al.* 2008), probably because observations of galaxies at significant look-back times are subject to systematic difficulties due to band-shifting and changing spatial resolution (see also Elmegreen *et al.* 2007).

A radical alternative is to regard bars as transient features that form and decay, and that the current fraction of barred galaxies represents the duty cycle (Bournaud *et al.* 2005). Bars in  $N$ -body simulations are dynamically rugged objects that appear to last indefinitely. Of course, they could be destroyed by a merger event, for example, although not much in the way of a cool disk would survive such an event. Norman and his co-workers (Pfeniger & Norman 1990; Norman *et al.* 1996) have argued that bars can be destroyed by the accumulation of mass at their centers, which may lead to a pseudo-bulge and/or a hot inner disk. However, Shen & Sellwood (2004) and Athanasoula *et al.* (2005) found that unreasonably large and/or dense mass concentrations were required to cause their bars to dissolve. The simulations by Bournaud *et al.* (2005) uniquely show that gas accretion may aid the dissolution of the bar and, with star formation, could recreate a cool disk that would be needed to make a new bar. Even if this behavior can be confirmed by others, the model requires very substantial gas infall. Moreover, the continued existence of the hot old disk and the build-up of a dense center makes every cycle of this speculative picture harder to achieve.

In the distant future, galaxy formation simulations (see chapter by Steinmetz) may have the quality and resolution perhaps to be able to predict the correct bar fraction, and thereby reveal their cause.

### 9.3. Structure of Bars

This section gives a brief description of a few important aspects of the orbital behavior in large-amplitude bars, and the reader is referred to Sellwood & Wilkinson (1993) for a more comprehensive discussion. Weak bars can be treated using epicycle theory (BT08 §3.3.3). Most early orbit studies in strongly barred potentials considered motion confined to the plane perpendicular to the rotation axis. Even though 3D motion is much richer, the fundamental structure of bars is most easily understood from in-plane orbits.

Since bars are believed to be steadily-rotating, long-lived objects, it makes sense to discuss their structure in a frame that co-rotates with the potential well at the angular rate  $\Omega_p$ . A rotating frame has the **effective potential**

$$\Phi_{\text{eff}} = \Phi - \frac{1}{2}\Omega_p^2 R^2, \quad (16)$$

where  $\Phi$  is the potential in an inertial frame.

The effective potential surface in the disk plane (eq. 16) resembles a volcano, with a central crater, a rim, and a steeply declining flank. The crater is elongated in the direction of the bar, and the rim has four **Lagrange points**: two maxima,  $L_4$  and  $L_5$ , on the bar minor axis and two saddle points,  $L_1$  and  $L_2$  on the bar major axis. (The fifth Lagrange point,  $L_3$ , is the local potential minimum at the bar center.) Because of Poisson's equation, the density contours of the bar must be more elongated than those of the inner  $\Phi_{\text{eff}}$ .

Neither  $E$  nor  $L_z$  is conserved in non-axisymmetric potentials, but Jacobi's invariant  $I_J$  (eq. 6) is conserved even for strong bars that rotate steadily. Since  $I_J = \frac{1}{2}|\mathbf{v}|^2 + \Phi_{\text{eff}}$ , where  $\mathbf{v}$  is the velocity in the rotating frame, contours of  $\Phi_{\text{eff}}$  bound the possible trajectories of stars having  $I_J$  less than that contour value. Stars that are confined to the bar, which also have  $I_J < \Phi_{\text{eff}}(L_1)$ , are of most interest here.

A **periodic orbit** is a possible path of a star in the rotating frame that retraces itself, usually after a single turn around the center, but always after a finite number of turns. Because the orbits close, the orbital period in the rotating frame is commensurable with the radial period and these orbits are also described as resonant orbits of the (strongly non-axisymmetric) potential.

Periodic orbits can be either **stable**, in which case a star nearby in phase space oscillates (librates) about its **parent** periodic orbit in an epicyclic fashion, or they are **unstable**, in which case the trajectory of a star nearby in phase space diverges exponentially (at first) from the periodic orbit. The orbits of stars that librate around a periodic orbit are known as **regular** orbits, those that diverge from it are known as **irregular** or **chaotic** orbits. Chaotic orbits have only a single integral,  $I_J$ , while regular orbits have an additional integral (two more in 3D) that confines their motion to a hypersurface of smaller dimension in phase space. Regular orbits are the more interesting because the star's orbit can be more elongated than the potential surface that confines it, which is of great value when building a self-consistent bar model.

The main features can be illustrated in the simple potential (*cf.* BT08 eq. 3.103)

$$\Phi_{\text{eff}}(x, y) = \frac{1}{2}v_0^2 \ln \left( 1 + \frac{x^2 + y^2/q^2}{R_c^2} \right) - \frac{1}{2}\Omega_p^2 R^2, \quad (17)$$

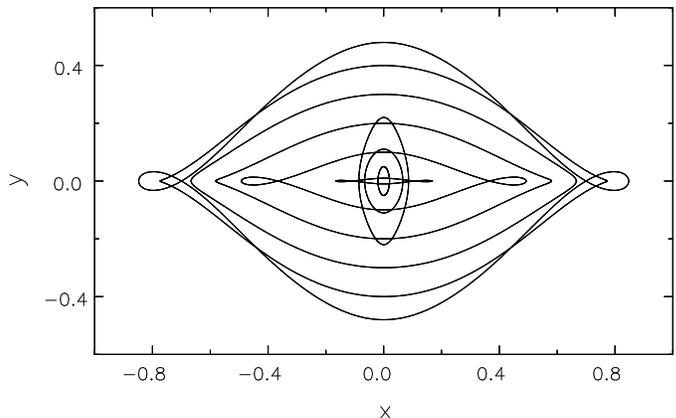


FIG. 13.— Examples of important periodic orbits in a rotating bar potential that close after two radial oscillations for every turn about the center, the 2:1 resonant families. Those orbits elongated parallel to the bar axis (horizontal) are members of the  $x_1$  family. The  $x_2$  orbits are elongated perpendicular to the bar.

where  $R^2 = x^2 + y^2$ ,  $R_c$  is a core radius inside of which the potential is approximately harmonic,  $q \leq 1$  is the flattening, and  $v_0$  is the circular speed at large  $R$  when  $q = 1$ . As in BT08, the values are:  $v_0 = 1$ ,  $q = 0.8$ ,  $R_c = 0.03$ , and  $\Omega_p = 1$ . With these parameters, the major-axis Lagrange points lie at a distance  $R_L \simeq 0.9996$  from the bar center.

The 2:1 resonant periodic orbits shown in Fig. 13 have a range of  $I_J$  values, but all close in the rotating frame after two radial oscillations for every turn about the center. Orbits of the main family, denoted  $x_1$ , are elongated parallel to the bar (horizontal in the Figure), and are referred to as the **backbone of the bar**, since the majority are stable. Stars on these orbits all move in the same direction as the bar rotates, but have shorter periods because they are interior to the Lagrange points (or more loosely, they are inside CR). Some  $x_1$  orbits are simple closed figures, but others have loops near the outer ends where the star progresses around the center of the galaxy more slowly than does the rotating frame. Sparke & Sellwood (1987) and Voglis *et al.* (2007) found that a large fraction of particles in an  $N$ -body bar librate around them, and the same behavior is expected for stars in real bars.

Another important family of 2:1 orbits appears near the center of the bar, but is elongated perpendicular to the bar axis, as also shown in Fig. 13. This family, denoted  $x_2$ , is almost always stable and is present in many realistic bar potentials. Since Lindblad resonances (§2.4) are defined for infinitesimal perturbations to axisymmetric potentials, the label ILR is very loose usage in strong bars. It is true that the  $x_2$  family appears in barred potentials only, but not always, when the axisymmetric mass distribution and pattern speed admit one or more ILRs. Furthermore,  $x_2$  orbits orient themselves perpendicular, while the  $x_1$  family is parallel, to the bar axis, which is directly analogous to the abrupt phase shift that occurs in the response of a forced harmonic oscillator as the forcing frequency crosses its natural frequency (*cf.* Sanders & Huntley 1976). At finite amplitude, the near circular orbit sequence acquires gaps at the 2:1 resonance bifurcations, which become broader as the bar amplitude rises. Thus the radial extent of the  $x_2$  family shrinks as the strength of the bar is increased, and it may disappear entirely. Even if

one is careful to say that the appearance of the  $x_2$  family is the generalization of the ILR to strong bars, the radius of this resonance is still badly defined because the orbits that appear inside it can be quite eccentric.

As  $I_J$  approaches the value of  $\Phi_{\text{eff}}(L_1)$ , the time to complete a full turn in the rotating frame lengthens and additional orbit families appear. Orbits that close after any number of radial oscillations can be found in principle, but of these only the 4:1 resonant orbits are of dynamical significance to bar structure. As the period lengthens, the proliferation of orbit families causes a precipitous decrease in the stable regions around each parent and chaotic behavior ensues (Chirikov 1979). The onset of chaos near CR led Contopoulos (1980) to expect the density of a self-consistent bar to drop steeply near the major axis Lagrange points, leading to the rule that the length of a bar is limited by CR. This rule predicts that the parameter (Elmegreen 1996)

$$\mathcal{R} \equiv R_L/a_B > 1, \quad (18)$$

where  $a_B$  is the semi-major axis of the bar. In principle, self-consistent bar dynamics could allow bars with  $\mathcal{R} \gg 1$ , although empirical bar pattern speed estimates (§9.5) mostly find that CR is in fact only slightly beyond the end of the bar.

There are many more in-plane orbit families, but few are of dynamical importance to the structure of the bar. See Sellwood & Wilkinson (1993) for a fuller account.

The extension to 3D allows for many more resonances between the in-plane motion and the vertical oscillations. While there is a rich variety of behavior (Pfenniger & Friedli 1991; Skokos *et al.* 2002), the backbone  $x_1$  family from 2D continues to be the most important, but now with a “tree” of orbits also librating vertically. The new periodic orbits that appear in 3D have a similar projected shapes as the in-plane  $x_1$  family, but they also librate vertically a small number of times over the same period as the motion in the plane. Patsis *et al.* (2002) highlighted the orbit families that they found to be of importance for the “boxy” appearance of edge-on bars (see also §9.7).

#### 9.4. Gas Flow

When pressure and magnetic forces can be neglected, any mild dissipation will drive gas to move on stable periodic orbits. An organized streaming gas flow pattern is expected wherever the simplest periodic orbits over a range of energies can be nested and intersect neither with neighboring orbits, nor with themselves. Shocks, where pressure ceases to be negligible, must occur in flows either where periodic orbits self-intersect, or where gas flows on two separate orbits cross. Fig. 13 shows that were gas to flow in that adopted bar potential, shocks would be inevitable because many orbits self-intersect (the loops) and, in particular,  $x_2$  orbits cross the  $x_1$  family. Thus shocks are a general feature of cool (low pressure) gas flows in bars.<sup>14</sup>

Full hydrodynamic simulations (*e.g.* Roberts *et al.* 1979; Athanassoula 1992; Fux 1999) are needed to determine the flow pattern. Shocks are offset to the leading sides of the bar major-axis in models having reasonable parameters.

<sup>14</sup> Shocks may be avoided when pressure is significant (Englmaier & Gerhard 1997).

Prendergast (1962) seems to have been the first to associate the dust lanes in bars with the locations of shocks in the gas.

Shocks convert some kinetic energy of bulk motion in the gas into heat, which is radiated efficiently. Furthermore, the offset location of the shock causes the gas to spend more than half its orbit on the leading sides of the bar, where it is attracted backwards towards the bar major axis, causing it to lose angular momentum.<sup>15</sup> Thus gas in the bar region must settle a little deeper into the potential well on each passage through a shock, *i.e.* the bar drives gas inwards, the angular momentum it loses being given to the bar.

The inflow stalls where gas settles onto the  $x_2$  orbit family, which is found in bar models that have dense centers, leading to a build up of gas. This behavior can be associated with nuclear rings of dense gas (Regan *et al.* 2002), which are often the sites of intense star formation also (Maoz *et al.* 2001; Benedict *et al.* 2002). If this were the whole story, the gas could not be driven any further inwards, but there is both observational evidence, in the form of spiral dust lanes and star formation (*e.g.* Carollo *et al.* 1998), and some theoretical work (Wada 2001) to suggest that self-gravity causes inflow to continue. However, the existence of high gas density and rapid star formation in the nuclear ring indicates that only a small fraction of the gas continues inwards.

Modeling the gas flow in a specific galaxy allows one to determine two properties of the bar that are hard to constrain otherwise. Estimating the gravitational potential of the galaxy from a photometric image plus a dark halo, Weiner *et al.* (2001) fitted for both the disk mass-to-light (M/L) and  $\Omega_p$  in the galaxy NGC 4123. These authors also described the procedure in detail. Results for a number of other galaxies were reported by Pérez *et al.* (2004), Weiner (2004), and Zánmar Sánchez *et al.* (2008), although these last authors were unable to obtain an entirely satisfactory fit. Lindblad *et al.* (1996) fixed the M/L and fitted only for  $\Omega_p$ . All these fits preferred rapidly rotating bars in heavy disks. Pérez (2008) confirmed that the best fit parameters of M/L and  $\Omega_p$  were the same for both 2D Eulerian (Godunov) and 3D Lagrangian (SPH) hydrodynamic methods.

Even though the quadrupole field of a bar decays quickly with radius, it can be strong enough to drive a spiral shock in the gas of the outer disk, as originally demonstrated by Sanders & Huntley (1976). Schwarz (1981) showed that when gas is modeled as inelastic particles, it is driven outwards to form an outer ring (see review by Buta & Combes 1996). However, it is unclear that spirals in the outer disks of real barred galaxies are the responses to the bar, and they may owe more to self-excited structures than to bar driving (§7, Sellwood & Sparke 1988; Buta *et al.* 2009). In addition, the outer spiral response to an imposed bar is not steady in modern simulations, with the shapes of the driven arms cycling through a broad range. For these reasons, the gas flow models fitted to individual galaxies should rely primarily on the fit within the bar and pay little attention to the outer disk.

Kranz *et al.* (2003) tried a similar approach, but fitted

<sup>15</sup> The opposite happens in shocks outside CR, where the gas gains angular momentum from the bar.

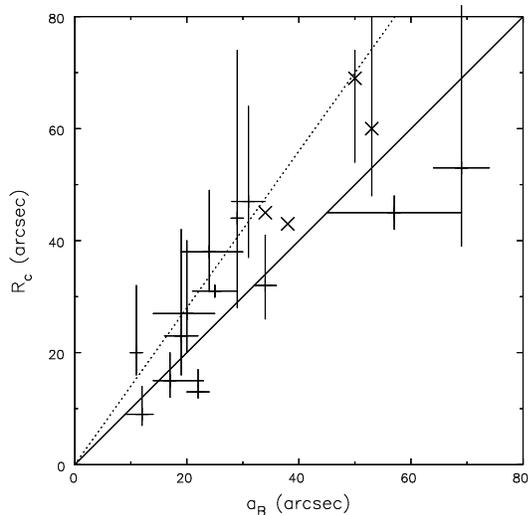


FIG. 14.— Summary of direct pattern speed measurements for bars collected by Corsini (2008). The diagonal line shows  $\mathcal{R} = 1$  and the dotted line  $\mathcal{R} = 1.4$ . The crosses mark values for which error bars are unavailable.

a spiral pattern instead of a bar, which may yield unreliable results for two reasons: (i) The lifetimes of spiral arms are believed to be short (§7) leading to broader resonances and stronger gas responses than would arise in simulations that assume a slowly evolving pattern, and (ii) the observed spirals could be the superposition of several features with different angular rotation rates. As bars undoubtedly last for longer than do spirals and dominate the non-axisymmetric potential, fits to bars in galaxies are likely to yield better disk mass estimates.

### 9.5. Bar Pattern Speeds

Tremaine & Weinberg (1984a) devised a method to measure the pattern speed of a bar directly from observations of a tracer component, which must obey the equation of continuity. Their original method assumes that the galaxy has but a single pattern, and would yield a misleading result were there more than one pattern, each rotating at a different angular rate.

The stellar component of early-type barred galaxies is believed to obey the equation of continuity because these galaxies have little dust obscuration and no star formation. They also rarely possess prominent spirals in the outer disk. Results of many studies using this method for early-type barred galaxies were summarized by Corsini (2008) and are shown in Fig. 14. While some individual measurements are quite uncertain, the data seem to favor  $1 < \mathcal{R} \lesssim 1.4$ . Chemenin & Hernandez (2009) found a counter-example in a low-luminosity galaxy.

Fathi *et al.* (2009) and Meidt *et al.* (2009) applied the method of Tremaine & Weinberg (1984a) to ionized and molecular gas, respectively. Both groups argue that this is valid, even though the separate gas components do not obey the continuity equation that underlies the method. Fathi *et al.* (2009) generally find fast bars. Meidt *et al.* (2008) generalized the method to attempt to measure radial variations in the pattern speed and Meidt *et al.* (2009) found suggestions of pattern speeds that are lower at large radii

than those near the center.

Other methods can yield indirect estimates of bar pattern speeds. Fits of models of the gas flow (§9.4) have been reported for a few galaxies, finding  $\mathcal{R} \sim 1.2$ . Athanassoula (1992) argued that the shapes and locations of dust lanes in bars also seem to suggest that  $\mathcal{R} \simeq 1.2$ . Associating a ring in a barred galaxy with the location of a major resonance with the bar (Buta & Combes 1996) yields, with kinematic information, an estimate of the pattern speed.

Rautiainen *et al.* (2008) computed models of the stellar and gas (sticky particles) responses to forcing by photometric models of 38 barred galaxies, in which they assumed that the entire non-axisymmetric structure rotated at the same pattern speed. They attempted to match the model to the visual morphology of the galaxy, and found a range of values for  $\mathcal{R}$ . However, in most cases where  $\mathcal{R} \gg 1$ , the fit is dominated by the outer spiral, which may have a lower angular speed than does the bar.

### 9.6. Bars Within Bars

Erwin & Sparke (2002) and others have found inner **secondary bars** within the inner parts of  $> 25\%$  of large-scale or **primary bars**. They reported that the secondary bar has a length some  $\sim 12\%$  of that of the primary bar and the deprojected angles between the principal axes of the two bars appeared to be randomly distributed, suggesting that the two bars may tumble at differing rates. This inference was supported by Corsini *et al.* (2003), who used the Tremaine & Weinberg (1984a) method to show that the two bars in NGC 2950 could not have the same rotation rates; Maciejewski (2006) used their data to argue that the secondary bar has a large retrograde pattern speed.

The theoretical challenge presented by these facts is substantial, and progress to understand the dynamics has been slow. Maciejewski & Sparke (2000) studied the orbital structure in a potential contain two nonaxisymmetric components rotating at differing rates. However, it is almost certain that the secondary bar can neither rotate at a uniform rate (Louis & Gerhard 1988) nor can it maintain the same shape at all relative phases to the primary.

Friedli & Martinet (1993) argued that gas was essential to forming secondary bars (see also *e.g.* Heller *et al.* 2001; Englmaier & Shlosman 2004). However, some of the collisionless simulations reported by Rautiainen & Salo (1999) and Rautiainen *et al.* (2002) manifested dynamically decoupled inner structures when the inner disk had high orbital frequencies due to a dense bulge. The structure was more spiral like in some models, but others appeared to show inner bars that rotated more rapidly than the main bar.

Debattista & Shen (2007) created long-lived, double-barred galaxy models in collisionless  $N$ -body simulations having dense inner disks, which they described as pseudo-bulges. They followed up with a more detailed study (Shen & Debattista 2009) that also made some predictions for observational tests. The secondary bars in their models indeed rotated at non-uniform rates, while their shape varied systematically with phase relative to that of the primary. These models prove that collisionless dynamics can support this behavior, but it is unclear that their initial conditions mimicked those that have given rise to double

barred galaxies in nature.

The possible consequence of gas inflow in these galaxies has attracted a lot of attention. Shlosman *et al.* (1989) speculated that bars within bars might lead to gas inflow over a wide dynamic range of scales, from global to the parsec scale where accretion onto a black hole might cause AGN activity. While inflows may have been observed (*e.g.* Haan *et al.* 2009), understanding of gas flow in these non-steady potentials remains rather preliminary (Maciejewski *et al.* 2002; Heller *et al.* 2007).

### 9.7. Buckling of Bars

Combes & Sanders (1981) first reported that the bars in their 3D simulations were thicker than the disk from which they had formed, and had acquired a pronounced “boxy” shape when viewed edge-on. Boxy isophotes in edge-on disk galaxies are now believed to be an indicator of a bar, as is supported by kinematic evidence in the gas (Merrifield & Kuijken 1999; Bureau & Athanassoula 2005).

The reason the bar thickened was explained by Raha *et al.* (1991), who showed that bars are subject to the buckling instability (§8.1). The bar buckles because the formation of the bar created a structure supported by elongated orbits that stream along the bar in the near radial direction. Even though the ingoing and outgoing stars stream on different sides of the bar, the effective averaged  $\sigma_R$  has risen as a result without changing  $\sigma_z$ . The simulation by Raha *et al.* (1991) revealed that the buckling instability produced a large amplitude arch just before it saturated, after which the bar became thicker. The energy to increase vertical motion in the bar appeared to have been released by the further concentration of mass towards the bar center (see also 2004). It is delightful that the evolution of one instability, the bar-forming mode, should create a new structure, the bar, that is itself unstable.

Bars still thicken in more recent simulations with grids having higher spatial resolution (*e.g.* Fig. 6), but do not seem to exhibit the spectacular arch reported by Raha *et al.* (1991) unless bi-symmetry is enforced. Low spatial resolution or significant gravity softening (which are equivalent) weakens the restoring force in eq. (13) and artificially increases  $\lambda_J$ , which is the characteristic length for instability. Stronger gravity causes the preferred buckling modes to have shorter wavelength allowing, say, an upward arch on one side of the center and downward arch on the other. Enforcing bi-symmetry prevents the bar from bending in this anti-symmetric manner, and forces it to buckle through the single arch mode.

### 9.8. Dynamical Friction on Bars

Friction between a rotating bar and a massive halo was first reported many years ago (Sellwood 1980), but the implications for dark matter halos have fueled a renewed intense study of the topic.

#### 9.8.1. Theory

Dynamical friction (Chandrasekhar 1943) is the retarding force experienced by a massive perturber as it moves through a background sea of low-mass particles. It arises, even in a perfectly collisionless system, from the vector sum of the impulses the perturber receives from the particles as they are deflected by its gravitational field (see

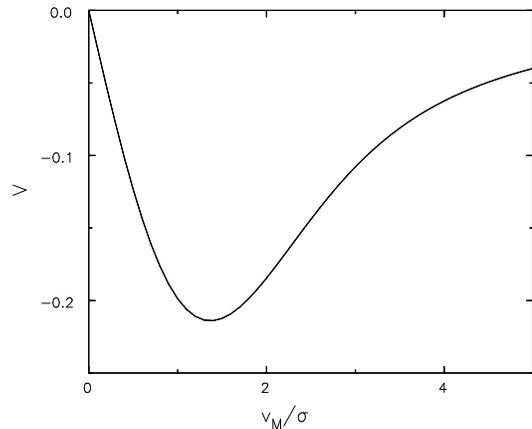


FIG. 15.— The dimensionless acceleration function  $V$  defined in eq. (19) for the case of a Gaussian distribution of velocities among the background particles. The function is negative because the acceleration is directed oppositely to the velocity.

Appendix). Equivalently, friction can be viewed as the gravitational attraction on the perturber of the density excess, or wake, that develops behind it as it moves, as was nicely illustrated by Mulder (1983).

Chandrasekhar’s formula (BT08 eq. 8.6) for the acceleration of a perturber of mass  $M$  moving at speed  $v_M$  through a uniform background, density  $\rho$ , of non-interacting particles having an isotropic velocity distribution with a 1D rms velocity spread  $\sigma$ , may be written as

$$\frac{dv_M}{dt} = 4\pi \ln \Lambda G^2 \frac{M\rho}{\sigma^2} V\left(\frac{v_M}{\sigma}\right). \quad (19)$$

The Coulomb logarithm is defined in the Appendix, and the dimensionless function  $V$  is drawn in Fig. 15 for a Gaussian distribution of velocities; other velocity distributions would yield a different functional form. Physically, the retarding acceleration must vanish when the perturber is at rest and it must also tend to zero when the perturber moves so fast that the background particles receive only small impulses and the feeble wake lies far downstream from the perturber. Friction is strongest when the speed of the perturber is somewhat greater than the rms speeds of the background particles.

The simplifying assumptions in its derivation invalidate application of eq. (19) to the physically more interesting problem of friction in a non-uniform medium in which the background particles are confined by a potential well and interact with the perturber repeatedly.

Repeated encounters between the perturber and the background particles require the more sophisticated treatment presented in Tremaine & Weinberg (1984b), who adopted a rotating potential perturbation in a gravitationally bound spherical halo of test particles. They showed, as did Lynden-Bell & Kalnajs (1972) for spiral waves, that lasting changes to the orbits of the halo particles appear to second order in the perturbing potential, and can occur only at resonances. They derived a daunting formula for the torque on the halo caused by the perturbation that sums the contributions from infinitely many resonances. The contribution at each resonance is proportional to the gradient of the DF near the phase-space location of the resonance, in a manner that is directly analogous to Lan-

dau damping. Weinberg (1985) computed the surprisingly large torque expected on a rotating bar, and his conclusion was confirmed in restricted tests (Little & Carlberg 1991; Hernquist & Weinberg 1992).

Weinberg & Katz (2007) pointed out that friction is dominated by a few important resonances. They estimated the widths of these resonances for a perturbation having constant pattern speed and finite amplitude, and argued that immense simulations would be needed to populate the resonance with sufficient particles to capture the correct net response. However, the loss of angular momentum from the perturber causes its pattern speed to change, and the resulting time-dependence of the forcing frequency is a much more important factor in broadening the resonances; thus friction can in fact be captured correctly in simulations having moderate numbers of particles (Sellwood 2008a). Note that the pattern speed of an orbiting satellite rises as it loses angular momentum, while that of a bar usually decreases.

Despite the complicated language of resonant dynamics, the upshot is simply that the perturber induces a wake-like response in the halo, as was beautifully illustrated by Weinberg & Katz (2007, their Fig. 1). As for the infinite medium, friction can be thought of more simply as the torque between the perturber and the induced halo response. Sellwood (2006, his Fig. 2b) shows the lag angle between the forcing bar and the halo response, which is about  $45^\circ$  when friction is a maximum and gradually decreases to zero as the bar slows, until eventually friction ceases when the halo response co-rotates with the bar.

Lin & Tremaine (1983) for an orbiting satellite, and Sellwood (2006) for a rotating bar, demonstrated that the frictional drag on the perturbation scales with the mass of the perturber,  $M$ , the halo density,  $\rho$ , and the halo velocity dispersion,  $\sigma$  exactly as in eq. (19). Furthermore, the dimensionless function that describes the dependence on the angular speed of the perturber shares the general properties with  $V(x)$  that it is negative (for reasonable non-rotating halos), and must  $\rightarrow 0$  as  $x \rightarrow \infty$ , and that it should be  $\propto x$  as  $x \rightarrow 0$ . Including self-gravity in the halo response causes a further slight increase in friction, but does not otherwise change the behavior.

### 9.8.2. Halo Density Constraint

Fully self-consistent simulations of bar formation in a live halo by Debattista & Sellwood (1998, 2000) showed that strong bars are indeed slowed rapidly. The fact that observed bars appear not to have been slowed (§9.5) may imply an upper bound to the density of the dark matter halo in barred disk galaxies. Valenzuela & Klypin (2003) claimed a counter-example of a bar that does not experience much friction in a “cosmologically-motivated” halo, even though their result disagreed with all others for strong bars (O’Neill & Dubinski 2003; Athanassoula 2003) and with theory!

Investigation of their anomalous result by Sellwood & Debattista (2006) revealed that friction can be avoided *temporarily* if the gradient of the DF at the most important resonance(s) has been flattened by earlier evolution, which they described as a metastable state. Lin & Tremaine (1983) reported similar behavior as a result of driving the perturber at constant frequency for a protracted period.

They showed, as did Sellwood & Debattista (2006) and Villa-Vargas *et al.* (2009), that the full frictional drag resumes after some delay, the duration of which seems to vary stochastically (Sellwood & Debattista 2009). Delayed friction can happen only in simulations of disks in isolated, smooth halos, since any reasonable amount of halo substructure, or a tidal encounter, disturbs the delicate metastable state of the halo, causing friction to appear with its full force. Thus simulations that do not find strong friction from moderately dense halos (*e.g.* Klypin *et al.* 2009) have simply not been run for long enough.

While Debattista & Sellwood (2000) argued for near maximal disks, and their requirement for a low halo density is in agreement with the disk masses derived from fitting bar flow models (§9.4), their constraint on the halo density may be specific to their adopted halo models. Thus additional careful studies of other halo models seem warranted.

### 9.8.3. Halo Density Reduction by Bars

While a full discussion of processes that may lower the dark matter density in the centers of halos is outside this review, a brief mention of the effect of bar friction is appropriate here.

Weinberg & Katz (2002) argued that the transfer of angular momentum from the bar to the halo could reduce the central density of the dark matter halo by a substantial factor. However, the possible density reduction is quite modest (Holley-Bockelmann *et al.* 2005; McMillan & Dehnen 2005; Sellwood 2008a) because the disk has only a finite amount angular momentum to give to the halo. Furthermore, as the disk loses angular momentum, its mass distribution contracts, and the deepening potential well further compresses the halo, which actually overwhelms the slight density reduction (Sellwood 2003; Colín *et al.* 2006).

## 10. SECULAR EVOLUTION WITHIN DISKS

Most of the behavior discussed so far, such as the non-linear evolution of instabilities, causes changes to the host galaxy on a dynamical time scale. The broad topic of secular evolution in galaxies describes changes that occur more gradually, such as the secular formation of pseudo-bulges (see Kormendy 1993; Kormendy & Kennicutt 2004, for excellent reviews, and also Kormendy’s chapter in this volume), the formation of rings (*e.g.* Buta & Combes 1996), or dynamical friction between components (§9.8). The discussion in this section concerns processes that scatter disk stars only.

It has long been realized that old stars in the solar neighborhood have larger peculiar motions relative to the local standard of rest (hereafter LSR) than do young stars (*e.g.* Wielen 1977; Nordström *et al.* 2004; Aumer & Binney 2009, see also Freeman’s chapter). Postulating that older stars were born with larger random speeds, say in a thicker disk, is unattractive because it makes the present epoch of low velocity dispersion special. Some mechanism to scatter stars must therefore be invoked to create the larger random speeds of older stars.

The trends presented in Fig. 16 use the revised ages (Holmberg *et al.* 2007) assigned to the GCS stars (see Soderblom 2010, for a review). Figs. 2, 3 & 4 of Aumer &

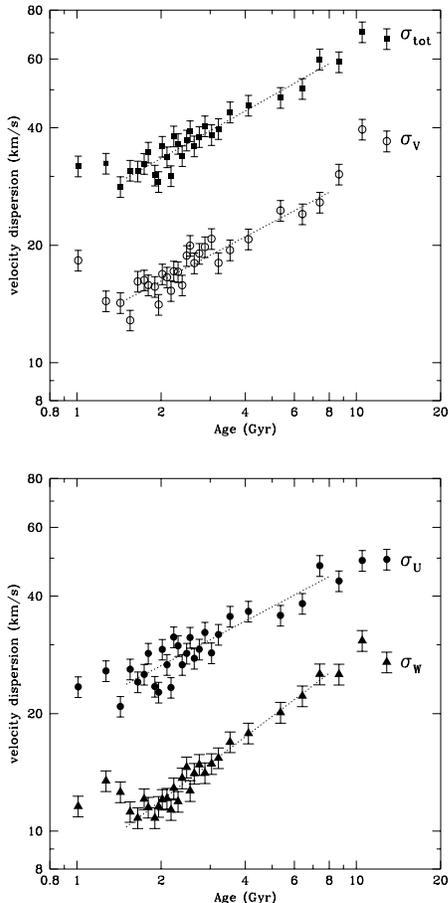


FIG. 16.— The dispersion of stars in all three components, and the total dispersion of the GCS sample of solar neighborhood stars using revised ages (Holmberg *et al.* 2007). The radial, azimuthal and vertical components are  $\sigma_U$ ,  $\sigma_V$ , and  $\sigma_W$  respectively, and  $\sigma_{\text{tot}}^2 = \sigma_U^2 + \sigma_V^2 + \sigma_W^2$ . The fitted straight lines do not include the first three or last two points.

Binney (2009) show the dispersions estimated as a function of color on the main sequence; while blue stars are necessarily young, red stars are expected to have a range of ages. The dispersions of the supposed oldest stars estimated by Holmberg *et al.* (2007) are no greater than those estimated by Aumer & Binney (2009) for their reddest stars, suggesting that the “older” bins in Fig. 16 include stars having a wide range of ages, as argued by Reid *et al.* (2007). It should be noted that the reported dispersions are simple second moments of the perhaps complex velocity distributions (see Fig. 1).

A reliable determination of the variation of dispersion as a function of time could provide another useful constraint on the scattering mechanism (*e.g.* BT08 §8.4). Quillen & Garnett (2000) and Seabroke & Gilmore (2007) argued that the dispersion may saturate for stars above a certain age; with a much older surge to account for the highest velocities. However, Aumer & Binney (2009) found better fits to the data with continuous acceleration, and deduced  $\sigma \sim t^{0.35}$ , with  $t$  being the current age of the stellar generation, in tolerable agreement, in fact, with 0.38 for the logarithmic slope of  $\sigma_{\text{tot}}$  in Fig. 16.

Three principal scattering agents have been discussed: dense gas clumps in the disk, massive black holes in the halo, and recurrent short-lived spirals. Note that the first two are essentially collisional processes that accelerate the relaxation rate (see the Appendix), while the changes caused by spirals can increase random motion without leading to a more relaxed DF. As none in isolation fits the data, a combination of spirals and gas clumps seems to be favored. Minor mergers and the effects of halo substructure are discussed in §11.

An orbiting mass clump induces a collective spiral wake in the surrounding disk that enhances its mass and size by substantial factors (Julian & Toomre 1966), a complication that is ignored in many studies of cloud scattering. Since molecular gas is mostly concentrated in spiral arms (Nieten *et al.* 2006; Gratier *et al.* 2010; Efremov 2010, see also the chapter by Bania), it is probably futile to draw a sharp distinction between spiral arms and the wakes of dense gas clumps, and a correct treatment would be to calculate the effects of the combined star and gas disk. Binney & Lacey (1988) took a step in this direction, but a full calculation may remain unreachable for some time if one tries to include a self-consistent treatment of the formation and dispersal of the gas clumps: molecular gas concentrations probably grow in the converging gas flow into a spiral arm, and are subsequently dispersed by star formation.

Treating spirals and mass clumps in the disk as distinct scattering agents may be justified, therefore, if the wakes of cloud complexes can be lumped with spirals into a single scattering agent that is distinct from the clouds that caused them. At the very least, this simplifying assumption separates the problem into tractable pieces.

### 10.1. Heating by Spirals

Lynden-Bell & Kalnajs (1972) showed that stars are scattered by a slowly changing potential perturbation only near resonances. More precisely, a spiral potential that grows and decays adiabatically, *i.e.* on a time-scale long compared with the orbital and epicyclic periods, will not cause a lasting change to a star’s  $E$  and  $L_z$ . Wave-particle interactions become important near the resonances, where stars experience secular changes through “surfing” on the potential variations at CR or through a periodic forcing close to their epicyclic frequency at the Lindblad resonances. Either case produces a lasting change to a star’s orbit.

The width of a resonance, *i.e.* the range of orbit frequencies of stars that are strongly affected, depends only on the amplitude of the potential when the perturbation is long-lived. But perturbations of shorter lifetimes have a broader range of frequencies and more stars experience lasting changes.

The discussion in §2.4 and Fig. 3 indicates that stars that lose (gain)  $L_z$  near the ILR (OLR) move onto more eccentric orbits, which is the root cause of heating by spirals. Exchanges at CR move stars to new orbits also, but with no change to the energy of non-circular motion, as discussed in §10.2.

Significant heating by spiral waves over a large part of a disk requires them to be transient; a quasi-steady pattern, of the type envisaged by Bertin & Lin (1996) say, will cause

localized heating at an exposed resonance, while stars elsewhere will move through the pattern without otherwise being affected. Barbanis & Woltjer (1967), Carlberg & Sellwood (1985) and Binney & Lacey (1988) calculated the heating caused by transient spirals. Jenkins & Binney (1990), De Simone *et al.* (2004), and Minchev & Quillen (2006) presented numerical studies of the consequences for a disk of test particles subject to some assumed set of spiral wave perturbations.

It is important to realize that the vertical oscillations of stars are little affected by spiral potential variations (§2.6 and Carlberg 1987). In the absence of heavy clumps that can redirect disk velocities (§10.3), the increasing in-plane motions in simulations of initially cool, thin disks may ultimately cause the velocity ellipsoid to become sufficiently anisotropic as to cause it to thicken through mild buckling instabilities (§8.1). This may account for claims (*e.g.* McMillan & Dehnen 2007) that disks thicken due to spiral heating.

### 10.2. Churning by Transient Spirals

Studies of the metallicities and ages of nearby stars (Edwardsson *et al.* 1993; Nordström *et al.* 2004) found that older stars tend to have lower metallicities on average. As the ages of individual stars are disputed (Reid *et al.* 2007; Holmberg *et al.* 2007), the precise form of the relation is unclear. However, there seems to be general agreement that there is a spread of metallicities at each age, which is also supported by other studies (Chen *et al.* 2003; Haywood 2008; Stanghellini & Haywood 2010). The spread seems to be more than twice that expected from simple blurring of the gradient by stellar epicyclic excursions. In the absence of radial mixing, a metallicity spread amongst coeval stars is inconsistent with a simple chemical evolution model in which the metallicity of the disk rises monotonically in each annular bin.

Sellwood & Binney (2002) showed that scattering at CR causes very effective mixing. In a few Gyr, multiple transient spirals caused stars to diffuse in radius. Churning of the stellar disk occurs at corotation of the spirals with no associated heating, and is able to account for the apparent metallicity spread with age. Roškar *et al.* (2008ab) presented more detailed simulations that included infall, star formation and feedback that confirmed this behavior. Schönrich & Binney (2009) developed the first chemical evolution model for the MW disk to include radial churning.

### 10.3. Cloud Scattering

Many years before the discovery of giant molecular gas clouds, Spitzer & Schwarzschild (1953) postulated their existence to account for the secular heating of disk stars. Lacey (1984) extended their calculation to 3D and concluded that cloud scattering should cause the vertical dispersion,  $\sigma_z$ , to be intermediate between the radial,  $\sigma_R$ , and azimuthal,  $\sigma_\phi$ , components.<sup>16</sup> This result seems physically plausible on energy equipartition grounds: scattering by massive clouds redirects the peculiar motions of stars through random angles, and therefore isotropizes the motions as far as the epicycle gyrations allow.

<sup>16</sup> The ratio  $\sigma_R/\sigma_\phi \approx 2\Omega/\kappa$  (BT08, eq. 8.117) is forced by epicyclic motions of disk stars.

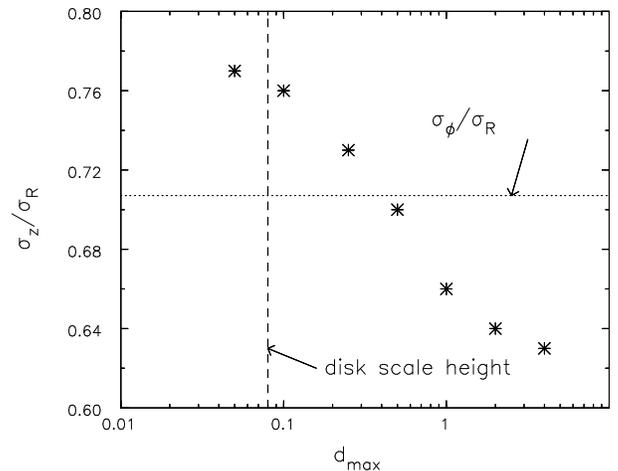


FIG. 17.— The equilibrium axis ratio of the velocity ellipsoid of particles plotted as a function of the limiting range of the perturbation forces from the heavy particles. See Sellwood (2008b) for a description of the calculations.

Despite the fact that redirecting peculiar motions happens much more rapidly than they can be increased by the same scatterers, the data do not reveal Lacey’s expected axis ratio. The second moments of the velocity distribution of solar neighborhood stars in the three orthogonal directions (Fig. 16 and Wielen 1977; Aumer & Binney 2009) satisfy the inequality  $\sigma_z < \sigma_\phi < \sigma_R$ . The ratio of the two in-plane components is in reasonable agreement with expectations from epicyclic motions, but the vertical component is the smallest, and this remains true for all groups when the stars are subdivided according to color or estimates of their ages. Gerssen *et al.* (2000) also observed a flattened ellipsoid in the disk of NGC 2895.

Carlberg (1987) and Jenkins & Binney (1990) therefore developed the plausible argument that spirals drive up the in-plane components more rapidly than scattering is able to redirect those motions into the vertical direction, thereby accounting for the observed axis ratios of the velocity ellipsoid. Sellwood (2000) cited their argument as offering strong support for the transient spiral picture, but it now seems to be incorrect.

Ida *et al.* (1993) claimed that cloud scattering alone would lead to the vertical component being the smallest, with the precise axis ratio depending on the local slope of the rotation curve. Their simulations (Shiidsuke & Ida 1999), and others (*e.g.* Villumsen 1985; Hänninen & Flynn 2002), confirmed their expectation.

Sellwood (2008b) resolved this disagreement using simulations of test particles in the sheared sheet (see §3.6). Scattering by randomly distributed co-orbiting mass clumps confirmed the flattened velocity ellipsoid predicted by Ida *et al.* (1993). Fig. 17 reveals why Ida’s prediction differs from Lacey’s: Lacey, as others (Spitzer & Schwarzschild 1953; Binney & Lacey 1988), assumed that cloud scattering is local, but the  $\ln \Lambda$  term in the formulae in the Appendix implies that distant encounters dominate any scattering process in 3D, at least to distances a few times the disk thickness. Distant scatterers in the flattened geometry of a disk must predominantly affect the in-plane star velocities, and couple much less strongly to

the vertical component. Fig. 17 shows the equilibrium ratio  $\sigma_z/\sigma_R$  when scatterers beyond the finite range  $d_{\max}$  artificially exert no forces. The ratio settles to something close to Lacey’s energy equipartition prediction when none but the closest heavy clumps perturb the stars, but the equilibrium ellipsoid flattens in separate experiments as more distant clouds are included, tending towards Ida’s result with no artificial cut off.

Thus the *shape* of the local velocity ellipsoid, Fig. 16 is apparently consistent with cloud scattering, and does not require concurrent spiral arm scattering, as seemed attractive. However, the data do not imply that spirals are unimportant: cloud scattering seems unable to generate the random speeds of the oldest stars (*e.g.* Lacey 1991; Hänninen & Flynn 2002), and there are hints in Fig. 16 of some evolution of the velocity ellipsoid shape that may demand a compound origin.

#### 10.4. Black Holes in the Halo

The possibility that the dark matter halos of galaxies are made up of massive black holes (BHs) is not yet excluded. Lacey & Ostriker (1985) calculated the consequences for the stars in the disk of the MW, assuming the BHs to have orbits characteristic of a pressure-supported halo and to impart impulses to disk stars as they pass through the disk. The high speeds of their encounters with disk stars would cause the velocity dispersion to rise as  $t^{1/2}$ , while the predicted shape of the velocity ellipsoid is in reasonable agreement with that observed.

Lacey & Ostriker (1985) addressed a number of issues with their model, such as the X-ray accretion luminosity as the BHs pass through the gas disk, the accumulation of BHs in the galactic center through dynamical friction, and the survival of dwarf galaxies. They also acknowledged that it does not predict the correct variation of  $\sigma_{\text{tot}}$  with Galactic radius. BT08 (§7.4.4) added that the idea could be ruled out on the grounds that wide binary star systems would be disrupted too quickly.

If disk scattering is dominated by spirals, as argued below, then scattering by BHs would be needed merely to redirect the peculiar motions into the vertical direction. This reviewer is not aware of any such calculation, but since passing BHs scatter stars in the direction perpendicular to their orbits, it seems unlikely that they could redirect peculiar motion without also increasing in-plane motions. However, if this expectation is too pessimistic and the desired axis ratio could be achieved with lower BH masses, many of the other objections are weakened.

#### 10.5. Discussion

The extraordinary phase-space structure of the solar neighborhood (Dehnen 1998; Nordström *et al.* 2004, and Fig. 1) indicates that there is little in the way of an underlying smooth component and the stellar distribution is broken into several “streams” (Bovy *et al.* 2009). The features are too substantial to have simply arisen from groups of stars that were born with similar kinematics (*e.g.* Eggen 1996), as confirmed in detailed studies (Famaey *et al.* 2007; Bensby *et al.* 2007; Bovy & Hogg 2009).

Thus it is clear that the entire DF has been sculptured by dynamical processes. Were the large part of the spread in velocities caused by multiple scatterings off molecular

clouds in the disk, or off black holes in the halo, the distribution should approximate the simple double Gaussian proposed by K. Schwarzschild (see BT08 §4.4.3). The vertical velocity distribution (Nordström *et al.* 2004), on the other hand, does have a relaxed appearance, as noted by Seabroke & Gilmore (2007).

Various dynamical agents have been proposed to account for kinematic features in the solar neighborhood. Kalnajs (1991) argued that the OLR of the bar in the MW might be close to the solar circle. Features in the subsequently-released HIPPARCOS data have also been attributed to the OLR of the bar (Raboud *et al.* 1998; Dehnen 2000; Fux 2001), while Sellwood (2010) attributed another to a recent ILR.

De Simone *et al.* (2004) were able to produce distributions of stars with a similar degree of substructure in simulations of test particles moving in the adopted potential perturbations representing a succession of short-lived spiral transients; see also Minchev & Quillen (2006). Other models that included both bars and spirals were presented by Quillen (2003), Chakrabarty (2007), and Antoja *et al.* (2009), while Helmi *et al.* (2006) suggested that some of the substructure may also be caused by satellite infall (see also §11).

Since spirals (and bars) are inefficient at exciting vertical motions, another scattering process needs to be invoked to redirect in-plane random motion into vertical motion. The influence of clumps in the disk, which are known to exist, seems consistent with the relaxed appearance of the vertical velocity distribution. Furthermore, the observed axes of the velocity ellipsoid, except possibly for the youngest stars (Fig. 16), are consistent with the prediction of cloud scattering. Thus cloud scattering seems to be just sufficient to redirect velocities, which they are good at, but not to contribute significantly to heating.

### 11. FRAGILITY OF DISKS

Most of the stars in spiral galaxies reside in remarkably thin disks. In a photometric study of edge-on galaxies, Yoachim & Dalcanton (2006) found that the fraction of the baryons in the thin disk component is in the range 70% to 90%, with higher fractions being characteristic of more massive galaxies. The so-called superthin disks are even more extreme; the disk of the low-luminosity galaxy UGC 7321, for example, has a radial scale-length some 14 times that of its characteristic thickness (Matthews 2000).

Hierarchical galaxy formation scenarios (see chapter by Steinmetz) predict that galaxy formation is far from monolithic, with occasional major mergers with other halos, and frequent minor mergers. Stellar disks that have formed in the centers of the halos are torn apart in major mergers (Barnes & Hernquist 1992), but the consequences of minor mergers are harder to determine from simulations (*e.g.* Walker *et al.* 1996). Tóth & Ostriker (1992) argued that the existence of thin disks in galaxies can be used to constrain the rate of minor mergers. Their numerical estimates have been criticized on various grounds (*e.g.* Huang & Carlberg 1997; Sellwood *et al.* 1998; Velázquez & White 1999), but it is clear that a tight constraint remains.

Wyse (this volume) stresses that the thick disk of the MW contains only very old stars, and that the ages of thin disk stars stretch back 10 Gyr. This fact would seem to

imply that the last galactic merger to have stirred the MW disk occurred some 10 Gyr ago, and that no comparable disturbance could have occurred since. The MW may not be unique in this respect, as Mould (2005) found that thick disks in four nearby galaxies also appear to be old.

Stewart *et al.* (2008) estimated the rate of mergers in the current  $\Lambda$ CDM cosmology. They concluded that 95% of parent halos of some  $10^{12}h^{-1} M_{\odot}$  will have merged with another halo of at least 20% of its mass in the last 10 Gyr, and consequently the disk hosted by the parent must somehow survive in most cases.

Possibly only a small fraction of infalling satellites pose a real threat to a disk. All satellites will be tidally stripped as they fall into the main halo, and some may even dissolve completely before they can affect the disk – the Sagittarius dwarf galaxy appears to be a good example (Law *et al.* 2005). A massive satellite loses orbital energy through dynamical friction (§9.8) causing it to settle deeper into the main halo. Its survival depends on its density (see BT08 §8.3); it will be stripped of its loosely bound envelope until its mean density,  $\bar{\rho}$ , is about one third the mean density of the halo interior to its orbital radius (BT08 eq. 8.92), and disrupted entirely as it reaches a radius in the main halo where even its central density falls below the mean interior halo density. This process has been studied in more detail by Boylan-Kolchin & Ma (2007) and by Choi *et al.* (2009). Thus the vulnerability of disks depends on the inner densities, in both dark matter and baryons, of the accreted sub-halos.

If even a single, moderately massive core survives to the inner halo, it could cause an unacceptable increase in the disk thickness. Vertical heating of the disk occurs when a passing or penetrating satellite is able to increase the vertical motions of the disk stars. High speed passages therefore deposit little energy into disk motions, but if the satellite’s orbit remains close to the disk mid-plane, then its vertical frequency will couple strongly to that of some of the disk stars and heating will be rapid. Indeed, Read *et al.* (2008) argued that the accretion event(s) that created the thick disk of the MW most probably resulted from the infall of a subhalo whose orbit plane was inclined at  $10^{\circ}$  –  $20^{\circ}$  to that of the disk. The energy deposited could take the form of exciting bending waves in the disk that can propagate radially until they are damped at vertical resonances (Sellwood *et al.* 1998).

Kazantzidis *et al.* (2009) simulated several minor mergers with sub-halos in detailed models with plausible parameters. They found that the disk is substantially thickened and heated by the mergers, although they noted that their simulations lacked a gaseous component. Simulations that include a gas component cannot resolve the small-scale physical processes of gas fragmentation, star formation, feedback, *etc.*, and therefore the behavior of the dissipative gas component necessarily includes somewhat *ad hoc* prescriptions for these aspects of sub-grid physics. Hopkins *et al.* (2009) stressed that gas in mergers settles quickly to begin to form a new disk; however, it is the fate of the stars that had been formed prior to the merger that is the principal concern. Kazantzidis *et al.* (2009) argued it is possible that a dissipative component in the disk could absorb some of the orbital energy of the satellite which would reduce the heating of the stellar disk,

and such an effect appears to have occurred in the simulations by Moster *et al.* (2010). A clear conclusion has yet to emerge from this on-going research effort, but exclusively old thick disks together with the prevalence of thin disks poses a substantial, though perhaps surmountable, challenge to the  $\Lambda$ CDM model.

It should be noted that ideas to reduce the density of the inner main halo through frictional energy loss to halo (*e.g.* El-Zant *et al.* 2001; Mashchenko *et al.* 2006; Romano-Díaz *et al.* 2008a) require the kinds of dense massive fragments that are themselves a danger to the disk. This will be of less importance in the early stages of galaxy assembly, but disk survival adds the requirement that any such process be completed quickly.

## 12. CONCLUSIONS

This review has focused on understanding the mechanisms that underlie the various instabilities and processes that affect the structure of galaxy disks. The discussion has referred extensively to simulations that are sufficiently simple that they clearly illustrate each particular aspect of the behavior. A complementary, and powerful, approach is to add additional physical processes to simulations, with the aim of improving realism to address otherwise inaccessible questions. However, the increased complexity of the behavior that comes with increasing realism, makes a deep understanding of the results harder to achieve. Hopefully, the behavior described here will form a solid basis on which to build as the challenges presented by more realistic systems are addressed.

A contrast with accretion disk theory seems appropriate here. Shakura & Sunyaev (1973) proposed a scaling relation for turbulent viscosity that led to rapid progress in modeling accretion disks some 25 years before the likely origin of the viscosity was identified (Balbus & Hawley 1998). A theory for the structural evolution of galaxies seems much farther away. While most of the important physical processes may have been identified, exactly how they drive evolution is still not fully understood, and even the evolutionary path remains vague. In the absence of this understanding, simplifying scaling laws cannot be identified with confidence and it seems best to work at improved understanding of the mechanisms at play.

Galaxy dynamics has made immense strides in the 45 years since the chapter by Oort (1965) in the corresponding volume of the preceding series. Understanding of local wave mechanics (§3) and the mechanisms for the principal global gravitational instabilities (§§4 & 6) is well advanced. Bending waves and global buckling modes (§8) are mostly understood, but not at quite the same level, while lop-sided modes (§5) perhaps still require some more concerted effort. Understanding of bar structure (§9) and the role of bars in galaxy evolution has developed beyond recognition, and disk heating (§10) seems more solidly understood, with our confidence being strongly boosted by the extremely valuable data from Nordström *et al.* (2004).

But answers to a number of major questions of galaxy dynamics are still incomplete. Although the bar-forming instability is now understood (§4.1), it does not provide a clear picture of why only a little over half of all bright disk galaxies are barred (§9.2). After bars, spirals arms are the second most prominent feature of disk galaxies

and are probably responsible for the most dynamical evolution (§10), yet a deep understanding of their origin (§7) remains elusive. The good progress made in recent years to understand the warps of galaxy disks (§8) still has not supplied a satisfactory account of their incidence.

Many of these outstanding issues may be bound up with how galaxies form, our understanding of which is currently making particularly rapid progress.

#### ACKNOWLEDGMENTS

The author is indebted to James Binney, Victor Debattista, Agris Kalnajs, Juntao Shen, Alar Toomre, and Scott Tremaine for numerous valuable comments on a draft of this review.

#### REFERENCES

- Agertz, O., Teyssier, R. & Moore, B. 2010, arXiv:1004.0005  
 Aguerri, J. A. L., Méndez-Abreu, J. & Corsini, E. M. 2009, *A&A*, **495**, 491  
 Antoja, T., Valenzuela, O., Pichardo, B., Moreno, E., Figueras, F. & Fernández, D. 2009, *ApJL*, **700**, L78  
 Aoki, S., Noguchi, M. & Iye, M. 1979, *PASJ*, **31**, 737  
 Araki, S. 1985, PhD thesis, MIT.  
 Araki, S. 1987, *AJ*, **94**, 99  
 Athanassoula, E. 1992, *MNRAS*, **259**, 345  
 Athanassoula, E. 2002, *ApJL*, **569**, L83  
 Athanassoula, E. 2003, *MNRAS*, **341**, 1179  
 Athanassoula, E., Bosma, A. & Papaioannou, S. 1987, *A&A*, **179**, 23  
 Athanassoula, E., Lambert, J. C. & Dehnen, W. 2005, *MNRAS*, **363**, 496  
 Athanassoula, E., Romero-Gómez, M. & Masdemont, J. J. 2009, *MNRAS*, **394**, 67  
 Athanassoula, E. & Sellwood, J. A. 1986, *MNRAS*, **221**, 213  
 Aumer, M. & Binney, J. J. 2009, *MNRAS*, **397**, 1286  
 Bailin, J. 2003, *ApJL*, **583**, L79  
 Balbus, S. A. & Hawley, J. F. 1998, *Rev. Mod. Phys.*, **70**, 1  
 Baldwin, J. E., Lynden-Bell, D., Sancisi R. 1980, *MNRAS*, **193**, 313  
 Barazza, F. D., Jogee, S. & Marinova, I. 2008, *ApJ*, **675**, 194  
 Barazza, F. D. *et al.* 2009, *A&A*, **497**, 713  
 Barbanis, B. & Woltjer, L. 1967, *ApJ*, **150**, 461  
 Barnes, J. E. & Hernquist, L. 1992, *ARAA*, **30**, 705  
 Benedict, G. F., Howell, A., Jorgensen, I., Kenney, J. & Smith, B. J. 2002, *AJ*, **123**, 1411  
 Bensby, T., Oey, M. S., Feltzing, S. & Gustafsson, B. 2007, *ApJL*, **655**, L89  
 Berentzen, I., Athanassoula, E., Heller, C. H. & Fricke, K. J. 2004, *MNRAS*, **347**, 220  
 Berentzen, I., Shlosman, I., Martínez-Valpuesta, I. & Heller, C. H. 2007, *ApJ*, **666**, 189  
 Bertin, G. & Lin, C. C. 1996, *Spiral Structure in Galaxies* (Cambridge, MA: The MIT Press)  
 Binney, J., Jiang, I. & Dutta, S. 1998, *MNRAS*, **297**, 1237  
 Binney, J. J. & Lacey, C. G. 1988, *MNRAS*, **230**, 597  
 Binney, J. & Tremaine, S. 2008, *Galactic Dynamics* (2nd ed.; Princeton: Princeton University Press) (BT08)  
 Block, D. L., Freeman, K. C., Jarrett, T. H., Puerari, I., Worthey, G., Combes, F. & Groess, R. 2004, *A&A*, **425**, L37  
 Bosma, A. 1991, in *Warped Disks and Inclined Rings Around Galaxies*, ed. S. Casertano, P. D. Sackett & F. H. Briggs (Cambridge: Cambridge University Press), 181  
 Bosma, A. 1996, in *IAU Colloq. 157, Barred Galaxies*, ed. R. Buta, D. A. Crocker & B. G. Elmegreen (San Francisco: ASP Conf series **91**), 132  
 Bottema, R. 1996, *A&A*, **306**, 345  
 Bornaud, F., Combes, F. & Semelin, B. 2005, *MNRAS*, **364**, L18  
 Bovy, J. & Hogg, D. W. 2009, arXiv:0912.3262  
 Bovy, J., Hogg, D. W. & Roweis, S. T. 2009, *ApJ*, **700**, 1794  
 Boylan-Kolchin, M. & Ma, C.-P. 2007, *MNRAS*, **374**, 1227  
 Briggs, F. H. 1990, *ApJ*, **352**, 15  
 Bureau, M. & Athanassoula, E., 2005, *ApJ*, **626**, 159  
 Buta, R. & Combes, F. 1996, *Fund. Cosmic Phys.*, **17**, 95  
 Buta, R. J., Knapen, J. H., Elmegreen, B. G., Salo, H., Laurikainen, E., Elmegreen, D. M., Puerari, I. & Block, D. L. 2009, *AJ*, **137**, 4487  
 Camm, G. L. 1950, *MNRAS*, **110**, 305  
 Carlberg, R. G. 1987, *ApJ*, **322**, 59  
 Carlberg, R. G. & Freedman, W. L. 1985, *ApJ*, **298**, 486  
 Carlberg, R. G. & Sellwood, J. A. 1985, *ApJ*, **292**, 79  
 Carollo, C. M., Stiavelli, M. & Mack, J. 1998, *AJ*, **116**, 68  
 Chakrabarty, D. 2007, *A&A*, **467**, 145  
 Chandrasekhar, S. 1943, *ApJ*, **97**, 255  
 Chemin, L. & Hernandez, O. 2009, *A&A*, **499**, L25  
 Chen, L., Hou, J. L. & Wang, J. J. 2003, *AJ*, **125**, 1397  
 Chirikov, B. V. 1979, *Phys. Rep.*, **52**, 265-379  
 Choi, J.-H., Weinberg, M. D. & Katz, N. 2009, *MNRAS*, **400**, 1247  
 Colin, P., Valenzuela, O. & Klypin, A. 2006, *ApJ*, **644**, 687  
 Combes, F. & Sanders, R. H. 1981, *A&A*, **96**, 164  
 Contopoulos, G. 1980, *A&A*, **81**, 198  
 Corbelli, E. & Walterbos, R. A. M. 2007, *ApJ*, **669**, 315  
 Corsini, E. M. 2008, in *Formation and Evolution of Galaxy Bulges*, *IAU Symp.* **245**, p. 125  
 Corsini, E. M., Debattista, V. P. & Aguerri, J. A. L. 2003, *ApJL*, **599**, L29  
 Courteau, S., Andersen, D. R., Bershady, M. A., MacArthur, L. A. & Rix, H.-W. 2003, *ApJ*, **594**, 208  
 Cox, A. L., Sparke, L. S., van Moorsel, G. & Shaw, M. 1996, *AJ*, **111**, 1505  
 Curir, A., Mazzei, P. & Murante, G. 2006, *A&A*, **447**, 453  
 Cuzzi, J. N., *et al.* 2010, *Science*, **327**, 1470  
 Davoust, E. & Contini, T. 2004, *A&A*, **416**, 515  
 Debattista, V. P. & Sellwood, J. A. 1998, *ApJL*, **493**, L5  
 Debattista, V. P. & Sellwood, J. A. 1999, *ApJL*, **513**, L107  
 Debattista, V. P. & Sellwood, J. A. 2000, *ApJ*, **543**, 704  
 Debattista, V. P. & Shen, J. A. 2007, *ApJL*, **654**, L127  
 Dehnen, W. 1998, *AJ*, **115**, 2384  
 Dehnen, W. 2000, *AJ*, **119**, 800  
 Dekel, A. & Shlosman, I. 1983, in *IAU Symposium 100, Internal Kinematics and Dynamics of Galaxies*, ed. E. Athanassoula (Dordrecht: Reidel) p 187  
 De Simone, R. S., Wu, X. & Tremaine, S. 2004, *MNRAS*, **350**, 627  
 Dobbs, C. L., Theis, C., Pringle, J. E. & Bate, M. R. 2010, *MNRAS*, (in press)  
 Donner, K. J. & Thomasson, M. 1994, *A&A*, **290**, 475  
 Dubinski, J., Berentzen, I. & Shlosman, I. 2009, *ApJ*, **697**, 293  
 Dubinski, J. & Chakrabarty, D. 2009, *ApJ*, **703**, 2068  
 Dubinski, J., Gauthier, J.-R., Widrow, L. & Nickerson, S. 2008, in *Formation and Evolution of Galaxy Disks*, ed. J. G. Funes SJ & E. M. Corsini (San Francisco: ASP **396**), p. 321  
 Dubinski, J. & Kuijken, K. 1995, *ApJ*, **442**, 492  
 Dury, V., de Rijcke, S., Debattista, V. P. & Dejonghe, H. 2008, *MNRAS*, **387**, 2  
 Earn, D. J. D. & Lynden-Bell, D. 1996, *MNRAS*, **278**, 395  
 Edvardsson, B., Andersen, B., Gustafsson, B., Lambert, D. L., Nissen, P. E. & Tomkin, J. 1993, *A&A*, **275**, 101  
 Efremov, Yu. N. 2010, *MNRAS*, to appear (arXiv:1002.4555)  
 Efsthathiou, G., Lake, G. & Negroponte, J. 1982, *MNRAS*, **199**, 1069  
 Eggen, O. J. 1996, *AJ*, **112**, 1595  
 Elmegreen, B. 1996, in *IAU Colloq. 157, Barred Galaxies*, ed. R. Buta, D. A. Crocker & B. G. Elmegreen (San Francisco: ASP Conf series **91**), 197  
 Elmegreen, B. G. & Thomasson, M. 1993, *A&A*, **272**, 37  
 Elmegreen, B. G., Elmegreen, D. M., Knapen, J. H., Buta, R. J., Block, D. L. & Puerari, I. 2007, *ApJL*, **670**, L97  
 Elmegreen, D. M., Elmegreen, B. G. & Bellin, A. D. 1990, *ApJ*, **364**, 415  
 El-Zant, A., Shlosman, I. & Hoffman, Y. 2001, *ApJ*, **560**, 636  
 Englmaier, P. & Gerhard, O. 1997, *MNRAS*, **287**, 57  
 Englmaier, P. & Shlosman, I. 2004, *ApJL*, **617**, L115  
 Erwin, P. 2005, *MNRAS*, **364**, 283  
 Erwin, P. & Sparke, L. S. 2002, *AJ*, **124**, 65  
 Eskridge, P. B., *et al.* 2000, *AJ*, **119**, 536  
 Evans, N. W. & Read, J. C. A. 1998, *MNRAS*, **300**, 106  
 Famaey, B., Pont, F., Luri, X., Udry, S., Mayor, M. & Jorriensen, A. 2007, *A&A*, **461**, 957  
 Fathi, K., Beckman, J. E., Piñol-Ferrer, N., Hernandez, O., Martínez-Valpuesta, I. & Carignan, C. 2009, *ApJ*, **704**, 1657  
 Fridman, A. M. & Polyachenko, V. L. 1984. *Physics of Gravitating Systems* (New York: Springer-Verlag)  
 Friedli, D. & Martinet, L. 1993, *A&A*, **277**, 27  
 Fuchs, B., Dettbarn, C. & Tsuchiya, T. 2005, *A&A*, **444**, 1  
 Fux, R. 1999, *A&A*, **345**, 787  
 Fux, R. 2001, *A&A*, **373**, 511  
 García-Ruiz, I., Kuijken, K. & Dubinski, J. 2002a, *MNRAS*, **337**, 459  
 García-Ruiz, I., Sancisi, R. & Kuijken, K. 2002b, *A&A*, **394**, 769  
 Gerssen, J., Kuijken, K. & Merrifield, M. R. 2000, *MNRAS*, **317**, 545  
 Goldreich, P. & Lynden-Bell, D. 1965a, *MNRAS*, **130**, 97  
 Goldreich, P. & Lynden-Bell, D. 1965b, *MNRAS*, **130**, 125  
 Goldreich, P. & Tremaine, S. 1978, *ApJ*, **222**, 850  
 Gratier, P. *et al.* 2010, *A&A*, to appear (arXiv:1003.3222)

- Grosbøl, P., Patsis, P. A. & Pompei, E. 2004, *A&A*, **423**, 849
- Haan, S., Schinnerer, E., Emsellem, E., Garca-Burillo, S., Combes, F., Mundell, C. G. & Rix, H.-W. 2009, *ApJ*, **692**, 1623
- Hänninen, J. & Flynn, C. 2002, *MNRAS*, **337**, 731
- Haywood, M. 2008, *MNRAS*, **388**, 1175
- Heller, C. H., Shlosman, I. & Athanassoula, E. 2007, *ApJ*, **657**, L65
- Heller, C., Shlosman, I. & Englmaier, P. 2001, *ApJ*, **553**, 661
- Helmi, A., Navarro, J. F., Nordström, B., Holmberg, J., Abadi, M. G. & Steinmetz, M. 2006, *MNRAS*, **365**, 1309
- Hénon, M. 1973, in *Dynamical Structure and Evolution of Stellar Systems*, ed. L. Martinet & M. Mayor (Sauverny: Geneva Observatory) p. 182
- Hernquist, L. & Weinberg, M. D. 1992, *ApJ*, **400**, 80
- Hill, G. W., 1878 *Am. J. Math.*, **1**, 5
- Hockney, R. W. & Brownrigg, D. R. K. 1974, *MNRAS*, **167**, 351
- Hohl, F. 1971, *ApJ*, **168**, 343
- Holley-Bockelmann, K., Weinberg, M. & Katz, N. 2005, *MNRAS*, **363**, 991
- Holmberg, J. & Flynn, C. 2004, *MNRAS*, **352**, 440
- Holmberg, J., Nordström, B. & Andersen, J. 2007, *A&A*, **475**, 519
- Holmberg, J., Nordström, B. & Andersen, J. 2009, *A&A*, **501**, 941
- Hopkins, P. F., Cox, T. J., Younger, J. D. & Hernquist, L. 2009, *ApJ*, **691**, 1168
- Huang, S. & Carlberg, R. G. 1997, *ApJ*, **480**, 503
- Hunter, C. & Toomre, A. 1969, *ApJ*, **155**, 747
- Ida, S., Kokuba, E. & Makino, J. 1993, *MNRAS*, **263**, 875
- Ideta, M. 2002, *ApJ*, **568**, 190
- Jalali, M. A. 2007, *ApJ*, **669**, 218
- James, R. A. & Sellwood, J. A. 1978, *MNRAS*, **182**, 331
- Jeans, J. H. 1923, *MNRAS*, **84**, 60
- Jeans, J. H. 1929, *Astronomy and Cosmogony* (Cambridge: Cambridge University Press)
- Jenkins, A. & Binney, J. J. 1990, *MNRAS*, **245**, 305
- Jiang, I. & Binney, J. 1999, *MNRAS*, **303**, L7
- Jog, C. J. & Combes, F. 2009, *Phys. Rep.*, **471**, 75
- Jog, C. J. & Solomon, P. M. 1992, *ApJ*, **387**, 152
- Jogee, S. *et al.* 2004, *ApJL*, **615**, L105
- Julian, W. H. & Toomre, A. 1966, *ApJ*, **146**, 810
- Kahn, F. D. & Woltjer, L. 1959, *ApJ*, **130**, 705
- Kalnajs, A. J. 1965, PhD thesis, Harvard University
- Kalnajs, A. J. 1972, *ApJ*, **175**, 63
- Kalnajs, A. J. 1976, *ApJ*, **205**, 751
- Kalnajs, A. J. 1978, in *IAU Symposium 77 Structure and Properties of Nearby Galaxies* eds. E. M. Berkhuisjen & R. Wielebinski (Dordrecht:Reidel) p. 113
- Kalnajs, A. J. 1991, in *Dynamics of Disc Galaxies*, ed. B. Sundelius (Gothenburg: Göteborgs University) p. 323
- Kazantzidis, S., Zentner, A. R., Kravtsov, A. V., Bullock, J. S. & Debattista, V. P. 2009, *ApJ*, **700**, 1896
- Khoperskov, A. V., Just, A., Korchagin, V. I. & Jalali, M. A. 2007, *A&A*, **473**, 31
- Klypin, A., Valenzuela, O., Colín, P. & Quinn, T. 2009, *MNRAS*, **398**, 1027
- Korchagin, V., Orlova, N., Kikuchi, N., Miyama, S. M. & Moiseev, A. V. 2005, *arXiv:astro-ph/0509708*
- Kornreich, D. A., Lovelace, R. V. E. & Haynes, M. P. 2002, *ApJ*, **580**, 705
- Kormendy, J. 1993, in *IAU Symposium 153, Galactic Bulges*, eds. H. Dejonghe & H. Habing (Dordrecht: Kluwer) p. 209
- Kormendy, J. & Kennicutt, R. C. 2004, *ARAA*, **42**, 603
- Kormendy, J. & Norman, C. A. 1979, *ApJ*, **233**, 539
- Kranz, T., Slyz, A. D. & Rix, H.-W. 2003, *ApJ*, **586**, 143
- Kuijken, K. 1991, *ApJ*, **376**, 467
- Kulsrud, R. M., Mark, J. W.-K. & Caruso, A. 1971, *Ap. Sp. Sci.*, **14**, 52
- Lacey, C. G. 1984, *MNRAS*, **208**, 687
- Lacey, C. G. 1991, in *Dynamics of Disc Galaxies*, ed. B. Sundelius (Gothenburg: Göteborgs University) p. 257
- Lacey, C. G. & Ostriker, J. P. 1985, *ApJ*, **299**, 633
- Law, D. R., Johnston, K. V. & Majewski, S. R. 2005, *ApJ*, **619**, 807
- Levine, E. S., Blitz, L. & Heiles, C. 2006, *ApJ*, **643**, 881
- Li, C., Gadotti, D. A., Mao, S. & Kauffmann, G. 2009, *MNRAS*, **397**, 726
- Lin, C. C. & Shu, F. H. 1966, *Proc. Nat. Acad. Sci.*, **55**, 229
- Lin, D. N. C. & Tremaine, S. 1983, *ApJ*, **264**, 364
- Lindblad, P. A. B., Lindblad, P. O. & Athanassoula, E. 1996, *A&A*, **313**, 65
- Little, B. & Carlberg, R. G. 1991, *MNRAS*, **250**, 161
- Louis, P. D. & Gerhard, O. E. 1988, *MNRAS*, **233**, 337
- Lovelace, R. V. E. 1998, *A&A*, **338**, 819
- Lovelace, R. V. E. & Hohlfeld, R. G. 1978, *ApJ*, **221**, 51
- Lovelace, R. V. E., Zhang, L., Kornreich, D. A. & Haynes, M. P. 1999, *ApJ*, **524**, 634
- Lowe, S. A., Roberts, W. W., Yang, J., Bertin, G. & Lin, C. C. 1994, *ApJ*, **427**, 184
- Lynden-Bell, D. 1965, *MNRAS*, **129**, 299
- Lynden-Bell, D. 1979, *MNRAS*, **187**, 101
- Lynden-Bell, D. & Kalnajs, A. J. 1972, *MNRAS*, **157**, 1
- Maciejewski, W. 2006, *MNRAS*, **371**, 451
- Maciejewski, W. & Sparke, L. S. 2000, *MNRAS*, **313**, 745
- Maciejewski, W., Teuben, P. J., Sparke, L. S. & Stone, J. M. 2002, *MNRAS*, **329**, 502
- Maoz, D., Barth, A. J., Ho, L. C., Sternberg, A. & Filippenko, A. V. 2001, *AJ*, **121**, 3048
- Marinova, I. & Jogee, S. 2007, *ApJ*, **659**, 1176
- Mark, J. W.-K. 1974, *ApJ*, **193**, 539
- Mark, J. W.-K. 1976, *ApJ*, **203**, 81
- Mark, J. W.-K. 1977, *ApJ*, **212**, 645
- Martinez-Valpuesta, I. & Shlosman, I. 2004, *ApJL*, **613**, L29
- Mashchenko, S., Couchman, H. M. P. & Wadsley, J. 2006, *Nature*, **442**, 539
- Masset, F. & Tagger, M. 1997, *A&A*, **322**, 442
- Masters, K. L., Nichol, R. C., Hoyle, B., Lintott, C., Bamford, S., Edmondson, E. M., Fortson, L., Keel, W. C., Schawinski, K., Smith, A. & Thomas, D. 2010, *arXiv:1003.0449*
- Mathur, S. D. 1990, *MNRAS*, **243**, 529
- Matthews, L. D. 2000, *AJ*, **120**, 1764
- McMillan, P. J. & Dehnen, W. 2005, *MNRAS*, **363**, 1205
- McMillan, P. J. & Dehnen, W. 2007, *MNRAS*, **378**, 541
- Meidt, S. E., Rand, R. J., Merrifield, M. R., Debattista, V. P. & Shen, J. 2008, *ApJ*, **676**, 899
- Meidt, S. E., Rand, R. J. & Merrifield, M. R. 2009, *ApJ*, **702**, 277
- Méndez-Abreu, J., Sánchez-Janssen, R. & Aguerri, J. A. L. 2010, *ApJL*, **711**, L61
- Merrifield, M. R. & Kuijken, K. 1999, *A&A*, **345**, L47
- Merritt, D. & Sellwood, J. A. 1994, *ApJ*, **425**, 551
- Merritt, D. & Stiavelli, M. 1990, *ApJ*, **358**, 399-417
- Mestel, L. 1963, *MNRAS*, **126**, 553
- Miller, R. H., Prendergast, K. H. & Quirk, W. J. 1970, *ApJ*, **161**, 903
- Minchev, I. & Quillen, A. C. 2006, *MNRAS*, **368**, 623
- Moster, B. P., Macciò, A. V., Somerville, R. S., Johansson, P. H. & Naab, T. 2010, *MNRAS*, **403**, 1009
- Mould, J. 2005, *AJ*, **129**, 698
- Mulder, W. A. 1983, *A&A*, **117**, 9
- Nelson, R. W. & Tremaine, S. 1995, *MNRAS*, **275**, 897
- Nieten, Ch., Neining, N., Guélin, M., Ungerechts, H., Lucas, R., Berkhuisjen, E. M., Beck, R. & Wielebinski, R. 2006, *A&A*, **453**, 459
- Noguchi, M. 1987, *MNRAS*, **228**, 635
- Nordström, B., Mayor, M., Andersen, J., Holmberg, J., Pont, F., Jørgensen, B. R., Olsen, E. H., Udry, S. & Mowlavi, N. 2004, *A&A*, **418**, 989
- Norman, C. A., Sellwood, J. A. & Hasan, H. 1996, *ApJ*, **462**, 114
- O'Neill, J. K. & Dubinski, J. 2003, *MNRAS*, **346**, 251
- Oort, J. H. 1965, in *Stars and Stellar Systems, 5 Galactic Structure*, ed. A. Blaauw & M. Schmidt (Chicago: University of Chicago Press), p. 455
- Oort, J. H., Kerr, F. J. & Westerhout, G. 1958, *MNRAS*, **118**, 379
- Ostriker, E. C. & Binney, J. J. 1989, *MNRAS*, **237**, 785
- Ostriker, J. P. & Peebles, P. J. E. 1973, *ApJ*, **186**, 467
- Papaloizou, J. C. B. & Lin, D. N. C. 1989, *ApJ*, **344**, 645
- Patsis, P. A., Contopoulos, G. & Grosbøl, P. 1991, *A&A*, **243**, 373
- Patsis, P. A., Skokos, Ch. & Athanassoula, E. 2002, *MNRAS*, **337**, 578
- Pérez, I. 2008, *A&A*, **478**, 717
- Pérez, I., Fux, R. & Freeman, K. 2004, *A&A*, **424**, 799
- Pfenniger, D. & Friedli, D. 1991, *A&A*, **252**, 75
- Pfenniger, D. & Norman, C. 1990, *ApJ*, **363**, 391
- Pichon, C. & Cannon, R. C. 1997, *MNRAS*, **291**, 616
- Polyachenko, E. V. 2004, *MNRAS*, **348**, 345
- Polyachenko, E. V. 2005, *MNRAS*, **357**, 559
- Polyachenko, V. L. 1977, *Sov. Ast. Lett.*, **3**, 51
- Prendergast, K. H. 1962, in *Interstellar Matter in Galaxies*, ed. L. Woltjer (New York: Benjamin), p. 217
- Quillen, A. C. 2003, *AJ*, **125**, 785
- Quillen, A. C. & Garnett, D. R. 2000, *arXiv:astro-ph/0004210*
- Quinn, T. & Binney, J. 1992, *MNRAS*, **255**, 729
- Raboud, D., Grenon, M., Martinet, L., Fux, R. & Udry, S. 1998, *A&A*, **335**, L61
- Rafikov, R. R. 2001, *MNRAS*, **323**, 445
- Raha, N., Sellwood, J. A., James, R. A. & Kahn, F. D. 1991, *Nature*, **352**, 411
- Rautiainen, P. & Salo, H. 1999, *A&A*, **348**, 737
- Rautiainen, P., Salo, H. & Laurikainen, E. 2002, *MNRAS*, **337**, 1233
- Rautiainen, P., Salo, H. & Laurikainen, E. 2008, *MNRAS*, **388**, 1803
- Read, J. I., Lake, G., Agertz, O. & Debattista, V. P. 2008, *MNRAS*, **389**, 1041
- Reese, A., Williams, T. B., Sellwood, J. A., Barnes, E. I. & Powell, B. A. 2007, *AJ*, **133**, 2846
- Regan, M. W., Sheth, K., Teuben, P. J. & Vogel, S. N. 2002, *ApJ*, **574**, 126

- Reid, I. N., Turner, E. L., Turnbull, M. C., Mountain, M. & Valenti, J. A. 2007, *ApJ*, **665**, 767
- Reshetnikov, V., Battaner, E., Combes, F. & Jiménez-Vicente, J. 2002, *A&A*, **382**, 513
- Reylé, C., Marshall, D. J., Robin, A. C. & Schultheis, M. 2009, *A&A*, **495**, 819
- Roberts, W. W., Huntley, J. M. & van Albada, G. D. 1979, *ApJ*, **233**, 67
- Romano-Díaz, E., Shlosman, I., Hoffman, Y. & Heller, C. 2008a, *ApJL*, **685**, L105
- Romano-Díaz, E., Shlosman, I., Heller, C. & Hoffman, Y. 2008b, *ApJL*, **687**, L13
- Romeo, A. B. 1992, *MNRAS*, **256**, 307
- Roškar, R., Debattista, V. P., Quinn, T. R., Stinson, G. S. & Wadsley, J. 2008, *ApJL*, **684**, L79
- Roškar, R., Debattista, V. P., Stinson, G. S., Quinn, T. R., Kaufmann, T. & Wadsley, J. 2008, *ApJL*, **675**, L65
- Rubin, V. C., Graham, J. A. & Kenney, J. D. P. 1992, *ApJL*, **394**, L9
- Rybicki, G. B. 1972, in *IAU Colloq.* **10**, *Gravitational N-body Problem*, ed. M. Lecar (Dordrecht: Reidel), 22
- Saha, K., Combes, J. & Jog, C. 2007, *MNRAS*, **382**, 419
- Saha, K., de Jong, R. & Holwerda, B. 2009, *MNRAS*, **396**, 409
- Salo, H. & Laurikainen, E. 1993, *ApJ*, **410**, 586
- Sancisi, R., Fraternali, F., Oosterloo, T. & van der Hulst, T. 2008, *A&A Rev.*, **15**, 189
- Sandage, A. & Humphreys, R. M. 1980, *ApJL*, **236**, L1
- Sanders, R. H. & Huntley, J. M. 1976, *ApJ*, **209**, 53
- Sawamura, M. 1988, *PASJ*, **40**, 279
- Schönrich, R. & Binney, J. 2009, *MNRAS*, **396**, 203
- Schwarz, M. P. 1981, *ApJ*, **247**, 77
- Seabroke, G. M. & Gilmore, G. 2007, *MNRAS*, **380**, 1348
- Sellwood, J. A. 1980, *A&A*, **89**, 296
- Sellwood, J. A. 1981, *A&A*, **99**, 362
- Sellwood, J. A. 1985, *MNRAS*, **217**, 127
- Sellwood, J. A. 1989a, *MNRAS*, **238**, 115
- Sellwood, J. A. 1989b, in *Dynamics of Astrophysical Discs*, ed. J. A. Sellwood (Cambridge: Cambridge University Press) p. 155
- Sellwood, J. A. 1994, in *Galactic and Solar System Optical Astrometry* ed L Morrison (Cambridge: Cambridge University Press) p. 156
- Sellwood, J. A. 1996a, in *IAU Symp.* **169**, *Unsolved Problems of the Milky Way*, ed. L. Blitz & P. Teuben (Dordrecht: Kluwer) p. 31
- Sellwood, J. A. 1996b, *ApJ*, **473**, 733
- Sellwood, J. A. 2000, in *Astrophysical Dynamics – in Commemoration of F. D. Kahn*, ed. D. Berry, D. Breitschwerdt, A. da Costa & J. E. Dyson, *Ap. Sp. Sci.*, **272**, p. 31 (astro-ph/9909093)
- Sellwood, J. A. 2003, *ApJ*, **587**, 638
- Sellwood, J. A. 2006, *ApJ*, **637**, 567
- Sellwood, J. A. 2008a, *ApJ*, **679**, 379
- Sellwood, J. A. 2008b, in *Formation and Evolution of Galaxy Disks*, ed. J. G. Funes SJ & E. M. Corsini (San Francisco: ASP **396**), p. 341 (arXiv:0803.1574)
- Sellwood, J. A. 2010, *MNRAS*, submitted (arXiv:1001.5197)
- Sellwood, J. A. & Binney, J. J. 2002, *MNRAS*, **336**, 785
- Sellwood, J. A. & Carlberg, R. G. 1984, *ApJ*, **282**, 61
- Sellwood, J. A. & Debattista, V. P. 2006, *ApJ*, **639**, 868
- Sellwood, J. A. & Debattista, V. P. 2009, *MNRAS*, **398**, 1279
- Sellwood, J. A. & Evans, N. W. 2001, *ApJ*, **546**, 176
- Sellwood, J. A. & Kahn, F. D. 1991, *MNRAS*, **250**, 278
- Sellwood, J. A. & Lin, D. N. C. 1989, *MNRAS*, **240**, 991
- Sellwood, J. A. & Merritt, D. 1994, *ApJ*, **425**, 530
- Sellwood, J. A. & Moore, E. M. 1999, *ApJ*, **510**, 125
- Sellwood, J. A., Nelson, R. D. & Tremaine, S. 1998, *ApJ*, **506**, 590
- Sellwood, J. A. & Sparke, L. S. 1988, *MNRAS*, **231**, 25P
- Sellwood, J. A. & Valluri, M. 1997, *MNRAS*, **287**, 124
- Sellwood, J. A. & Wilkinson, A. 1993, *Rep. Prog. Phys.*, **56**, 173
- Shakura, N. I. & Sunyaev, R. A. 1973, *A&A*, **24**, 337
- Shen, J. & Debattista, V. P. 2009, *ApJ*, **690**, 758
- Shen, J. & Sellwood, J. A. 2004, *ApJ*, **604**, 614
- Shen, J. & Sellwood, J. A. 2006, *MNRAS*, **370**, 2
- Sheth, K. *et al.* 2008, *ApJ*, **675**, 1141
- Shetty, R., Vogel, S. N., Ostriker, E. C. & Teuben, P. J. 2007, *ApJ*, **665**, 1138
- Shiidsuke, K. & Ida, S. 1999, *MNRAS*, **307**, 737
- Shlosman, I., Frank, J. & Begelman, M. C. 1989, *Nature*, **338**, 45
- Shu, F. H., Tremaine, S., Adams, F. C. & Ruden, S. P. 1990, *ApJ*, **358**, 495
- Skokos, Ch., Patsis, P. A. & Athanassoula, E. 2002, *MNRAS*, **333**, 847
- Soderblom, D. R. 2010, *ARAA*, to appear (arXiv:1003.6074)
- Sparke, L. S. & Casertano, S. 1988, *MNRAS*, **234**, 873
- Sparke, L. S. & Sellwood, J. A. 1987, *MNRAS*, **225**, 653
- Spitzer, L. 1942, *ApJ*, **95**, 329
- Spitzer, L. & Schwarzschild, M. 1953, *ApJ*, **118**, 106
- Stanghellini, L. & Haywood, M. 2010, *ApJ*, **714**, 1096
- Stewart, K. R., Bullock, J. S., Wechsler, R. H., Maller, A. H. & Zentner, A. R. 2008, *ApJ*, **683**, 597
- Sygnnet, J. F., Tagger, M., Athanassoula, E. & Pellat, R. 1988, *MNRAS*, **232**, 733
- Tagger, M., Sygnnet, J. F., Athanassoula, E. & Pellat, R. 1987, *ApJL*, **318**, L43
- Thomasson, M., Elmegreen, B. G., Donner, K. J. & Sundelius, B. 1990, *ApJL*, **356**, L9
- Toomre, A. 1964, *ApJ*, **139**, 1217
- Toomre, A. 1966, in *Geophysical Fluid Dynamics*, notes on the 1966 Summer Study Program at the Woods Hole Oceanographic Institution, ref. no. 66-46
- Toomre, A. 1969, *ApJ*, **158**, 899
- Toomre, A. 1981, in *The Structure and Evolution of Normal Galaxies*, ed. S. M. Fall & D. Lynden-Bell (Cambridge: Cambridge University Press), p. 111
- Toomre, A. 1983, in *IAU Symposium 100, Internal Kinematics and Dynamics of Galaxies*, ed. E. Athanassoula (Dordrecht: Reidel) p. 177
- Toomre, A. 1989, in *Dynamics of Astrophysical Discs*, ed. J. A. Sellwood (Cambridge: Cambridge University Press) p. 153
- Toomre, A. 1990, in *Dynamics & Interactions of Galaxies*, ed. R. Wielen (Berlin, Heidelberg: Springer-Verlag), p. 292
- Toomre, A. 1995, unpublished notes
- Toomre, A. & Kalnajs, A. J. 1991, in *Dynamics of Disc Galaxies*, ed. B. Sundelius (Gothenburg: Göteborgs University) p. 341
- Tóth, G. & Ostriker, J. P. 1992, *ApJ*, **389**, 5
- Tremaine, S. 2005, *ApJ*, **625**, 143
- Tremaine, S. & Weinberg, M. D. 1984a, *ApJL*, **282**, L5
- Tremaine, S. & Weinberg, M. D. 1984b, *MNRAS*, **209**, 729
- Tsoutsis, P., Kalapotharakos, C., Efthymiopoulos, C. & Contopoulos, G. 2009, *A&A*, **495**, 743
- Valenzuela, O. & Klypin, A. 2003, *MNRAS*, **345**, 406
- Vandervoort, P. O. 1970, *ApJ*, **161**, 87
- Vauterin, P. & Dejonghe, H. 1996, *A&A*, **313**, 465
- Velázquez, H. & White, S. D. M. 1999, *MNRAS*, **304**, 254
- Villa-Vargas, J., Shlosman, I. & Heller, C. 2009, *ApJ*, **707**, 218
- Villumsen, J. V. 1985, *ApJ*, **290**, 75
- Voglis, N., Harsoula, M. & Contopoulos, G. 2007, *MNRAS*, **381**, 757
- Wada, K. 2001, *ApJL*, **559**, L41
- Walker, I. R., Mihos, J. C. & Hernquist, L. 1996, *ApJ*, **460**, 121
- Weinberg, M. D. 1985, *MNRAS*, **213**, 451
- Weinberg, M. D. 1991, *ApJ*, **373**, 391
- Weinberg, M. D. 1994, *ApJ*, **421**, 481
- Weinberg, M. D. & Blitz, L. 2006, *ApJL*, **641**, L33
- Weinberg, M. D. & Katz, N. 2002, *ApJ*, **580**, 627
- Weinberg, M. D. & Katz, N. 2007, *MNRAS*, **375**, 425
- Weiner, B. J. 2004, in *IAU Symp.* **220**, *Dark Matter in Galaxies*, ed. S. Ryder, D. J. Pisano, M. Walker & K. C. Freeman (Dordrecht: Reidel), p. 35
- Weiner, B. J., Sellwood, J. A. & Williams, T. B. 2001, *ApJ*, **546**, 931
- Wielen, R. 1977, *A&A*, **60**, 263
- Yoachim, P. & Dalcanton, J. J. 2006, *AJ*, **131**, 226
- Zang, T. A. 1976, PhD thesis, MIT
- Zang, T. A. & Hohl, F. 1978, *ApJ*, **226**, 521
- Zánmar Sánchez, R., Sellwood, J. A., Weiner, B. J. & Williams, T. B. 2008, *ApJ*, **674**, 797
- Zhang, X. 1996, *ApJ*, **457**, 125
- Zhang, X. 1998, *ApJ*, **499**, 93
- Zibetti, S., Charlot, S. & Rix, H.-W. 2009, *MNRAS*, **400**, 1181

## APPENDIX

## RELAXATION TIME IN SPHEROIDS AND DISKS

A test particle moving at velocity  $\mathbf{v}$  along a trajectory that passes a stationary field star of mass  $m$  with impact parameter  $b$  is deflected by the attraction of the field star. For a distant passage, it acquires a transverse velocity component  $|\mathbf{v}_\perp| \simeq 2Gm/(bv)$  to first order (BT08 eq. 1.30). Encounters at impact parameters small enough to produce deflections where this approximation fails badly are negligibly rare and relaxation is driven by the cumulative effect of many small deflections.

If the density of field stars is  $n$  per unit volume, the test particle will encounter  $\delta n = 2\pi b \delta b n v$  stars per unit time with impact parameters between  $b$  and  $b + \delta b$ . Assuming stars to have equal masses, each encounter at this impact parameter produces a randomly directed  $\mathbf{v}_\perp$  that will cause a mean square net deflection per unit time of

$$\delta v_\perp^2 \simeq \left(\frac{2Gm}{bv}\right)^2 \times 2\pi b \delta b n v = \frac{8\pi G^2 m^2 n \delta b}{v b}. \quad (\text{A1})$$

The total rate of deflection from all encounters is the integral over impact parameters, yielding

$$v_\perp^2 = \frac{8\pi G^2 m^2 n}{v} \int_{b_{\min}}^{b_{\max}} \frac{db}{b} = \frac{8\pi G^2 m^2 n}{v} \ln \Lambda, \quad (\text{A2})$$

where  $\ln \Lambda \equiv \ln(b_{\max}/b_{\min})$  is the Coulomb logarithm. Typically one chooses the lower limit to be the impact parameter of a close encounter,  $b_{\min} \simeq 2Gm/v^2$ , for which  $|\mathbf{v}_\perp|$  is overestimated by the linear formula, while the upper limit is, say, the half-mass (or effective) radius,  $R$ , of the stellar distribution beyond which the density decreases rapidly. The vagueness of these definitions is not of great significance to an estimate of the overall rate because we need only the logarithm of their ratio. The Coulomb logarithm implies equal contributions to the integrated deflection rate from every decade in impact parameter simply because the diminishing gravitational influence of more distant stars is exactly balanced by their increasing numbers.

Note that the first order deflections that give rise to this steadily increasing random energy come at the expense of second order reductions in the forward motion of the same particles that we have neglected (Hénon 1973). Thus the system does indeed conserve energy, as it must.

We define the **relaxation time** to be the time needed for  $v_\perp^2 \simeq v^2$ , where  $v$  is the typical velocity of a star. Thus

$$\tau_{\text{relax}} = \frac{v^3}{8\pi G^2 m^2 n \ln \Lambda}. \quad (\text{A3})$$

To order of magnitude, a typical velocity  $v^2 \approx GNm/R$ , where  $N$  is the number of stars each of mean mass  $m$ , yielding  $\Lambda \approx N$ . Defining the dynamical time to be  $\tau_{\text{dyn}} = R/v$  and setting  $N \sim R^3 n$ , we have

$$\tau_{\text{relax}} \approx \frac{N}{6 \ln N} \tau_{\text{dyn}}, \quad (\text{A4})$$

which shows that the collisionless approximation is well satisfied in galaxies, which have  $10^8 \lesssim N \lesssim 10^{11}$  stars. Including the effect of a smooth dark matter component in this estimate would increase the typical velocity,  $v$ , thereby further lengthening the relaxation time.

This standard argument, however, assumed a pressure-supported quasi-spherical system in several places. Rybicki (1972) pointed out that the flattened geometry and organized streaming motion within disks affects the relaxation time in two important ways. First, the assumption that the typical encounter velocity is comparable to the orbital speed  $v = (GNm/R)^{1/2}$  is clearly wrong; stars move past each other at the typical random speeds in the disk, say  $\beta v$  with  $\beta \sim 0.1$ , causing larger deflections and decreasing the relaxation time by a factor  $\beta^3$ .

Second, the distribution of scatterers is not uniform in 3D, as was implicitly assumed in eq. (A1). Assuming a razor thin disk, changes the volume element from  $2\pi v b \delta b$  for 3D to  $2v \delta b$  in 2D, which changes the integrand in eq. (A2) to  $b^{-2}$  and replaces the Coulomb logarithm by the factor  $(b_{\min}^{-1} - b_{\max}^{-1})$ . In 2D therefore, relaxation is dominated by close encounters. Real galaxy disks are neither razor thin, nor spherical, so the spherical dependence applies at ranges up to the typical disk thickness,  $z_0$ , beyond which the density of stars drops too quickly to make a significant further contribution to the relaxation rate. Thus we should use  $\Lambda \simeq z_0/b_{\min}$  for disks. More significantly, the local mass density is also higher, so that  $N \sim R^2 z_0 n$ . These considerations shorten the relaxation time by the factor  $(z_0/R) \ln(R/z_0)$ . An additional effect of flattened distribution of scatterers is to determine the shape of the equilibrium velocity ellipsoid, as discussed in §10.3.

A third consideration for disks is that the mass distribution is much less smooth than is the case in the bulk of pressure supported galaxies. A galaxy disk generally contains massive star clusters and giant molecular clouds whose influence on the relaxation rate turns out to be non-negligible (see §10).