Mon. Not. R. astr. Soc. (1988) **234,** 873–898

A model for persistent galactic warps

Linda S. Sparke and Stefano Casertano *Kapteyn Astronomical Institute, Postbus 800, 9700 AV Groningen, The Netherlands*

Accepted 1988 March 17. Received 1988 February 29; in original form 1987 December 18

Summary. We examine the idea that observed galactic warps may represent discrete modes of bending in a self-gravitating disc. If the disc is subject to the potential of a flattened unseen halo, then discrete modes may exist which are not sensitive to details of the disc edge; the most important parameter is the core radius of the halo. We discuss the variety of warp shapes which can arise in the presence of an oblate dark halo, construct models for two observed warps, and consider how warps may be used to constrain the distribution of unseen matter in galaxies. Isolated discs may also have discrete modes of bending, but only when the density falls to zero over quite a narrow region at the edge; these modes are probably not relevant to the interpretation of observed warps.

1 Introduction

In many spiral galaxies, including our own, the outer parts of the disc are warped. This bending is most often seen in the neutral hydrogen layer (Sancisi 1976, 1983; Bosma 1978) though it is sometimes observed in the stellar disc as well (as in M31, Innanen *et al.* 1982; and NGC 4565, van der Kruit & Searle 1981a). The warps have a characteristic 'integral sign' shape; on one side the disc is bent up above the central plane, and on the opposite side it curves downwards. Bending usually begins at radii of about three to five times the exponential scale length measured for the light distribution; sometimes, but not invariably, the neutral hydrogen layer begins to warp where the stellar disc comes to an abrupt edge. The warps usually do not have a spiral form; the line of nodes, where the warped outer disc crosses the plane of the inner galaxy, is approximately straight (Kerr 1985).

Warps are common, which suggests that, once present, they persist for a long time. However, the bending is observed at radii where a free particle precesses quite rapidly in the gravitational field of the disc (Avner & King 1967). In general the precession rate in a galaxy varies with radius, so that warped structures at different radii precess at different rates; the warp tends to wind up into a tight spiral corrugation. For example, in M83 Rogstad, Lockhart & Wright (1974) estimate that the warp would become noticeably wound within about 2×10^9 yr. This winding can be slowed, though not eliminated, if the galactic disc is surrounded by a spherical massive halo (Tubbs & Sanders 1979), but very large halo masses would be required to maintain a coherent

warp at radii of 10–20 kpc, as in our Galaxy (Henderson, Jackson & Kerr 1982; Kulkarni, Blitz & Heiles 1982).

The visible portions of a galaxy probably contain less than half the mass (Faber & Gallagher 1979; van Albada & Sancisi 1986; Begeman 1987). A number of mechanisms have been put forward by which the unseen mass could produce or maintain a warp. Kahn & Woltjer (1959) suggested that the plane of the Galaxy could be bent by the motion through a hot intergalactic medium; their idea had at least the merit of predicting that the warp of our own Galaxy should be asymmetric (Henderson *et al.* 1982; Kulkarni *et al.* 1982).* Binney (1978, 1981) tried to use a triaxial halo or a bar to pump warps through parametric amplification of the vertical oscillations of the disc; but Sparke (1984b) showed that warps thus generated are unlikely to grow fast enough in self-gravitating discs. Petrou (1980) proposed that the halo might be flattened, with ellipticity varying with radius in such a way as to make the precession frequency of the warp constant. In the absence of a dynamical mechanism to set up the delicate balance required, this idea must be discounted as a general explanation for warps. Bertin & Mark (1980) pointed out that a fast rotating disc in a slowly or non-rotating halo may be unstable to the gravitational equivalent of the Kelvin–Helmholz instability, which could then cause bending waves to grow at the expense of the angular momentum of the disc. Though appealing, their idea has never been made to work in detail (see also Bertin & Casertano 1982).

Dekel & Shlosman (1983) suggested that the observed warps could be produced by a misalignment between the plane of the inner disc and the symmetry plane of a flattened halo; Toomre (1983), putting forward the same idea, pointed out that such a misalignment could persist only if the disc had a discrete normal mode of bending.

A galactic disc with a discrete mode of vertical oscillation could vibrate like a drum under its own gravity. External forces on the disc could excite the bending mode, and once a warp had been set up it would persist for many rotation periods. Thus discrete modes would offer an attractive explanation for the observed galactic warps (Lynden-Bell 1965; Hunter & Toomre 1969). Here we consider the question of bending modes both in an isolated disc and in a disc subject to the potential of a flattened unseen halo. We find that a smooth-edged disc *can* have a discrete bending mode in the presence of an oblate halo, provided that the halo is not too flattened and the disc is not extended too far. From this we proceed to give a *quantitative* basis to the idea that flattened halo potentials can be responsible for many of the observed warps. We are able to produce detailed models for the observed warps of two galaxies, using halo parameters that are consistent with the constraints posed by the rotation curves. In principle, the shape of the observed warp could be used to constrain the distribution of the dark mass in these systems.

2 Bending modes in discs

Isolated self-gravitating discs can have two families of integral-sign (i.e. with azimuthal wavenumber $m=1$) warping modes: the direct modes, which precess faster than the local rotation speed, and the retrograde modes, which regress quite slowly (Hunter & Toomre 1969). The retrograde modes represent a slow precession of the disc, while the direct modes correspond to rapid nutation; the two classes of modes can be easily understood in the limit when self-gravity is weak (see Toomre 1983 for a physical discussion). The slowly precessing retrograde modes are probably most relevant to galactic warps. If the retrograde modes are listed in order of decreasing frequency, the first (R1) is simply a tilt of the whole disc, with zero precession frequency. The next, R2, has one radial node, and is the first non-trivial mode.

*The predicted asymmetry was pointed out by Hunter & Toomre (1969); curiously enough, this was at the time one of the main reasons to dismiss the idea.

Lynden-Bell (1965) found a discrete warping mode in the highly flattened Maclaurin spheroid; but when Hunter (1969a, b) and Hunter & Toomre (1969) investigated the bending modes of a sequence of isolated self-gravitating discs, they found that those discs for which the density tapered smoothly to zero at the outer edge had no discrete warping modes. Their fundamental work, on a sequence of finite discs derived from flattened Maclaurin spheroids, established that discs in which the density $\sigma(r)$ falls smoothly at the edge admit a continuum of warping modes. They gave an argument based on the WKB approximation, showing that a continuum should exist if the quantity $\int dr/\sigma(r)$ was divergent at the disc edge, i.e. if the density decreased linearly or more gradually there. These authors found no discrete modes in the tapered discs which they studied, apart from 'one barely discrete R2 mode for Model 2X' (Hunter & Toomre 1969, p. 762; *cf.* also Hunter 1969a). Their result suggested that real galactic discs, which are unlikely to have sharp edges, probably do not have discrete warping modes, but rather support a continuum of dispersive waves. Time-dependent calculations showed that, if such a disc was initially forced to warp, corrugation waves would be set up, carrying the energy of bending away towards the edge. The warp would become increasingly spiral and confined to the very outer edge of the disc. The only discrete mode of oscillation they found was the trivial mode, a tilt in the plane of the whole disc.

If the disc is surrounded by an unseen halo which is flattened, then the trivial tilt is no longer a mode of the system; a discrete warped mode might be present instead. Toomre (1983) showed that, if the oblate halo was idealized as a distant massive fixed ring, the disc could have a discrete bending mode, which gave reason to hope that the same might be true in a more realistic case. Sparke (1984a) considered the problem in greater detail, looking for discrete modes of bending of a disc within a halo in which the rotation speed was constant at large radii. She found that, for an oblate halo potential, the bending mode always ceases to be discrete as the disc size is increased.

All these warping modes are neutrally stable; they neither grow nor decay with time. Hunter & Toomre (1969) showed that $m=1$ warping modes in an isolated disc were stable, and could not grow; their proof carries over essentially unmodified to a disc in an oblate halo. Even in prolate potentials Sparke (1984a) never found growing modes. Consequently the warp must have a straight line of nodes and no spiral form. This is a prediction of the hypothesis that galactic warps represent *modes* of vertical oscillation, and can be used to test this hypothesis in its general form.

3 Equation of motion and description of the model

3.1 THE EQUATION OF MOTION

This is essentially the same as used by Sparke (1984a), except for the introduction of softening, but we derive it briefly in order to discuss some aspects of its structure.

Consider a cold, thin axisymmetric disc of surface density $\sigma(r)$, rotating in the plane $z=0$ with angular speed $\Omega(r)$ about the axis $r=0$; r, φ, z are cylindrical polar coordinates. When the disc is displaced vertically by a small amount $Z(r,\varphi,t)$, the subsequent motion is described by

$$\frac{D^2Z}{Dt^2} \equiv \left(\frac{\partial}{\partial t}+\Omega(r)\frac{\partial}{\partial \varphi}\right)^2 Z = F_{\text{ext}}+F_{\text{self}} \tag{1}$$

where F_{ext} and F_{self} represent the force per unit mass due to the external gravity and to the self-interaction of the disc material, respectively. This ignores any diffusion of matter from one radius to another, such as might be caused by epicyclic motions in a stellar disc. It also assumes that the disc remains thin, without considering the balance of the cohesive forces which keep it so.

The mass distribution which gives rise to the external potential is taken to be fixed in space; if it

is axially symmetric about $r=0$ and mirror symmetric about the plane $z=0$, the vertical restoring force can be approximated near that plane by

$$F_{\rm ext}(r,\varphi,t)=-\mu_{\rm ext}^2(r)\,Z(r,\varphi,t). \tag{2}$$

The vertical force $F_{\rm self}$ due to the warped disc itself at the point $[r, \varphi, Z(r,\varphi,t)]$ is given, to first order in Z/r, by

$$F_{\rm self}(r,\varphi,t)=G\int_0^\infty \sigma(r')r'\,dr'\int_0^{2\pi} d\varphi'\,\frac{[Z(r,\varphi,t)-Z(r',\varphi',t)]}{[r^2+r'^2-2rr'\cos(\varphi-\varphi')+z_0^2]^{3/2}} \tag{3}$$

where the small softening parameter z_0 is introduced to make the integrand regular at $\mathbf{r}=\mathbf{r}'$. Numerical results given below have been extrapolated to the limit $z_0=0$, which is meaningful provided $Z(r, \varphi, t)$ has continuous spatial derivatives. A form of equation (3) valid when $z_0=0$ exactly is derived in the Appendix and discussed in Section 4.1.

When expressions (2) and (3) are substituted in equation (1), a linear equation for $Z(r,\varphi,t)$ results. A steady warp has the form

$$Z(r,\varphi,t)=Re\{h(r)\exp[i(\omega t-m\varphi)]\}. \tag{4}$$

Since we will always find that ω is real, $h(r)$ can be chosen real as well, representing a pattern with m-fold symmetry that precesses without change of shape at angular frequency ω/m; the observed 'integral-sign' warps correspond to $m=1$, and all our computations refer to this case. Substituting the form (4) for Z into equation (1) yields

$$\{[\omega-m\Omega(r)]^2-\mu_{\rm ext}^2(r)\}h(r)=Gh(r)\int_0^\infty \sigma(r')H(r,r')r'\,dr'$$
$$-G\int_0^\infty \sigma(r')[H(r,r')-K(r,r')]h(r')r'\,dr', \tag{5}$$

where

$$H(r,r')=\int_0^{2\pi}\frac{d\psi}{(r^2+r'^2-2rr'\cos\psi+z_0^2)^{3/2}} \tag{5a}$$

and

$$K(r,r')=\int_0^{2\pi}\frac{(1-\cos m\psi)\,d\psi}{(r^2+r'^2-2rr'\cos\psi+z_0^2)^{3/2}}. \tag{5b}$$

For computational convenience, H and K can be written as sums of complete elliptic integrals. The rotation rate $\Omega(r)$ includes a contribution from the softened self-gravity of the disc, which can be found in terms of integrals over the functions H and K.

3.2 DISC-HALO MODEL

We model the disc as a truncated exponential of unit scale length and unit total mass; in our calculations the velocity scale is chosen so as to set the gravitational constant G to unity. The disc density follows an exponential law out to a radius $r_{\rm trun}$, and is zero beyond a radius $r_{\rm out}>r_{\rm trun}$; between these two radii, the disc tapers smoothly:

$$\sigma(r)=\begin{cases}\exp(-r) & (r\le r_{\rm trun})\\ \exp(-r)\cos^2\left(\dfrac{\pi}{2}\dfrac{r-r_{\rm trun}}{r_{\rm out}-r_{\rm trun}}\right) & (r_{\rm trun}\le r\le r_{\rm out})\\ 0 & (r_{\rm out}\le r)\end{cases}. \tag{6}$$

The density and its first derivative are continuous at both $r_{\rm trun}$ and $r_{\rm out}$, and $\sigma(r)$ falls quadratically to zero at the edge. The taper becomes more abrupt as the radius $r_{\rm trun}$ approaches $r_{\rm out}$.

The disc lies in a flattened unseen halo with density $\varrho_{\rm ext}(r, z)$ approximated by

$$\varrho_{\rm ext}(r, z)=\frac{\varrho_0}{1+a^2/r_{\rm c}^2}, \quad a^2=r^2+\frac{z^2}{(1-\epsilon)^2} \tag{7}$$

where a is the elliptic radius, ϵ the ellipticity of the density contours, and $r_{\rm c}$ the core radius of the halo, inside which the density is nearly constant; far outside this radius the halo rotation curve is flat with rotation velocity v_∞, where

$$v_\infty^2=4\pi G\varrho_0 r_{\rm c}^2\,(1-\epsilon)\,\arcsin\left[\sqrt{1-(1-\epsilon)^2}\right]/\sqrt{1-(1-\epsilon)^2}. \tag{8}$$

The ellipticity ϵ is assumed constant, so that the halo material is stratified on concentric, similar spheroids. The gravitational potential $\Phi_{\rm ext}$ corresponding to the mass distribution of expression (7) can be found by a one-dimensional integral (Schmidt 1956; Burbidge & Burbidge 1975).

4 Analytic results

The dynamical equation (5), in its generality, must be solved numerically. However, the exact results can often be understood better, and even approximately reproduced, by analytic methods. Here we derive an alternative formulation of the dynamical equation for bending modes which is more amenable to analytic study, and use this to understand the bending modes in both isolated discs and discs subject to the force of a flattened halo.

4.1 THE REGULARIZED EQUATION

Using Fourier–Bessel transforms, we can derive a different ('regularized') equation which applies specifically to the unsoftened problem. The detailed steps, which closely follow the treatment of Casertano (1983b), are given in Appendix A. The resulting regularized equation (A18) reads:

$$\{[\omega-m\Omega(r)]^2-\mu_{\rm ext}^2(r)\}h(r)=Gh(r)\int_0^\infty \sigma(r')B_0(r, r')r'\,dr'$$

$$-G\int_0^\infty h(r')\sigma(r')B_m(r, r')r'\,dr'. \tag{9}$$

The kernels B_0, B_m are obtained as integrals over Bessel functions (equation A11) and have only integrable singularities for $r=r'$. Thus it is sufficient that h be continuous for equation (9) to be meaningful.

Equations (5) and (9) are equivalent only if some additional assumptions are made about the disc corrugation. More precisely, the right-hand side of equation (9) is equivalent to the limit of the right-hand side of equation (5) as $z_0\to 0$, provided h is differentiable and has a continuous first derivative. This may be seen by substituting into equation (5) the first-order Taylor expansion for $h(r')$ near $r=r'$; the contributions which are singular in z_0 arise from a neighbourhood of this point, and can be shown to cancel as the softening vanishes. The remaining regular contributions are then easily shown to be the same.

This mathematical statement has direct correspondence to the physical situation. The two separate integrals on the right-hand side of equation (5) represent the (linearized) vertical force of the unperturbed disc at the perturbed position and of the perturbed disc at the unperturbed position, respectively. Both integrands are very large for $r\approx r'$ because they represent the

derivative of the force across the disc; in fact, neither term separately has a meaningful limit as the softening parameter z_0 vanishes. In deriving equation (9), we have avoided the singularities implicitly by taking into account only the curvature of the potential, neglecting the term in the force which is discontinuous across the plane (see equation A13). The rationale for this treatment is that the infinite force gradient associated with the discontinuity affects only the *internal* structure of the disc – it is the 'cohesion' force of Section 3.1 – and will be integrated away in the motion of the disc as a whole [although resonances associated with the vertical structure of the disc might be present, as Bertin & Coppi (1985) find in the somewhat similar problem of the bending of the current sheet associated with the solar wind].

The two force terms on the right-hand side of equation (9) correspond to the terms which Hunter & Toomre (1969, p. 751, equations [12] and [11]) called F_2 and F_1, respectively; these terms do *not* correspond to the terms $\mu(r')h(r, t)$ and $\mu(r')h(r', t)$ in their equation (2). An alternative form for the F_1 term, previously derived by Hunter & Toomre (1969), is:

$$G\int_0^\infty \sigma(r')r'B_0(r, r')\,dr' = 2\Omega_{\rm disc}^2(r) - \kappa_{\rm disc}^2(r) \equiv \mu_{\rm disc}^2(r) \tag{10}$$

where $\Omega_{\rm disc}$ is the equilibrium angular velocity in the gravitational field of the disc alone, and $\kappa_{\rm disc} \equiv \Omega_{\rm disc}\sqrt{4+2d\ln\Omega_{\rm disc}/d\ln r}$ is the corresponding epicyclic frequency. The quantity $\mu_{\rm disc}^2(r)$ measures the vertical curvature of the disc potential away from the disc itself, where the contribution from the local surface density is constant; it is in other words the *non-local* portion of the vertical oscillation frequency. This contribution is well behaved across the disc, except at a sharp edge, where the rotation frequency $\Omega_{\rm disc}$ is not continuous. Note that $\mu_{\rm disc}^2$ is negative in the inner disc, where the rotation velocity $r\Omega_{\rm disc}$ is rising, and positive where $r\Omega_{\rm disc}$ falls.

With the aid of equation (10), and specializing from now on to the case $m=1$ which is most relevant to large-scale warps, we can rewrite the regularized dynamical equation (9) as

$$\{[\omega-\Omega(r)]^2-\mu_{\rm tot}^2(r)\}h(r) = -G\int_0^\infty \sigma(r')r'B_1(r, r')h(r')\,dr' \tag{11}$$

where $\mu_{\rm tot}^2 = \mu_{\rm disc}^2 + \mu_{\rm ext}^2$ is the total vertical frequency of oscillation *except for the local disc contribution*.

4.2 DISCRETE MODES AND THE CONTINUUM

In the short-wavelength limit, bending disturbances can be described in terms of the WKB picture. Hunter (1969b) has shown that in this limit a wave with frequency ω obeys the local dispersion relation

$$(\omega-\Omega)^2 = \mu_{\rm tot}^2 + 2\pi G\sigma|k| \tag{12}$$

where k is the local radial wavenumber and $\mu_{\rm tot}$ has the same meaning as in equation (11); the term $2\pi G\sigma|k|$ represents the self-gravity of the corrugated disc. In this description, a *mode* is a standing wave, which restricts $\int k\,dr$ to a number of discrete values; from equation (12), $|k| \propto 1/\sigma(r)$. As the density $\sigma(r)$ at the edge of the disc is reduced, changing the value of $\int k\,dr$ by an integer number of half-periods has progressively less effect on the frequency ω; the frequency interval separating the Nth mode from the $(N+1)$th decreases. In the limit that the integral $\int dr/\sigma(r)$ diverges, the spectrum of possible eigenfrequencies becomes continuous. The continuum eigenvectors are not smooth, but oscillate infinitely rapidly with increasing amplitude at the edge of the disc.

If the disc is finite, the continuum has a gap. Waves can propagate only where their frequency

$|\omega-\Omega|$ exceeds the free restoring frequency $\mu_{\rm tot}$ of equation (12). If for given ω this condition is not met at the edge, waves with that frequency cannot reach it and do not extend into the region where $\int dr/\sigma(r)$ diverges; the interval

$$\mu_l<\omega<\mu_{\rm u} \tag{13}$$

is free from continuum modes. The bounding frequencies of equation (13) are:

$$\mu_l\equiv(\Omega-\mu_{\rm tot})|_{\rm edge};\quad \mu_{\rm u}\equiv(\Omega+\mu_{\rm tot})|_{\rm edge}. \tag{14}$$

The two frequencies μ_l and $\mu_{\rm u}$ are at the top of the retrograde and the bottom of the prograde continuum respectively. These predictions were confirmed by the numerical experiments of Hunter & Toomre (1969).

The disc may have discrete modes in addition to the continuous spectrum. In a finite disc, the lower frequency μ_l is negative and the upper frequency $\mu_{\rm u}$ is positive; since the orientation of the disc is arbitrary, there is always an isolated point of the spectrum at $\omega=0$, corresponding to the trivial tilt mode. As the disc is extended to infinity, all the frequencies tend to zero; the gap between μ_l and $\mu_{\rm u}$ will disappear, and no other mode can remain discrete.

We have applied the above considerations to finite discs in which the density is truncated smoothly but over a narrow edge region. For these discs, Casertano (1983a) has shown that, near the edge, the rotation velocity has a sudden rise, followed by a sharp fall. By equation (10), the non-local component of the disc-restoring force, identified with the frequency $\mu_{\rm disc}$, will also have a glitch; in particular, its value at the edge will increase. This increases the value of $\mu_{\rm tot}$, and depresses the frequency μ_l which bounds the gap from below. As the width of the truncation region is reduced, the dip in $\Omega-\mu_{\rm tot}$ near the edge becomes increasingly pronounced; over a wider range in frequency, bending waves are prevented from reaching the disc edge. The number of discrete modes should then increase, since there is more room for modes to satisfy the WKB quantization condition with a frequency which does not allow them to penetrate to the edge. In Section 5.2 below we present numerical results on discrete modes in isolated discs; these show that our WKB arguments are qualitatively correct, even when the modes concerned do not have the required rapid spatial variation.

4.3 THE MODIFIED TILT MODE IN ASPHERICAL POTENTIALS

If the disc is placed in an external gravitational field which is not spherically symmetric, the trivial tilt mode of zero frequency will no longer be an exact solution of the dynamical equation. If the asphericity is sufficiently small, however, one might expect a discrete mode to exist with slightly altered shape and frequency. We can estimate the frequency $\omega_{\rm t}$ of this *modified tilt mode* in an axisymmetric flattened potential using standard perturbation theory, with the ellipticity ϵ of the halo density (7) as the expansion parameter. We assume that the change in the tilt mode and its perturbed frequency are both linear in ϵ when this is small; thus $\omega_{\rm t}=\omega_1$, and $h_{\rm t}=h_0+h_1(r)$, where the subscript 1 denotes terms which are of first order in ϵ, and $h_0(r)\propto r$. In equation (11) the only explicit change is in the vertical oscillation frequency $\mu_{\rm ext}$, which, to first order in ϵ, can be written as

$$\mu^2_{\rm ext}(r)=\Omega^2_{\rm ext}(r)+\epsilon\delta^2(r); \tag{15}$$

here $\delta^2(r)$ is a positive-definite function and $\epsilon>0$ for an oblate mass distribution, $\epsilon<0$ for a prolate one. For convenience we rewrite the dynamical equation (11) in the form

$$\{[\omega-\Omega(r)]^2-\mu_0^2(r)-\epsilon\delta^2(r)\}h(r)=\mathscr{T}h(r) \tag{16}$$

where $\mu_0^2(r)=\mu^2_{\rm disc}(r)+\Omega^2_{\rm ext}(r)$, and $\mathscr{T}$ represents the action of the integral operator on the right-

hand side of equation (11). We notice that $\mathscr{T}$ is self-adjoint with respect to the mass-weighted scalar product defined as

$$(f, g)\equiv\int_0^\infty f(r)\bar{g}(r)\sigma(r)r\,dr \tag{17}$$

(the bar indicates the complex conjugate), since, as is easily shown by direct substitution,

$$(f, \mathscr{T}g)=(\mathscr{T}f, g). \tag{18}$$

We now substitute the modified tilt mode solution into equation (16) and linearize to obtain

$$2\Omega\omega_1 h_0=-\epsilon\delta^2 h_0+[\Omega^2-\mu_0^2-\mathscr{T}]h_1. \tag{19}$$

In order to eliminate the term in h_1, we take the scalar product (17) of both sides of (19) with h_0 and make use of the self-adjointness of $\mathscr{T}$. Because h_0 is an eigenfunction of the unperturbed operator with zero eigenvalue, the square bracket applied to it vanishes, and

$$\omega_1(h_0, 2\Omega h_0)=-\epsilon(h_0, \delta^2 h_0). \tag{20}$$

Finally we substitute the zero-order tilt mode $h_0\propto r$ and the definitions of ω_1, δ^2 to derive the linear estimate for ω_t, the frequency of the perturbed mode:

$$\omega_t=\frac{\displaystyle\int_0^\infty[\Omega_{ext}^2(r)-\mu_{ext}^2(r)]\sigma(r)r^3\,dr}{\displaystyle\int_0^\infty 2\Omega(r)\sigma(r)r^3\,dr.}+\mathscr{O}(\epsilon^2). \tag{21}$$

So long as equation (21) is approximately valid, the frequency ω_t of the modified tilt mode is proportional to the flattening ϵ; in an oblate potential $\omega_t<0$ and the mode is retrograde, while in a prolate halo $\omega_t>0$ for a prograde mode. The estimate (21) for ω_t is rather insensitive to the disc parameters, such as its radial size (provided the density vanishes sufficiently rapidly, say as r^{-4}), or to edge details.

A smoothly truncated disc in a flattened halo potential still has a continuum of modes, with a gap in frequency where only discrete modes are possible. The gap is given by equation (13), where the frequencies Ω and μ_{tot} now include a contribution from the halo potential. The WKB argument implies that the modified tilt mode remains discrete if its frequency ω_t falls in the gap between μ_l and μ_u, defined by equation (14); it is discrete if and only if its frequency

$$\omega_t>\mu_l\equiv(\Omega-\mu_{tot})|_{edge}. \tag{22}$$

As the radial size of the disc is extended, $|\mu_l|$ eventually tends to zero. Since, according to equation (21), the mode frequency ω_t is insensitive to the outer parts of the disc, condition (22) will be violated if the disc is sufficiently large.

In Fig. 1 we show the bounding frequency μ_l and the value of ω_t for a sequence of discs subject to the halo potential of equation (7) with $r_c=2.0$, $v_\infty=0.71$ and $\epsilon=0.2$. When the disc is sufficiently large, the modified tilt mode disappears into the continuum.

As for the discrete modes of isolated discs, the finite size of the disc plays a fundamental role here; as Sparke (1984a) found, no infinitely extended disc will have a discrete warping mode in a realistic flattened halo potential, where the frequencies Ω and μ_{ext} vanish at infinity. Toomre (1983) did find an apparently discrete warping mode for an extended disc in the tidal field of a distant massive ring; but that field exerts a torque on the twisted disc material which increases with radius. The frequencies μ_l and μ_u do not tend to zero as the disc becomes larger, and the gap remains open.

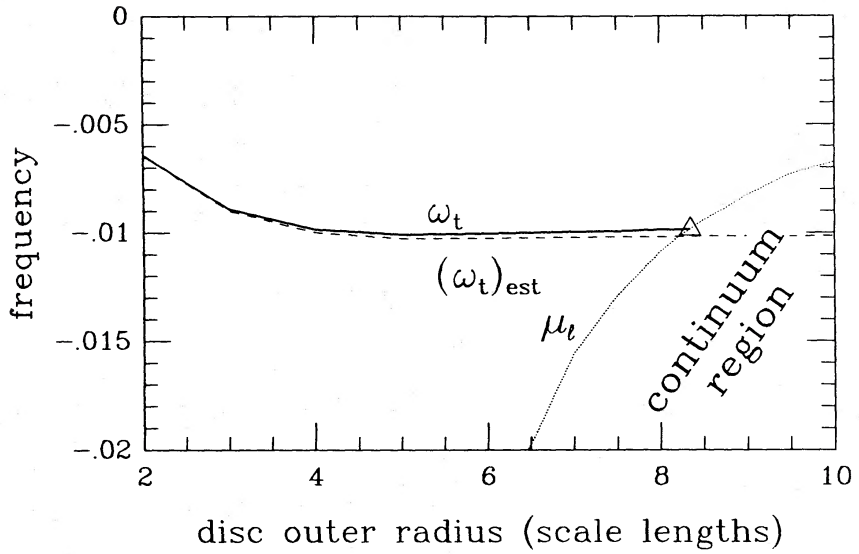

Figure 1. The frequency ω_t of the modified tilt mode as a function of the outer radius of the disc r_{out}, with $r_{trun}=r_{out}-1$. As the disc is extended, the frequency μ_l which bounds the retrograde continuum from above (dotted line) approaches zero. When $\omega_t<\mu_l$, the modified tilt mode is no longer discrete. For this model, the analytic prediction for ω_t (dashed) matches its numerical value, computed as described in Section 5 (solid line), within 5 per cent.

Increasing the flattening ϵ of the halo has the same effect as extending the disc; the tilt mode merges into the continuum. This is because the mode frequency ω_t decreases faster with flattening than the edge frequency of the retrograde continuum μ_l. At some ellipticity ϵ_{crit} the condition (22) will be violated and the disc has no discrete mode. In Fig. 2 we plot the approximate analytic prediction (21) and the numerical value of ω_t, with the value of μ_l, for the same halo potential as in Fig. 1, except that the ellipticity is varied; the perturbation theory implies that the modified tilt mode is no longer discrete if the ellipticity ϵ exceeds a value of about 0.5.

Notice that the retrograde continuum bound μ_l is much closer to zero than the prograde continuum limit μ_u, so that the negative side of the gap is much narrower than the positive. Thus in a prolate potential the disc can be extended to larger radii than in an oblate halo while retaining a discrete mode.

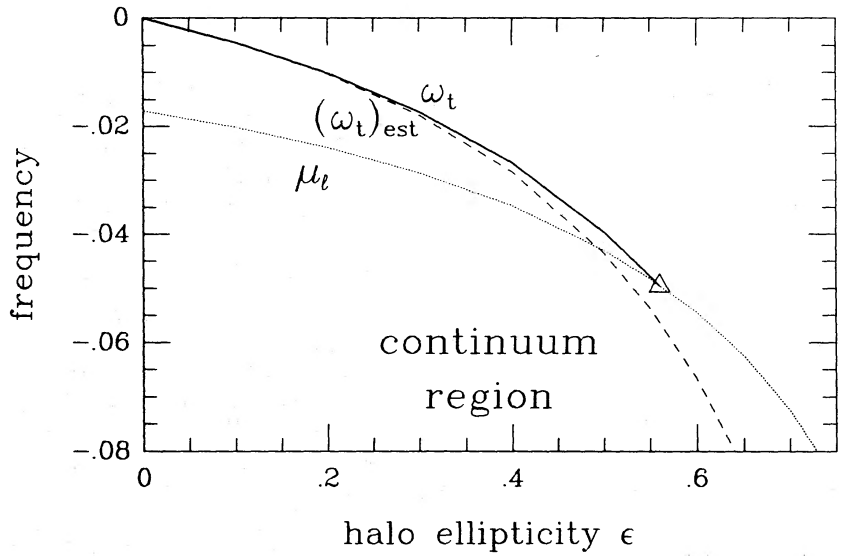

Figure 2. As Fig. 1, but now the disc size is kept constant ($r_{out}=6$), while the halo ellipticity ϵ is varied. The value ϵ_{crit} of the eccentricity at which the tilt mode merges into the continuum region is predicted with about 10 per cent accuracy.

5 Numerical results

5.1 NUMERICAL PROCEDURE

The integral equation (5) is an eigenvalue equation for the eigenfrequency ω and the eigenfunction $h(r)$ which describes the disc bending. The numerical treatment is exactly as described in Sparke (1984a). If the solution is required at a number of radial points $\{r_1, r_2, \ldots, r_N\}$, then the integral can be approximated as a sum over $\{h(r_1), h(r_2), \ldots, h(r_N)\}$; integrating by the trapezoidal rule amounts to treating the continuous disc as a collection of massive rings. The result is an N-dimensional eigenvalue equation quadratic in the eigenvalue ω; this can be transformed into a linear problem of twice the dimension and solved by standard methods. The rings need not be evenly spaced in radius; in the calculations reported below, they were concentrated near the outer edge of the disc to improve resolution there.

The spectrum of our eigenmode problem is partly continuous. The eigenmode $h(r_i)$ corresponding to a discrete mode for the discretized system will converge to a continuous function $h(r)$ as the number N of rings is increased. However, eigenvectors in the continuum are singular at the edge of the disc (see Section 4.2); numerically, this means that the solutions to the N-ring problem fail to converge as N increases, and the eigenvectors are large only at the last few radial points. All the precession rates given below *for discrete eigenmodes* have been extrapolated to the limit of an infinite number of rings.

5.2 ISOLATED DISC

The WKB analysis of Section 4.2 suggests that if an isolated disc is truncated ever more sharply at its edge, an increasing number of discrete warping modes will arise. We show that this is indeed the case for a sequence of discs described by equation (6), with the outer edge set at $r_{\rm out}=5$. As the radius $r_{\rm trun}$, at which the disc density starts to fall away from an exponential law, approaches $r_{\rm out}$, the taper becomes increasingly abrupt. For these discs, the density falls quadratically to zero at the edge, so we expect to see a continuum of warping modes for any value of $r_{\rm trun}$.

When the truncation occurs over a sufficiently narrow region at the edge, a few of the retrograde modes stay well apart in frequency and converge to smoothly curving shapes as the number of rings in the calculation increases; their limiting forms are shown in Fig. 3 for three representative values of the truncation radius. It appears that these smooth-edged discs, together with a continuum of warping modes, *can* also have discrete modes, provided the taper is confined to the very edge of the disc.

Once a mode has separated itself from the continuum, its frequency and shape remain approximately the same as the tapering length decreases further. This behaviour agrees with the WKB analysis described in Section 4.2. Retrograde bending modes of frequency ω cannot propagate into regions of the disc where $\omega>\Omega-\mu_{\rm tot}$; a mode with $\omega>\mu_i\equiv\Omega-\mu_{\rm tot}|_{\rm edge}$ cannot reach the disc edge and must remain discrete. In Fig. 4 we show the limiting frequency $\Omega-\mu_{\rm tot}$ for the same sequence of discs as in Fig. 3, with the frequencies of the three discrete modes marked for comparison.

We now understand why Hunter & Toomre (1969) did not see these discrete modes. The density taper in their smoothly truncated models (Series X, XX, XXX) was specified by removing the appropriate polynomial in $\xi=\sqrt{1-r^2}$ from the density law. Thus the truncation was never confined to a narrow radial region; in fact, from their fig. 2 we see that the tapering affects at least 40 per cent of the disc. Such a taper width is too large to produce a cavity sufficiently deep for modes to be trapped.

If the discs of our sequence are extended using *test particles only* beyond 5 scale lengths, the frequency of the R2 mode, and its shape in the inner mass-carrying portion of the disc, persists

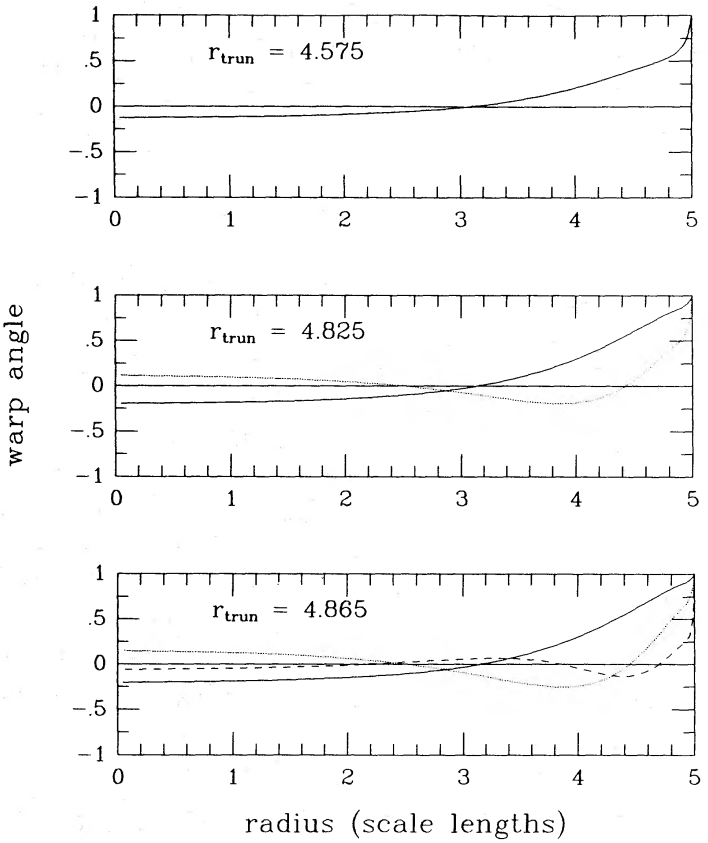

Figure 3. Shape of the discrete retrograde modes of bending for isolated disc models with truncation confined to a narrow edge region (r_{trun}=4.575, 4.825 and 4.865, from top to bottom; r_{out}=5 for all). The warp angle has been normalized to unity at the outer edge. Besides the trivial tilt mode, these models have 1, 2 and 3 discrete modes, respectively. Once separated from the continuum, the modes change little in shape and frequency as the width of the truncation region decreases (top to bottom).

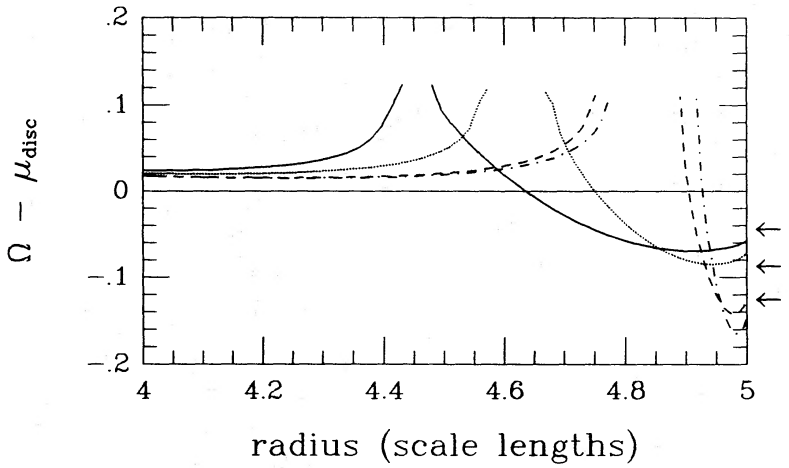

Figure 4. Plot of the limiting frequency $\Omega-\mu_{disc}$ in the edge region of the isolated discs of Fig. 3, plus one with a wider edge region and no non-trivial discrete mode. The curves correspond to r_{trun}=4.4, 4.575, 4.825 and 4.865 for the solid, dotted, dashed, and dot-dashed lines respectively. In the WKB limit, retrograde bending waves can only propagate where their frequency ω is smaller than $\Omega-\mu_{disc}$. The increasingly deep dip in the latter quantity near the edge prevents more waves from reaching the continuum-generating smooth edge. Arrows mark the frequencies of the three modes found for r_{trun}=4.865. The curves are interrupted where $\mu^2_{disc}=\mu^2_{tot}$ is negative; in those regions there is no limit on the frequency of a propagating wave.

unchanged until the disc reaches a radius $r_* \approx 5.2$. The massless outer disc becomes increasingly curved as $r \to r_*$; eventually the gravity in the inner disc ceases to be sufficient to pull the outermost particles around with it in its precession. If more test particles are added outside the radius r_*, there is no longer a mode of the whole disc with the frequency characteristic of the inner regions.

5.3 DISC WITHIN A FLATTENED HALO

If a non-spherically symmetric external gravitational field, such as that of an oblate halo, is applied to the disc, the first retrograde mode R1 is no longer a uniform tilt, but becomes warped. According to the results of Section 4.3, the R1 mode in a tapered disc will remain discrete if the halo is only slightly flattened. A massive disc in an axisymmetric but somewhat oblate unseen halo might then have a discrete warping mode, even if the isolated disc had only the trivial tilt mode.

Consider the smooth-edged exponential disc of equation (6) with the outer edge at $r_{\rm out}=6$ and density tapering to zero beyond a radius $r_{\rm trun}=5$; when this is isolated, it has no discrete bending modes apart from the trivial tilt mode. Now place it in a halo described by equation (7), with ellipticity $\epsilon=0.2$, core radius $r_c=2.0$ and $v_\infty=0.71$; these are the same parameters as used for the model presented in Fig. 2. The total rotation curve is then approximately flat over a large fraction of the disc. The tilt mode R1 is still the only discrete mode, but now it has become warped (solid line in Fig. 5a). The dashed lines show how this modified tilt mode changes shape as the edge radius $r_{\rm out}$ changes; the shape and frequency of the mode are fairly insensitive to the radius of the disc edge, until a limiting value ($r_{\rm out}\approx 8.0$) is reached at which the tapered disc has no discrete mode. This is as expected from the arguments of Section 4.3; from Fig. 1, the mode frequencies are within 5 per cent of those predicted using the perturbation theory. Similarly, if the halo is made increasingly heavy or flattened, the mode will eventually disappear. Fig. 2 showed the analytic prediction for the way the mode frequency ω_t varies as the ellipticity ϵ of the halo is increased; the modified tilt mode ceases to be discrete at a critical ellipticity $\epsilon_{\rm crit}$ which matches the analytic estimate to within 10 per cent. Thus a disc subject to the potential of a flattened halo can have a discrete warping mode, even if the same disc would have only the trivial tilt mode when isolated; the frequency of this mode, and the regions of parameter space where it exists, are close to the predictions of the perturbation theory of Section 4.3.

The *shape* of the warped mode depends on how the halo mass is distributed. In Fig. 5b we show again the disc of Fig. 5a with $r_{\rm out}=6$, this time in a halo of much larger core radius ($r_c=6$ scale lengths); again, the rotation curve is approximately flat. The warp shape is *qualitatively* different in the two cases. If the core radius r_c is small, the disc is bent *upward* with respect to a constant slope; if r_c is large, we have a *downward* bend, in which the tilt angle with respect to the halo plane decreases outward. We call these two types of warps Type I and Type II, respectively.

The physical reason for the different behaviour of the disc in Type I and II warps can be qualitatively understood by considering the disc as a set of concentric massive rings in the halo field. If the halo potential is aspherical, it exerts a torque on each ring that will make it precess with some angular frequency $\Omega_{\rm prec}(r)$. This frequency is approximately given by $\Omega_{\rm prec}(r)\approx(\Omega_{\rm ext}^2-\mu_{\rm ext}^2)/2\Omega$, and is negative when the potential is oblate. If the warp is to precess rigidly, the bending of the disc must compensate for the radial variation of $\Omega_{\rm prec}$. In an external potential with a small core radius, the rings will tend to regress rapidly in the centre ($\Omega_{\rm prec}$ large and negative), and more slowly further out. The self-gravity of the disc must make this regression slower in the inner parts, and faster in the outer parts. Since a tilted ring causes a *regression* of a more inclined one and a *progression* of a less inclined one, this balance is obtained by increasing the disc tilt outward; this is a Type I warp. On the other hand, if the halo core radius is large, the quantity $\Omega_{\rm ext}^2-\mu_{\rm ext}^2$ will be nearly constant, while the rotation rate Ω decreases outwards because

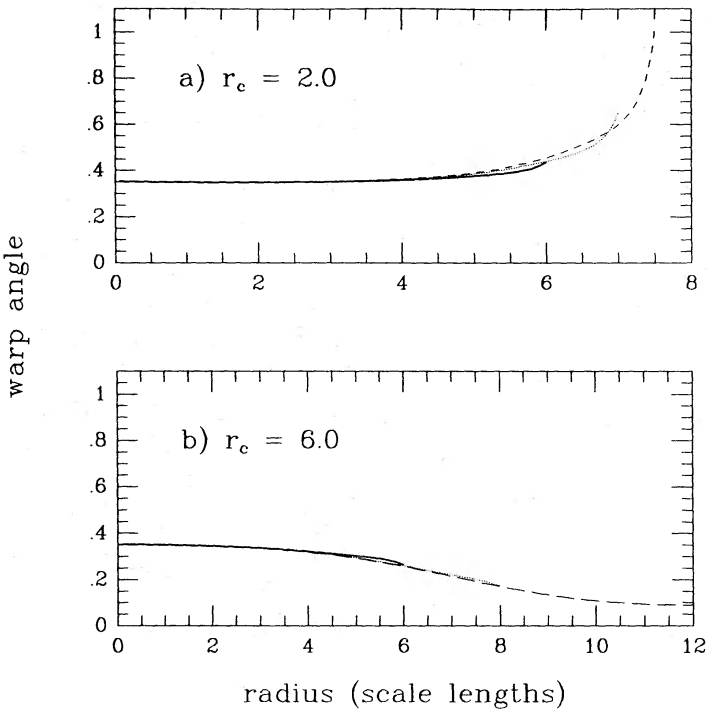

Figure 5. Shape of the modified tilt mode in the disc-halo models described in the text. In panel (a), the halo has r_c=2.0, v_∞=0.71, ϵ=0.2; the disc has unit mass and extends to R_{out}=6, 7, and 7.5 scale lengths (solid, dotted and dashed line respectively), with $r_{trun}=r_{out}-1$. The shape of the mode changes little as the disc extends outwards, until a critical radius is reached. In panel (b), the halo has a larger core radius (r_c=6, v_∞=1.05, ϵ=0.2), and the disc can be extended further; the three models shown are truncated at 6, 8 and 12 scale lengths respectively.

the disc contribution falls. The regression is then slower in the centre than at one or two disc scale lengths; the balancing disc contribution will have the opposite trend, and the disc bends down towards the plane of the halo in a Type II warp. If the outer radius of the disc r_{out} is increased, this trend is eventually reversed; the outer edge of the disc begins to curl upward again, since the derivative of the free precession rate changes sign. The downturn begins where the rotation curve of the disc turns over; the upturn starts at about two halo core radii, where the halo frequency drops significantly from its central value.

If the disc is extended further, even using *test particles only,* the warped mode eventually disappears, exactly like the modes in isolated discs discussed in Section 5.2 above. The frequency of the mode, and the shape in the mass-carrying inner portion, persist unchanged until the disc reaches some radius r_* ($\approx$7.3 for the model in Fig. 5a, 15.4 for that in Fig. 5b). The massless outer disc becomes increasingly sharply bent as $r \rightarrow r_*$, and eventually the outer material can no longer keep up with the precession of the inner warp. The radius r_* is smaller in a halo with a small core radius because the warp precession frequency, which is characteristic of the inner disc (see Section 4.3), is larger in absolute value; this fast regression is more difficult to sustain at large radii.

Our results show that realistic galactic discs which lie in a massive flattened halo can have an isolated bending mode, and so sustain a long-lived warp. The rest of this paper contains a detailed investigation of this mode, the modified tilt mode, and its possible astrophysical relevance. Most of our calculations are based on the assumptions discussed in Section 3 above; we should mention here some of their weaknesses.

The most important difficulty is perhaps the assumption of a rigid halo, whose shape and orientation would survive unchanged the gravitational field of a precessing, flat massive disc. This

is probably not true in the central regions, where the disc field is most important; as Toomre (1983) pointed out, even a hot pressure-supported halo reacts to the potential of the thin precessing disc. It is likely that this reaction will effectively reduce the precession due to the halo in the inner galaxy; the halo may become more spherical and its plane may realign partially with that of the disc. We can mimic this in our models by reducing the ellipticity of the fixed halo in the inner parts. The results are similar to those found when the core radius is increased, and no new behaviour is seen. But it should be kept in mind that in a fully self-consistent kinetic treatment new phenomena might appear; in particular, resonances may cause the warp to grow or decay.

There is also a problem with the sharply-curved Type I warps (Fig. 3a), which would be sensitive to dissipative processes in the gas, or to radial mixing if a stellar population was affected. Our tilted-ring model disregards such phenomena; it remains to be seen what fraction of the parameter space for Type I warps is actually accessible to more realistic discs.

Finally, all our calculations are linear in amplitude, while some of the observed warps turn through relatively large angles. But similar work on polar rings (Sparke 1986) suggests that this is not a worry; the shape changes only slightly as the warp angle varies up to 30° from the halo equator.

While the above problems should be investigated, we feel that it is worth while at this stage to consider the implications of our simple model, particularly to discuss the feasibility of the theory and suggest possible observational tests.

6 The modified tilt mode and observed warps

In this section we discuss whether the modified tilt mode can provide the basis for a general theory of warps in galactic discs. If we are to explain a particular observed warp in terms of the modified tilt mode, then for the combination of disc and halo chosen, two rules must be satisfied. They are:

R1: The modified tilt mode must be a *discrete* mode of warping in the disc even when the disc edge is not sharp. This is a mathematical condition which implies that the warp can be long-lived.

R2: The modified tilt mode must be sufficiently different from a rigid tilt, so that we see a warp when we observe the system.

Even when the two rules above are satisfied, we may still have to reject the modified tilt mode as an explanation: for example, the predicted shape may not look like the observed warp. But the two conditions above are necessary; systems which violate them cannot maintain a warp corresponding to the modified tilt mode. On the other hand, a system which satisfies both **R1** and **R2** will not actually show a warp unless the long-lived mode has also been excited at some time, perhaps as the galaxy was formed.

We shall first discuss when **R1** is satisfied; then we turn to **R2** and its implications, indicating the shape of a typical modified tilt mode. For definiteness, we shall restrict our discussion to the combination of a smoothly truncated exponential disc and an oblate halo of constant ellipticity, with the density law given in equation (7).

6.1 DISCRETE?

A long-lived warp can correspond only to a discrete bending mode (rule **R1**). Hunter & Toomre (1969) illustrate what happens when no discrete mode is present. If a disc which has only a continuum of bending modes is initially forced to warp, it becomes progressively more strongly curved at the edges, while the inner region becomes flat. Dissipative effects in the stars or the gas will rapidly eliminate this outer, highly corrugated warp. By contrast, a disc with a discrete bending mode can oscillate without change of appearance once this mode has been excited.

The perturbation theory equations (21) and (22) can be used to determine whether the

modified tilt mode is discrete in any particular combination of disc and flattened halo. To illustrate this, we consider our standard model in which the disc density is given by equation (6) truncated over the last scale length ($r_{\rm trun}=r_{\rm out}-1$) while the halo density follows equation (7). This model has three parameters: the core radius $r_{\rm c}$ and the ellipticity ϵ of the halo, and the ratio of halo to disc mass, measured within the fiducial radius of 6 disc scale lengths: $f\equiv M_{\rm h}(6)/M_{\rm d}$.

For fixed values of the quantities $r_{\rm c}$ and f, the modified tilt mode will remain discrete if the ellipticity of the halo does not exceed a critical value $\epsilon_{\rm crit}$; in Fig. 6 we plot our estimate of $\epsilon_{\rm crit}$ as a function of the halo mass fraction f, at fixed core radius $r_{\rm c}$. The solid curve is for a model with a small disc and a small halo core radius ($r_{\rm out}=6$, $r_{\rm c}=2$). In general, the more massive the halo, the smaller the critical value of the ellipticity because increasing either the halo mass or its flattening has qualitatively the same effect on the mode. For this particular model, a mode is present for moderate halo flattening ($\epsilon<0.3$) unless the halo mass is more than four times that of the disc. The dotted curve corresponds to a model with a larger core radius ($r_{\rm c}=3.5$); this increases the value of $\epsilon_{\rm crit}$, so that even when $f=5$ a mode is present for $\epsilon<0.4$. If the disc is extended (dashed curve, corresponding to $r_{\rm out}=9$, $r_{\rm c}=3.5$), the critical ellipticity is reduced. When the disc is large, the mode rapidly disappears as the halo core radius is decreased.

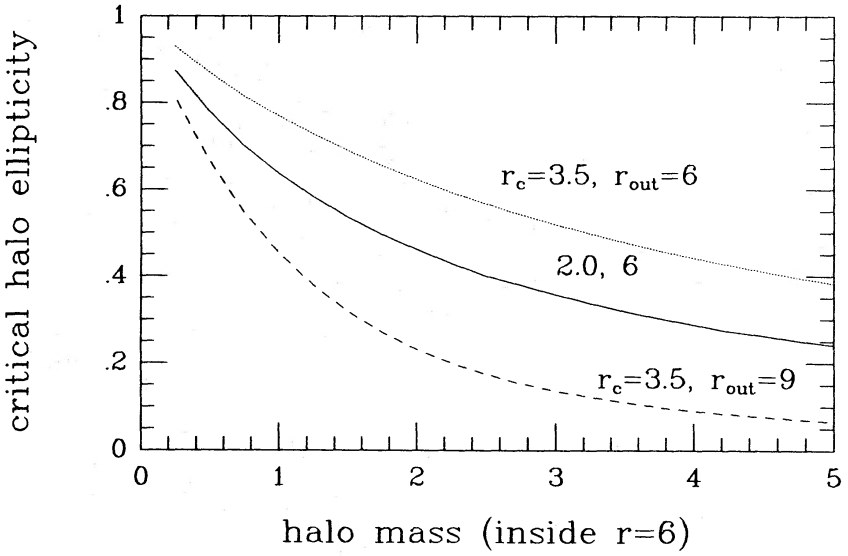

Figure 6. The critical halo ellipticity $\epsilon_{\rm crit}$ at which the modified tilt mode ceases to be discrete, as a function of halo mass, for different disc-halo models. The solid curve corresponds to halo core radius $r_{\rm c}=2$ and disc size $r_{\rm out}=6$; the dotted curve to $r_{\rm c}=3.5$, $r_{\rm out}=6$; and the dashed curve to $r_{\rm c}=3.5$, $r_{\rm out}=9$.

By rule **R1**, we require that the modified tilt mode should be discrete in a galaxy where a warp is observed. In the framework of our model, this requirement puts explicit constraints on the distribution of mass in the halo.

6.2 WARPED?

In a flattened halo, the shape of the warping mode differs from a rigid tilt; this is fortunate, as otherwise no warp will be observed! Note that the symmetry plane of the external potential cannot be seen; all warp observations are referred to the plane of the inner disc. To illustrate the importance of this distinction, we show in Fig. 7 the modified tilt mode for a number of models.

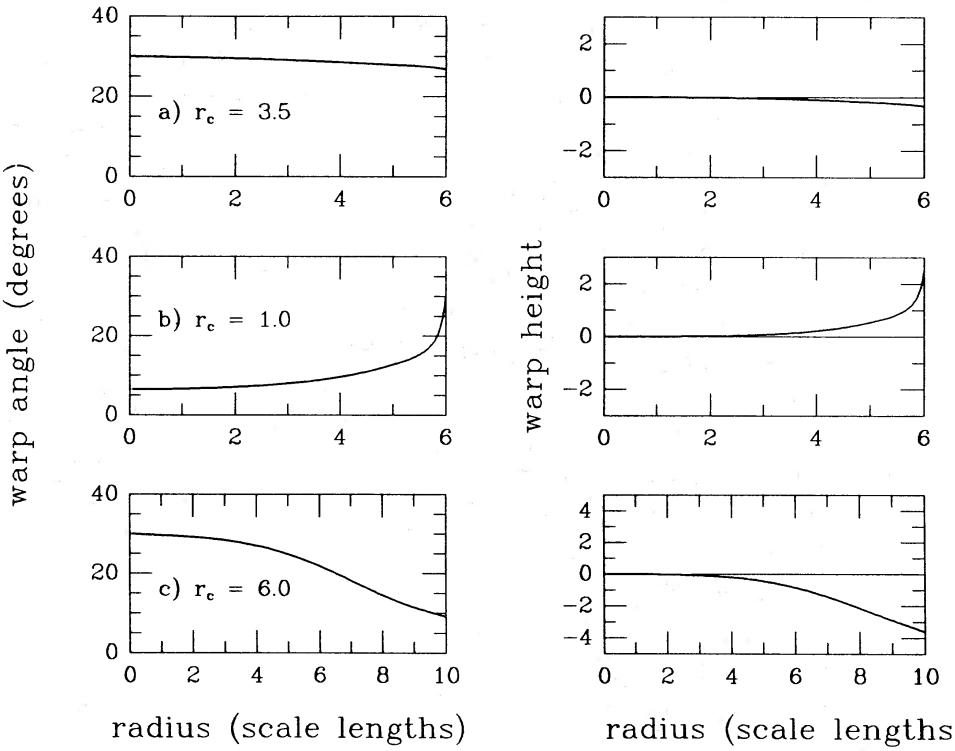

Figure 7. Modified tilt mode in three disc-halo models. In the left panel, we give the angular amplitude of the warp with respect to the (invisible) midplane of the halo. In the right-hand panel, we show the height above the midplane of the inner disc, which depicts the warp as it would be seen by an optimally placed observer, viewing the warped disc from its line of nodes. All these warps are scaled so that the maximum deviation from the halo midplane has an arbitrary value of 30°. In each model, the disc extends to the last point shown, with $r_{trun}=r_{out}-1$; the halo core radius r_c is as shown, and the halo mass is chosen so as to obtain a flat rotation curve. The three cases show: (a) an almost planar mode, corresponding to a scarcely observable warp; (b) an upward-tilting mode (Type I); (c) a downward-tilting mode (Type II). Note that even in case (c) the warp would be difficult to observe if the disc extended only to $r_{out}=6$.

On the left-hand side, these warps are drawn with respect to the (invisible) symmetry plane of the halo. On the right-hand side, the reference plane has been rotated so that the warps are measured from the (observable) plane of the inner disc. In all cases, the disc extends out to the last point plotted, with $r_{trun}=r_{out}-1$.

In the linear theory of this paper, the warp amplitude is arbitrary; the warps we show are given a peak amplitude of 30° with respect to the halo midplane, a rather dramatic tilt. In model (a) (top row), even such a large intrinsic tilt is observed as only a 5° bend; a real galaxy with this combination of disc and halo could never show a long-lasting warp of more than a few degrees. The other two models show much greater deviations from the rigid tilt, and significant amplitude when referred to the plane of the inner disc; galaxies like these might retain large warps as a relic of past disturbance.

Rule **R2** states that, if the warp observed in a particular system is to be explained in terms of the modified tilt mode, then the curvature of the mode shape must be sufficiently large. The amount of bending required depends on the specific galaxy; some systems, like NGC 3198 (Begeman 1987), have warps of only a few degrees, and even a mode like that of Fig. 7a would satisfy rule **R2**. Others, like NGC 4013 (Bottema, Shostak & van der Kruit 1987), show an observed warp as large as 30°; for these the deviation from a rigid tilt must be far greater. This deviation can be either away from (Fig. 7b) or towards (Fig. 7c) the halo equator; we call these Type I and Type II warps respectively (see Section 5.3).

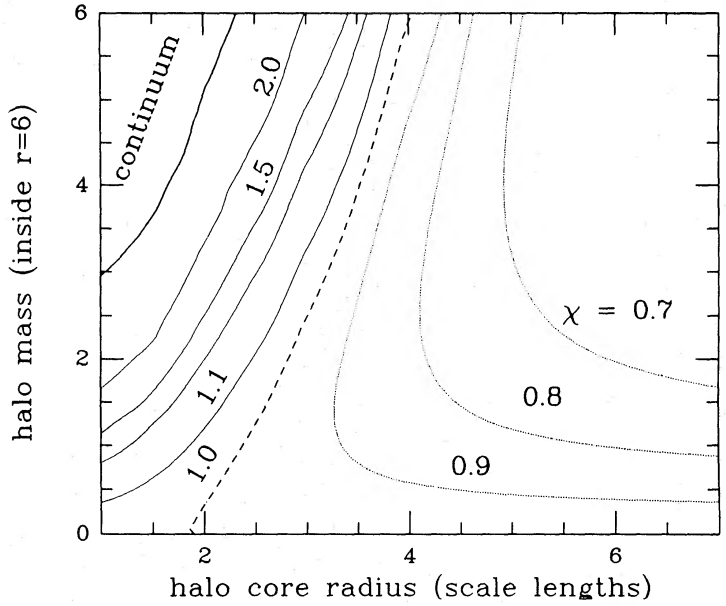

Figure 8. Contour levels of the shape parameter $\chi=\theta(\text{edge})/\theta(0)$ of the modified tilt mode in a system made of a disc smoothly truncated at $r_{out}=6$ ($r_{trun}=5$) and a flattened halo with ellipticity $\epsilon=0.2$, as a function of the halo core radius and mass. The bending $1-\chi$ is proportional to the deviation of the mode from a rigid tilt, and is positive for a Type I mode, negative for a Type II. The dashed contour separates the two types; systems near this line will show very little warping. The heavy solid contour corresponds to the continuum; systems beyond this line do not have a discrete modified tilt mode.

The bending of the modified tilt mode can be described in a first approximation by the quantity $\chi=\theta(\text{edge})/\theta(0)$. This dimensionless number will be larger than unity for an upward tilt (Type I), and below unity for a downward one (Type II). For a slightly oblate halo with $\epsilon \ll \epsilon_{crit}$, we expect that the bending $(1-\chi) \propto \epsilon$. Rule **R2** requires that χ be sufficiently far from unity.

In Fig. 8 we illustrate the dependence of χ on the core radius and the mass of the halo for a disc extending to $r_{out}=6$ and a plausible fixed value $\epsilon=0.2$ for the halo ellipticity. Three main regions can be identified. On the upper left (heavy halo with small core radius) we have the 'continuum' region; here $\epsilon_{crit}<0.2$, so the modified tilt mode is not discrete in these potentials. Systems falling in this region violate rule **R1**, and cannot develop persistent warps. Adjacent to this we have a region of Type I warps (upward-bending, $\chi>1$); these correspond to a halo with a small core radius but which is not too heavy. The approach to the continuum is through very large values of χ, which reflects the fact that continuum modes are singular at the edge. On the right we have a region of Type II warps (downward-bending, $\chi<1$), corresponding to a halo with a large core radius. These two regions are separated by the line $\chi=1$ (heavy dashed in the plot); near this line, the deviation from the rigid tilt is very small, and rule **R2** is violated: even large tilts away from the halo equator correspond to very small observable warps. Thus our warp model allows a specific prediction: galaxies with significant warps must fall in either region I or region II, and sufficiently far from the dividing line.

To test this prediction in a particular observed galaxy, we must construct a model for the disc and halo. In general, the rotation curve of a particular galaxy can be fit by a one-parameter family of disc-halo models; the mass and core radius of the halo cannot be constrained separately. Models in this sequence correspond to a one-dimensional curve in the plane of a diagram like Fig. 8. If this curve intersects either of regions I or II sufficiently far from the dividing line $\chi=1$, then we predict that the galaxy *may* show an observable warp; otherwise, it *cannot* do so in the framework of our theory. The predicted shape can then be compared with the detailed observed shape.

6.3 THE SHAPE OF THE MODE

A disc in an oblate halo can show two different types of bending mode; the outer parts of the disc may deviate *away* from the midplane of the halo or *towards* it. In Section 5.3 we have called these two families Type I and II warps respectively; if the halo has a small core radius then the warp is of Type I, while if the core is large the warp is usually of Type II. [The warps envisaged by Dekel & Shlosman (1983) and Toomre (1983) are basically Type II warps.]

The intrinsic difference between these two classes, whether the disc bends towards the halo equator or away from it, is not observable, as the halo plane itself is not seen. But there are secondary differences which help in discriminating between them. Type I warps usually show an abrupt change in angle near the outer edge of the disc, while Type II warps begin around 3 or 4 scale lengths where the disc rotation curve falls, and can extend over half the disc. Galaxies with a small disc are likely to show Type I warps; our model below for NGC 4565, in which the disc ends within 6 scale lengths, shows an appreciable warp only if the halo has a small core radius. With a large core radius, the warp would be of Type II, but the bending would be too small to fit the observations.

A galaxy with a large disc cannot maintain a long-lived warp unless the halo core is large; otherwise the warped mode is not isolated. If a warp is present it is likely to be of Type II; we shall see this in our model for NGC 4013, which has a disc extending to 9 scale lengths.

If the core radius does not fall in the required range, a very small warp or none at all will be present. This might be the case for galaxies like NGC 891 or 3198, which appear similar to NGC 4565 and 4013 respectively, but show a much less substantial warp.

7 Two examples: models for NGC 4565 and 4013

Here we apply our model for long-lived galactic warps to the warps actually observed in the galaxies NGC 4565 and 4013. These are both fairly isolated galaxies where the warp has a significant amplitude, and are nearly edge-on, so that the warping shape can be inferred without using kinematic deprojection.

With the aid of the numerical procedure described in Section 3.2, we have computed warp modes for discs corresponding to the specifications of these two galaxies. For each we have generated a family of models, parametrized by the core radius r_c of the halo. The halo mass is then determined by fitting the rotation curve. In order to limit the number of parameters, we have considered only the ellipticity $\epsilon=0.2$; the shape of the modified tilt mode is not very sensitive to ϵ.

7.1 NGC 4565

NGC 4565 is a relatively large Sb galaxy, viewed nearly edge-on; the stellar disc has a scale length of 5.5 kpc and appears to be truncated at about 4.5 scale lengths (van der Kruit & Searle 1981a). The H I extends about 1.2 scale lengths farther out, and rotates with a maximum velocity in the outer galaxy of about 250 km s^{-1}. Near the edge of the luminous disc, the gas layer deviates rather abruptly from the plane of the inner disc; the warp reaches a maximum amplitude of about 9° (Sancisi 1976; Sancisi, van Albada & Casertano, in preparation). The optical disc also appears slightly warped near its edge (van der Kruit 1979).

The warp is clearly visible only in the NW side of the galaxy; on the SE side, the H I runs out before reaching the radius at which the warping is most prominent on the NW side. The observations are consistent (Sancisi *et al.* in preparation) with the same warped plane on both sides of the galaxy, for both stars and gas; only the densities differ. Since the gas at the outermost radii is dynamically unimportant, we will treat the warp as if it were symmetric; the problem of

lopsidedness of the gas disc is beyond the scope of the present work (see Baldwin, Lynden-Bell & Sancisi 1980).

To model the warp in NGC 4565, we assume that the mass distribution in the disc follows the exponential law (6), with $r_{trun}=4.5$ scale lengths and $r_{out}=4.7$ scale lengths; the disc is continued to 5.7 scale lengths using test particles to represent the gas. The core radius of the halo is varied between $r_c=4$ and 32 kpc; for each value of r_c, the halo mass is chosen by a best-fit to the outer part of the rotation curve (Sancisi *et al.* in preparation). No bulge component was included in our model; the bulge is likely to be more closely coupled with the disc than with the halo, and thus behave like a thickened central region of the disc. This would have little effect on the inferred warp shape.

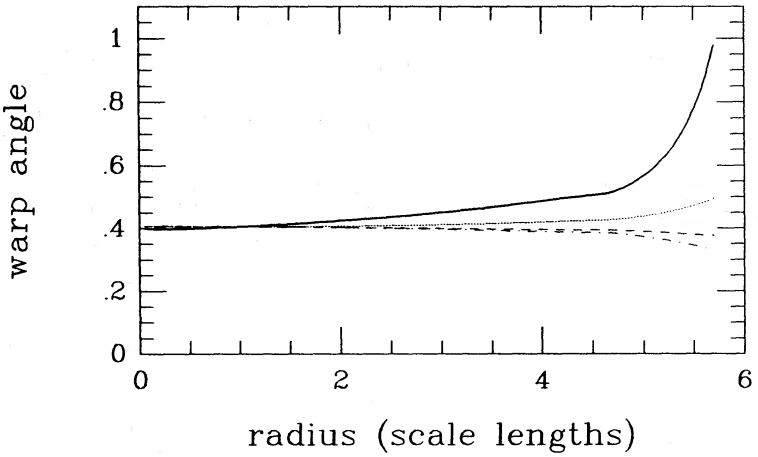

Figure 9. Modified tilt mode for various model disc-halo systems designed to represent the warped galaxy NGC 4565. For a given halo core radius, the halo mass is determined by fitting the rotation curve. Large core radii (8, 16 and 32 kpc; dotted, dashed, and dot-dashed line respectively) lead to very small bending $1-\chi$, so that no warp would be observed. A small core radius (4 kpc, solid line) yields a better warp model (*cf.* Fig. 10).

A selection of the model warps is shown in Fig. 9. With a large core radius, the modified tilt mode barely deviates from a rigid tilt, and little warping can be observed ($|1-\chi|<0.2$): rule **R2** is violated. A small core radius not only gives a sizeable warp amplitude, but, because the model falls close to the continuum boundary of Fig. 8, the model warp shows an abrupt change in curvature rather similar to that observed (see Fig. 10). In our picture, NGC 4565 presents a clear case of a Type I warp. However, only the general features of the warp can be reproduced; the detailed shape is very sensitive to small changes in the model parameters. We illustrate this by showing in Fig. 10 both the best model of Fig. 9, rescaled to match the observed warp amplitude (dotted), and a model in which the truncation in the disc has been moved inwards by 0.5 scale length (dashed). The latter would produce a significantly better match to the details of the warp shape, and still be compatible with the observations. We have not attempted to optimize the model parameters.

A peculiarity of the warp of NGC 4565 is that it appears to turn down at the very edge of the disc. This feature, apparent only on one side of the galaxy, cannot be reproduced by our models with constant halo ellipticity. However, the observed downturn involves only a very small amount of H I, and the dynamical significance of this whiff of gas is not obvious.

Notice that our best-fitting model, that with the smallest core radius, predicts that, if the gas disc were extended more than about 0.5 scale length further than its present observed outer radius, the warping mode would cease to be isolated (rule **R1** would be violated) and thus no warping would be observable. This conclusion is not very sensitive to the disc parameters.

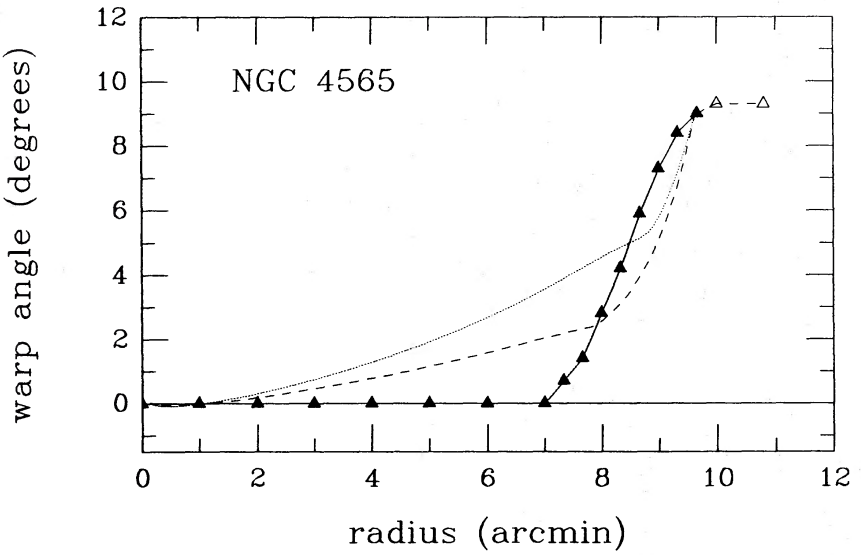

Figure 10. Observed and model warp of NGC 4565. The solid line and the triangles show the observed warp (Sancisi *et al.* in preparation; the last two points, shown open, are uncertain and have not been used in the fit). The dotted line is the best-fitting model of Fig. 9. The match is much improved if the edge is moved inward by 0.5 disc scale lengths (dashed line).

7.2 NGC 4013

NGC 4013 is a smaller galaxy, of Hubble type Sbc, with an optical scale length of 2.3 kpc (van der Kruit & Searle 1982) and a maximum rotation velocity of 195 km s^{-1} (Bottema *et al.* 1987). As is often the case for small galaxies, the H I disc extends well past the optical edge, out to about 9 scale lengths. The gas layer shows an obvious, smooth tilt from the inner to the outer regions, again starting near the edge of the stellar disc; the peak observed amplitude is about 30°. The shape of this warp is qualitatively different from that of NGC 4565.

In our models, the disc follows the exponential law of equation (6), with $r_{\rm trun}$=4.0 scale lengths and $r_{\rm out}$=4.25 scale lengths; the disc is extended to 8.7 scale lengths using test particles. We considered a range in halo core radius $r_{\rm c}$ from 4 to 24 kpc (1.75–10.5 scale lengths). For each value of $r_{\rm c}$, the halo mass was chosen by requiring the total rotation speed to drop by 15 per cent between a radius of 3 and 6 scale lengths, as indicated by the observations by Bottema *et al.* (1987). In the model with the smallest core radius (4 kpc), the modified tilt mode was not isolated; for larger core radii, the predicted warps are shown in Fig. 11.

Because the gas disc of this galaxy extends further than that of NGC 4565, configurations with a small core radius violate rule **R1**: the modified tilt mode is not isolated. Only a very narrow range of core radii results in an upward-tilting mode. Larger cores produce a downward-tilting mode with reasonably large amplitude ($|1-\chi|\sim0.7$). The model with $r_{\rm c}$=16 kpc (solid line in Fig. 11) shows a gradual tilt that matches well that observed in this galaxy. To illustrate the agreement between observed and model warp shape we have produced, with the kind assistance of R. Bottema, a synthetic map of the galaxy based on our best-fitting warp model; the agreement with the observed map is quite satisfactory (Fig. 12). The warp of NGC 4013 is a good Type II candidate.

In both these galaxies, independent attempts have been made at determining the core radius of the halo from the rotation curve alone. For NGC 4565, Sancisi *et al.* (in preparation) suggest that the halo core radius should be about 1 disc scale length; for NGC 4013, Bottema *et al.* (1987)

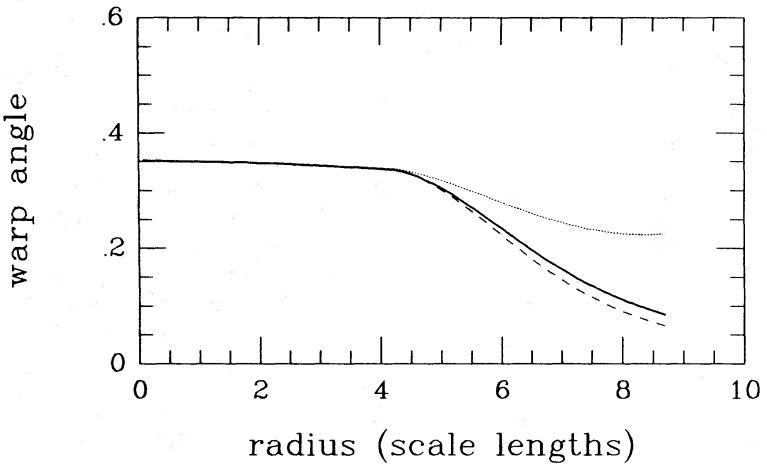

Figure 11. Modified tilt mode for various models for NGC 4013. Here, a small core radius is not acceptable: the mode would not be discrete. The models shown have r_c=3.5, 7 and 10.5 scale lengths and v_∞=0.65, 0.89 and 1.22 (dotted, solid and dashed line, respectively). The first shows too little warping. The other two models cannot be easily distinguished; the halo has nearly constant density in the observed region. The model shown with the solid line is in good agreement with the observations (see also Fig. 12).

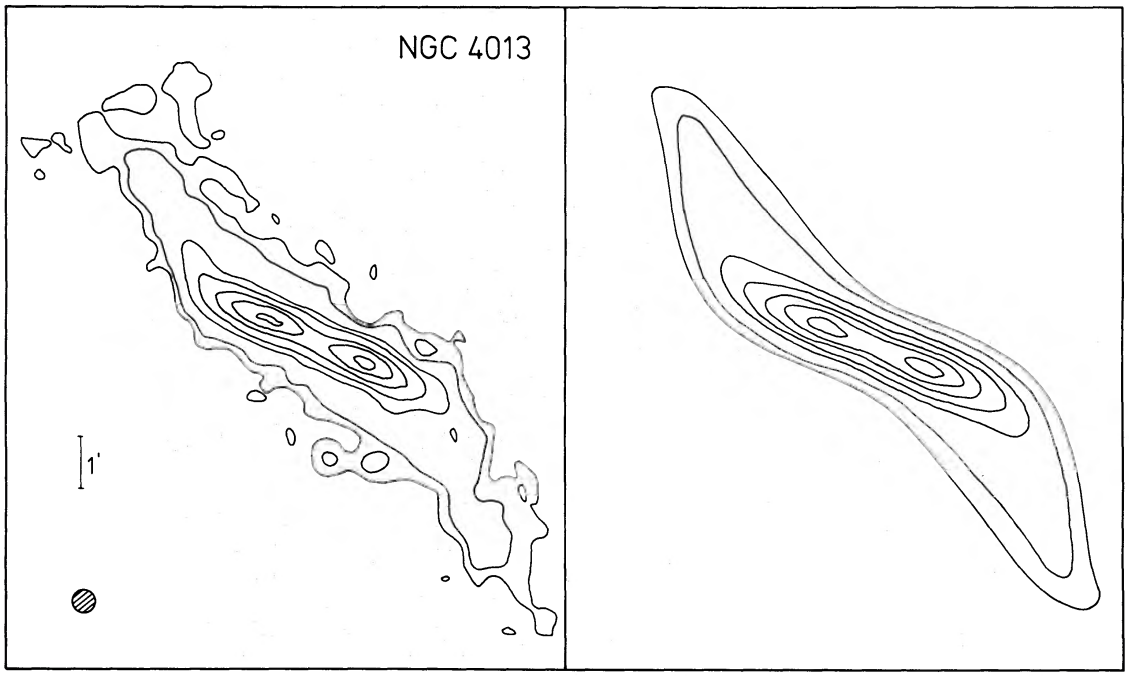

Figure 12. Comparison between observed and model warp in NGC 4013. The appearance of the H I in the plane of the sky (left) is well matched by the model image produced with our best-fitting warp (right).

adopt a large core radius, of about 6.5 disc scale lengths. Although the constraints are not very tight in either case, it is remarkable that these values are in good agreement with those derived from the best-fitting warp models.

8 Discussion

We have searched for discrete m=1 modes of vertical oscillation in a family of self-gravitating discs. A disc which has such a discrete mode can sustain a long-lived warp of the 'integral-sign'

shape found in many disc galaxies; the warp would have a straight line of nodes, with no sense of spirality. We have found two classes of these modes.

The first appears in isolated discs. If the density falls smoothly to zero (so that its derivative is continuous at the edge), then the system shows a continuum of bending modes, as predicted by Hunter & Toomre (1969). But in addition a number of discrete modes, which lead to long-lived warps, are present when the truncation takes place (smoothly) over a sufficiently narrow region. As this edge region is made smaller, more discrete modes appear; this can be understood using a WKB analysis.

It is not clear whether these discrete modes have astrophysical relevance. They arise only if the truncation region is very narrow, usually a few tenths of a scale length. It might be argued that the optical discs of some edge-on galaxies are in fact sharply truncated (van der Kruit & Searle 1981a, b); but the neutral gas in these galaxies continues significantly further out than the optical edge, and shows a significant warping (see Section 7). Even if its self-gravity is neglected, this gas can take part in the warp only if the massive disc exerts enough torque to make it precess at the same rate as the inner galaxy; that means that the gas layer cannot extend too far.

The optical disc of NGC 4565, for example, appears to end sharply at a radius of 4.5 scale lengths, and indeed a disc with $r_{\rm trun}=4.5$, $r_{\rm out}=4.7$ does have two discrete modes. But the H I gas extends to 5.7 scale lengths (Sancisi *et al.* in preparation), while the lowest bending mode R2 of our disc would persist only if the disc were extended with test particles to represent the gas out to about 4.9 scale lengths; beyond this radius, the outer material can no longer keep up with the precession of the inner warp. In conclusion, even though discrete modes may exist, galactic warps could only be explained as discrete bending modes of an isolated disc if the whole warped disc (stars *and* gas) was truncated abruptly, within a fraction of a scale length.

The second class of modes appears when the disc lies within an oblate distribution of halo material. In a spherical halo, the disc plane can take an arbitrary orientation; if the halo is flattened, this trivial 'tilt mode' may be modified to become a discrete warping mode. This mode does not depend on exactly how the disc is truncated, provided that it is not too extended nor the halo too flattened. The mode frequency can be found by a perturbation analysis.

We have thus given a quantitative basis to the suggestion (Dekel & Shlosman 1983, Toomre 1983) that flattened halo potentials can be responsible for many of the observed warps. For a reasonable oblateness (in most of our calculations, the halo density followed ellipsoids of constant ellipticity $\epsilon=0.2$), a discrete warping mode exists for values of the halo mass and core radius consistent with rotation curve analyses. The shape of the observable warp is most sensitive to the halo core radius; the mass and flattening of the halo are much less important.

If galactic warps represent modes of vertical oscillation, then the line of nodes of the warp must be straight; there must be no sense of spirality. This is a testable prediction. The warp in the Milky Way has a straight line of nodes (Henderson *et al.* 1982), although it is far from having the requested $m=1$ symmetry. The lines of nodes also appear to be nearly straight in the observed edge-on warps. Where warps are inferred by kinematic modelling, their spirality is much more difficult to check. We are not aware of any kinematic models accurate enough for a meaningful test of the hypothesis of a straight line of nodes.

Another directly testable prediction is that of the warp shape; we have successfully reproduced the main features of the warps observed in NGC 4565 and 4013, and derived estimates for the core radius of the halo mass distribution. In both cases our values for the core radius are close to previous estimates based on their rotation curves and luminosity distributions. Similar tests can be performed with observations of other warped systems; work on this is in progress. It is important to realize that some features, such as the exact shape of a Type I warp near the disc edge, are very sensitive to the details of the model; it is not desirable to fit such features. The predictive power of our model is in the correlation between the halo parameters and the main

properties of a warp, namely whether it is sharp or gentle and the radial range over which it extends.

The present study has addressed the question of whether a misalignment between disc and halo can produce a long-lived observable warp. Our predicted warps are independent of what process (disc angular momentum initially misaligned with the halo, gas accreted from out-of-plane orbits, tidal interaction) might have caused this tilt. The problem of the *generation* of warps remains open.

Should the interpretation of persistent galactic warps as the modified tilt mode of a disc within an oblate massive halo be confirmed by further work, observed warps could provide insight into the distribution of dark matter. The rotation curve of a galaxy can generally be fitted by a one-parameter sequence of disc-halo models; the mass and core radius cannot be found separately. If a good description of the warp shape and accurate rotation velocities are available, it should be possible to determine both these parameters in our model to within a factor of two. Even without detailed modelling, we expect that abrupt (Type I) warps will be associated with a small core radius in the dark component, and gradual (Type II) warps with a large core; a statistical study of warp shapes might then be used to constrain the core sizes.

The analysis of warps would then provide one of the very few available tests of conjectures about the distribution of 'missing mass' around galaxies

Acknowledgments

During the course of this work, we have benefited from frequent discussions with a number of colleagues, especially Tjeerd van Albada, Giuseppe Bertin and Renzo Sancisi. They and Bob Sanders have offered valuable advice on the presentation. Roelof Bottema has provided us with his observations of NGC4013 in advance of publication, and has helped us in producing the model of the warp in this galaxy shown in Fig. 12. Finally, we acknowledge an extremely useful exchange of ideas (and a careful scrutiny of an early draft) with Alar Toomre.

References

Avner, E. S. & King, I. R., 1967. *Astr. J.*, **72,** 750.
Baldwin, J. E., Lynden-Bell, D. & Sancisi, R., 1980. *Mon. Not. R. astr. Soc.*, **193,** 313.
Begeman, K., 1987. *PhD thesis,* Groningen University, The Netherlands.
Bertin, G. & Casertano, S., 1982. *Astr. Astrophys.*, **106,** 274.
Bertin, G. & Coppi, B., 1985. *Astrophys. J.*, **298,** 387.
Bertin, G. & Mark, J. W.-K., 1980. *Astr. Astrophys.*, **88,** 289.
Binney, J., 1978. *Mon. Not. R. astr. Soc.*, **183,** 779.
Binney, J., 1981. *Mon. Not. R. astr. Soc.*, **196,** 455.
Bosma, A., 1978. *PhD thesis,* Groningen University, The Netherlands.
Bottema, R., Shostak, G. S. & van der Kruit, P. C., 1987. *Nature,* **328,** 401.
Burbidge, E. M. & Burbidge, G. R., 1975. In: *Stars and Stellar Systems,* Vol. IX: *Galaxies and the Universe,* eds Sandage, A., Sandage, M. & Kristian, J., University of Chicago Press, Chicago, USA.
Casertano, S., 1983a. *Mon. Not. R. astr. Soc.*, **203,** 735.
Casertano, S., 1983b. *Thesis,* Scuola Normale Superiore, Pisa, Italy.
Dekel, A. & Shlosman, I., 1983. In: *Internal Kinematics and Dynamics of Galaxies, IAU Symp. No. 100,* ed Athanassoula, E., Reidel, Dordrecht, The Netherlands.
Faber, S. M. & Gallagher, J. S., 1979. *Ann. Rev. Astr. Astrophys.*, **17,** 135.
Henderson, A. P., Jackson, P. D. & Kerr, F. J., 1982. *Astrophys. J.*, **263,** 116.
Hunter, C., 1969a. *Studies Appl. Math.*, **48,** 55.
Hunter, C., 1969b. *Astrophys. J.*, **157,** 183.
Hunter, C. & Toomre, A., 1969. *Astrophys. J.*, **155,** 747.
Innanen, K. A., Kamper, K. W., Papp, K. A. & van den Bergh, S., 1982. *Astrophys. J.*, **254,** 515.
Kahn, F. D. & Woltjer, L., 1959. *Astrophys. J.*, **130,** 705.

Kerr, F. J., 1985. In: *The Milky Way Galaxy, IAU Symp. No. 106,* eds van Woerden, H., Allen, R. J. & Burton, W. B., Reidel, Dordrecht, The Netherlands (discussion after paper by L. S. Sparke, p. 501).
Kulkarni, S. R., Blitz, L. & Heiles, C., 1982. *Astrophys. J.*, **259,** L63.
Lynden-Bell, D., 1965. *Mon. Not. R. astr. Soc.*, **129,** 299.
Petrou, M., 1980. *Mon. Not. R. astr. Soc.*, **191,** 767.
Rogstad, D. H., Lockhart, I. A. & Wright, M. C. H., 1974. *Astrophys. J.*, **193,** 309.
Sancisi, R., 1976. *Astr. Astrophys.*, **53,** 159.
Sancisi, R., 1983. In: *Internal Kinematics and Dynamics of Galaxies, IAU Symp. No. 100,* ed. Athanassoula, E., Reidel, Dordrecht, The Netherlands.
Schmidt, M., 1956. *Bull. astr. Inst. Neth.*, **13,** 15.
Sparke, L. S., 1984a. *Astrophys. J.*, **280,** 117.
Sparke, L. S., 1984b. *Mon. Not. R. astr. Soc.*, **211,** 911.
Sparke, L. S., 1986. *Mon. Not. R. astr. Soc.*, **219,** 657.
Toomre, A., 1983. In: *Internal Kinematics and Dynamics of Galaxies, IAU Symp. No. 100,* ed Athanassoula, E., Reidel, Dordrecht, The Netherlands.
Tubbs, A. D. & Sanders, R. H., 1979. *Astrophys. J.*, **230,** 736.
van Albada, T. S. & Sancisi, R., 1986. *Phil. Trans. R. Soc. Lond.*, **A320,** 305.
van der Kruit, P. C., 1979. *Astr. Astrophys. Suppl.*, **38,** 15.
van der Kruit, P. C. & Searle, L., 1981a. *Astr. Astrophys.*, **95,** 105.
van der Kruit, P. C. & Searle, L., 1981b. *Astr. Astrophys.*, **95,** 116.
van der Kruit, P. C. & Searle, L., 1982. *Astr. Astrophys.*, **110,** 61.

Appendix: The regularized dynamical equation for a thin disc

Here we derive the Fourier–Bessel version of the dynamical equation for bending perturbations for a thin disc, which is essentially equivalent to that derived by Hunter & Toomre (1969) using associate Legendre functions. The advantages of Fourier–Bessel transforms are a more direct relationship with the non-regularized equations of Section 3 and the possibility of dealing with an infinite disc.

In the following we consider a cold, infinitely thin disc, with a warp described by the height $Z=Re\,[h(r,\varphi,t)]$ of the disc above the plane $z=0$. The treatment is linear in h, so we can extract a single Fourier component in time and azimuth; all quantities linear in h have the implicit dependence $\exp[i(\omega t-m\varphi)]$.

The equation we derive is 'regularized', in the sense that we use the 'small-slope' approximation of Hunter & Toomre (1969) to drop the most singular terms in the self-gravity of the disc. The small-slope approximation requires that both the warp height h and its derivative be small. Within this approximation, the vertical field $F_{\rm self}$ experienced by the bent disc can be decomposed into two terms: F_1, the field due to the perturbed disc, and F_2, the unperturbed field of the disc measured at the perturbed position. These two terms have the same meaning as in Hunter & Toomre (1969), equations (11) and (12).

Our dynamical equation (1) can be written:

$$\{[\omega-m\Omega(r)]^2-\mu^2_{\rm ext}(r)\}h(r)=-F_1[h]-F_2[h], \tag{A1}$$

where $\mu_{\rm ext}$ is the vertical frequency in the external potential, as in equation (2).

The term F_1 represents the change in the vertical force at the plane $z=0$ due to the perturbation in the disc. In the small-slope approximation, the density of the warped disc can be decomposed into the sum of the unperturbed disc and of a dipole distribution in the plane $Z=0$, with surface density

$$\nu(r)=h(r)\sigma(r). \tag{A2}$$

The potential $\Phi_1(r,z)$ generated by this dipole perturbation must satisfy the following conditions:

(i) For $z\neq 0$, it is a solution of the Laplace equation

$$\frac{\partial^2\Phi_1}{\partial r^2}+\frac{1}{r}\frac{\partial\Phi_1}{\partial r}+\frac{\partial^2\Phi_1}{\partial z^2}-\frac{m^2}{r_2}\Phi_1=0; \tag{A3}$$

(ii) It is odd in z $[\Phi_1(z)=-\Phi_1(-z)]$;
(iii) At the plane $z=0$ it is given by

$$\lim_{z\to 0\pm}\Phi_1(r,z)=\pm 2\pi G\not{\rho}(r); \tag{A4}$$

(iv) It must vanish for $z\to\infty$.

The explicit solution of equation (A3) for Φ_1 is immediate with the aid of Fourier–Bessel transforms. The perturbed potential Φ_1 can be expressed in terms of its radial transform $\tilde{\Phi}_1$ as:

$$\Phi_1(r,z)=\int_0^\infty \tilde{\Phi}_1\ (k,z)J_m(kr)k\,dk, \tag{A5}$$

where the definition of the transformed quantity is:

$$\tilde{\Phi}_1\ (k,z)=\int_0^\infty \Phi_1(r',z)J_m(kr')r'\,dr'. \tag{A6}$$

The z-dependence of the potential is found by substituting the transformed potential (A5) into the Laplace equation (A3), with the boundary condition (iv):

$$\Phi_1(r,z)=\mathrm{sgn}(z)\int_0^\infty \tilde{\Phi}_1\ (k,0)J_m(kr)\exp(-k|z|)\,k\,dk. \tag{A7}$$

This satisfies conditions (i), (ii) and (iv). Applying the definition (A6) of the inverse transform and the condition (A4) for the value of the potential at the plane $z=0$, we find that:

$$\Phi_1(r,z)=2\pi G\,\mathrm{sgn}(z)\int_0^\infty J_m(kr)\exp(-k|z|)\,k\,dk\int_0^\infty J_m(kr')\not{\rho}(r')r'\,dr'. \tag{A8}$$

This satisfies conditions (i)–(iv) and is thus the required solution. The force F_1 can now be computed:

$$F_1=-\left.\frac{\partial\Phi_1}{\partial z}\right|_{z=0\pm}=\lim_{z\to 0\pm}2\pi G\int_0^\infty J_m(kr)\exp(-k|z|)\,k\,dk\int_0^\infty J_m(kr')\not{\rho}(r')r'\,dr'. \tag{A9}$$

For $z\neq 0$ we can reverse the order of integration to obtain:

$$F_1=G\int_0^\infty \sigma(r')h(r')B_m(r,r')r'\,dr' \tag{A10}$$

where we have substituted the definition (A2) of the dipole density $\not{\rho}$ and the kernel $B_m(r,r')$ is defined as

$$B_m(r,r')=\lim_{z\to 0\pm}\int_0^\infty J_m(kr)J_m(kr')\exp(-k|z|)k^2\,dk. \tag{A11}$$

The term F_2 is the vertical component of the unperturbed field at the perturbed position of the disc. However, a distinction is necessary. The unperturbed vertical force due to the flat disc can be decomposed into two parts: a portion linearly increasing near the plane $z=0$, due to distant parts of the disc, and a portion which is constant near the surface of the disc and discontinuous

across $z=0$, due to the local matter. Only the first term is required here, since the local matter has now moved along with the perturbed disc. We must then subtract the second term away from the vertical field. If $\Phi_0(r, z)$ is the gravitational potential of the unperturbed disc, the *total* vertical field F_z at the location $z=Z\equiv Re(h)$ is

$$F_z(r, Z)=-\left.\frac{\partial\Phi_0}{\partial z}\right|_{z=Z}. \tag{A12}$$

The contribution of the local material is $\lim_{z\to 0\pm} F_z(r, z)\equiv\mp 2\pi G\sigma(r)$. The vertical force F_2 actually experienced by the disc is obtained by subtracting out this 'self-force' contribution:

$$F_2(z, Z)=F_z(r, Z)-\lim_{z\to 0\pm} F_z(r, z), \tag{A13}$$

where the plus sign is chosen if $Z>0$, the minus otherwise. In the linear limit for Z (or h), equation (A13) is equivalent to

$$F_2[h]=-h\left.\frac{\partial^2\Phi_0}{\partial z^2}\right|_{z=0\pm}; \tag{A14}$$

here either sign can be chosen, since the second derivative of the potential is continuous across the disc (although it is not defined at $z=0$).

The potential $\Phi_0(r, z)$ can be obtained by a slight adaptation of the treatment in Casertano (1983a) in terms of an integral involving the Bessel function J_0:

$$\Phi_0(r, z)=-2\pi G\int_0^\infty J_0(kr)\exp(-k|z|)\,dk\int_0^\infty \sigma(r')J_0(kr')r'\,dr'. \tag{A15}$$

Differentiating twice with respect to z we obtain for F_2:

$$F_2=2\pi Gh(r)\lim_{z\to 0\pm}\int_0^\infty J_0(kr)\exp(-k|z|)\,k^2\,dk\int_0^\infty \sigma(r')J_0(kr')r'\,dr'. \tag{A16}$$

Since we only evaluate the double integral for $z\neq 0$, the order of integration can be reversed to obtain:

$$F_2=Gh(r)\int_0^\infty \sigma(r')B_0(r, r')r'\,dr', \tag{A17}$$

where the kernel B_0 is defined as in equation (A11), with $m=0$.

Combining equations (A1), (A10) and (A17) we obtain the regularized form of the dynamical equation:

$$\{[\omega-m\Omega(r)]^2-\mu_{\rm ext}^2(r)\}h(r)=Gh(r)\int_0^\infty \sigma(r')B_0(r, r')r'\,dr'$$

$$-G\int_0^\infty h(r')\sigma(r')B_m(r, r')r'\,dr'. \tag{A18}$$