1997

Dynamics of an Interacting Luminous Disk, Dark Halo, and Satellite Companion

Martin D. Weinberg
University of Massachusetts - Amherst

Dynamics of an interacting luminous disk, dark halo, and satellite companion

Martin D. Weinberg[*]

Department of Physics and Astronomy, University of Massachusetts, Amherst, MA 01003-4525, USA

ABSTRACT

This paper describes a method for determining the dynamical interaction between extended halo and spheroid components and an environmental disturbance. One finds that resonant interaction between a galaxy and passing interlopers or satellite companions can carry the disturbance inward, deep inside the halo, where it can perturb the disk.

Applied to the Milky Way for example, the LMC and SMC appear to be sufficient to cause the observed Galactic warp and possibly seed other asymmetries. This is a multi-scale interaction in which the halo wake has a feature at roughly half the satellite orbital radius due to a 2:1 orbital resonance. The rotating disturbance then excites an $m = 1$ vertical disk mode which has the classic integral-sign morphology. A polar satellite orbit produces the largest warp and therefore the inferred LMC orbit is nearly optimal for maximum warp production.

Both the magnitude and morphology of the response depend on the details of the disk and halo models. Most critically, a change in the halo profile will shift the resonant frequencies and response location and consequently alter the coupling to the bending disk. Increasing the halo support relative to the disk, a sub-maximal disk model, decreases the warp amplitude.

Finally, the results and prognosis for N-body simulations are discussed. Discreteness noise in the halo, similar to that due to a population of $10^6 M_\odot$ black holes, can produce observable warping. Details are discussed in an associated paper.

Key words: Galaxy: halo, structure—galaxies: halos, kinematics and dynamics—Magellanic Clouds

1 INTRODUCTION

Even casual examination shows that most disk galaxies are not truly symmetric but exhibit a variety of morphological peculiarities of which spiral arms and bars are the most pronounced. After decades of effort, we know that these features may be driven by environmental disturbance acting directly on the disk, in addition to self-excitation of a local disturbance (e.g. by swing amplification, Toomre 1981, Sellwood & Carlberg 1984). However, all disks are embedded within halos and therefore are not dynamically independent and will respond to asymmetries and distortions in the halo, as well.

Until recently, conventional wisdom was that halos acted to stabilize disks but otherwise remained relatively inert. The argument behind this assumption is as follows. Halos, spheroids and bulges are supported against their own gravity by the random motion of their stars—a so-called "hot" distribution (e.g. Binney & Tremaine 1987). On all but the largest scales, they look like nearly homogeneous thermal baths of stars. Because all self-sustaining patterns or waves in a *homogeneous* universe of stars with a Maxwellian velocity distribution are predicted to damp quickly (e.g. Ikeuchi, Nakamura & Takahara 1974), one expects that any pattern will be strongly damped in halos and spheroids as well. However, recent work suggests that halos *do* respond to tidal encounters by companions or cluster members and *are* susceptible to induction of long-lived modes due to their inhomogeneity. These modes are at the largest scales for which self gravity is most effective. In particular, if halos are large as many currently estimate, halo-halo interactions in groups or clusters will be frequent and much more common than disk-disk interactions. Because non-local coupling can carry a disturbance to small radii, transient halo structure may be sufficient to trigger disk structure even if the halo pat-

tern subsequently damps away. The non-locality of the response was predicted by Weinberg (1986, 1989) and verified in n-body simulations by Hernquist & Weinberg (1989), and Prugniel & Combes (1992). Application to self-gravitating fluctuations is described in an associated paper (Weinberg 1997).

Similarly, we expect that a companion can continuously re-excite structure. The orbit of the companion decays by dynamical friction. The response of the dark halo to the interloping satellite can be thought of as a gravitational wake (e.g. Mulder 1983); since the wake trails and has mass, it exerts a backward pull on the satellite. This view of dynamical friction reproduces the standard approach but shows that the halo responds with structure whose mass comparable to the satellite. This general approach has been tested and compared with n-body simulations in a variety of contexts with good agreement. Recent work (Weinberg 1995b, Paper I) suggests that the Magellanic Clouds use this mechanism to produce distortions in the Galactic disk sufficient to account for both the radial location, position angle and sign of the HI warp and observed anomalies in stellar kinematics. In clusters, this mechanism is mostly likely the culprit behind *galaxy harassment* (Moore et al. 1996).

Here, we develop this suggestion and present a formalism for exploring the dynamics of low-amplitude interactions that can lead to significant long-term evolution. As an example throughout, I will focus on disk warping and on the Milky Way—LMC interaction presented in Paper I. Although there are some general trends, a relatively large amplitude disk response tends to be the result of a conspiracy between frequencies. Rather than present an exhaustive parameter survey, we explore some simple scenarios. Even though the warp amplitudes vary with the details of the galaxy profiles, astronomically interesting amplitudes are generally produced. Sections 2 and 3 describe the galaxian models and the method. We will use a numerical perturbation theory which is well suited to describing weak coherent perturbations. The main results are in §4 which begins by examining an example distortion of the halo by a companion satellite and traces its influence on both the disk warp and in-plane disk distortions. Readers may skip the technical detail without loss of continuity by turning to §4 after the introduction to §sec:method. The results section is followed by a discussion of the range of effects using other models and rough generalizations (§5). It is tempting and desirable to follow this up with n-body simulation but simulations of weak distortions with large dynamic range is a significant challenge and will require very large particle numbers to recover signal (§6). We end with a summary and outline for future work (§7).

2 GALAXY MODELS

In order to explore interactions between components, we need self-consistent multi-component galaxy models. Fully self-consistent disk-in-halo models are generally made prescriptively rather than constructively because of the difficulty in determining distribution functions. For example, a regular disk profile embedded in an a regular spherical halo will generally not yield a self-consistent regular system demanding computationally intensive techniques such as Schwarzschild's method (1979). Most often in the literature, the Jeans' moment equations are used to construct an n-body disk in a given halo or spheroid and the resulting distribution is allowed to phase mix to equilibrium. In this paper, we adopt a hybrid approach suited to both the analytic perturbation theory described in §3 and the n-body simulations described in §6. The prescription is as follows:

(i) Choose a spherical halo and two- or three-dimensional disk profile.

(ii) Assume that the disk does not affect the halo profile and construct a disk phase-space distribution function for the disk in the halo. This would rigorously obtain for $M_{halo} \gg M_{disk}$. The disk distribution function is computed by a quadratic programming scheme similar to that discussed by Dejonghe (1989, see Appendix A). The term halo here includes the entire *hot* component: bulge, stellar spheroid and dark matter. The halo distribution function is assumed to be known. If it is not, it may be constructed using any of the established techniques (e.g. Eddington inversion, generalized integral inversion, atlas method, etc.), ignoring the disk.

(iii) The approximate distribution functions for each component are now known. For an n-body simulation, these may be directly realized by a Monte Carlo procedure.

The realized phase space distribution will not be in strict equilibrium. However, as long as the force of the halo dominates the disk at large radii, and the disk dominates its own gravity inside its scale length for realistic parameters, the initial construction is mostly likely close to equilibrium. Numerical experiments support this conjecture. Nevertheless, a mild inconsistency is of minor consequence for the numerical method presented in §3 (see discussion in §3.2).

For analytic convenience, we use King models of varying concentration and scale to represent the dark halo, a Hernquist model to represent the bulge/spheroid, and Hunter (1963) polynomial disk profiles. The latter choice is motivated by the numerical analysis required to compute the vertical disk response following Hunter & Toomre (1969, hereafter HT). Hunter's polynomial disk models are less centrally concentrated and fall of more steeply then the standard exponential disk. Hunter & Toomre modified these disks to better fit observed profiles by subtracting off a low-order contribution; they denoted these models with the suffix "X". The Hunter-Toomre 16X model is fair approximation to the exponential disk with a scale length of $R_{max}/6$; this corresponds to $a = 3.5, 4.5$ kpc for $R_{max} = 21, 28$ kpc. These profiles are compared in Figure 1. For the n-body tests described in §6, we used a rigid bulge component to stabilize the inner disk (which is otherwise bar unstable).

The extent of Milky Way disk remains under debate. Robin, Crézé & Mohan (1993) present evidence for an edge to the stellar disk at 14 kpc. This is consistent with estimates based on molecular tracers (Wouterloot et al. 1990, Digel 1991, Heyer et al. 1997) which imply an edge of about 14 kpc or somewhat greater. The outer atomic gas matches on to the inner component continues to 30 kpc at significant surface densities (Kulkarni, Heiles & Blitz 1982). A more recent analysis (Diplas & Savage 1991) reports an exponentially distributed neutral hydrogen disk out to *at least* 30 kpc. We will adopt the $R_{max} = 28$ kpc model because it better represents the global extent of the disk even
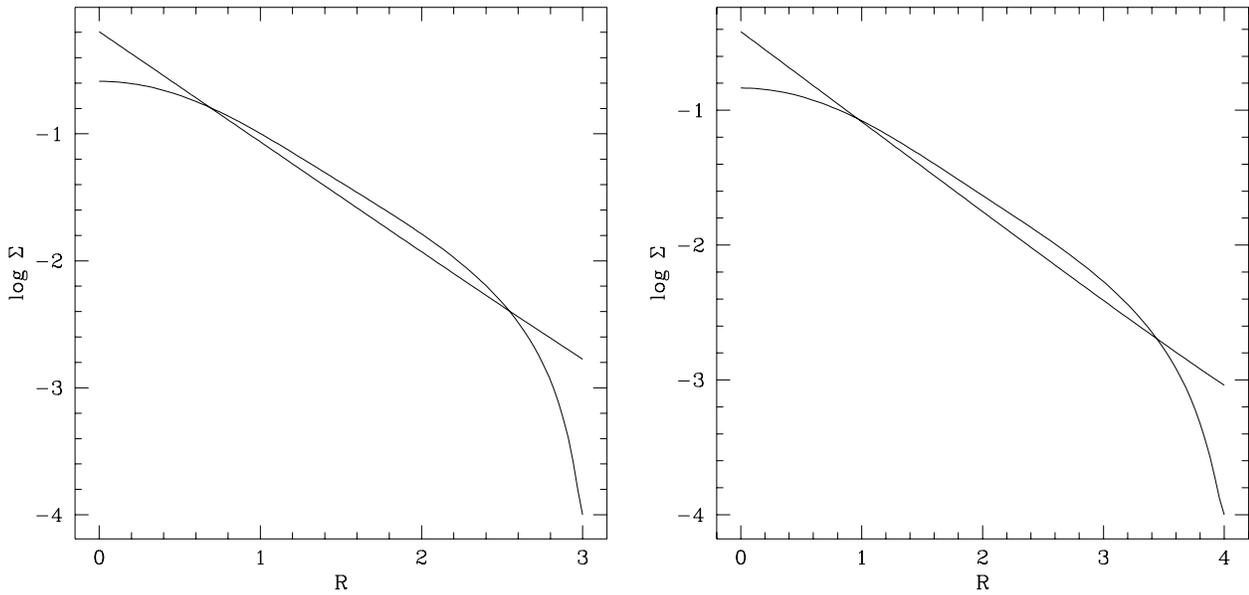
Figure 1. Left-hand (right-hand) panel compares the exponential disk to the Hunter-Toomre 16X model with identical mass inside of $R = 3$ ($R = 4$) corresponding to approximately 21 kpc (28 kpc).

though the scale length required is a bit large. In calculations below, the disk model is assigned $M = 1$ and this corresponds to the estimated 6.0×10^{10} M$_\odot$ disk (Binney & Tremaine 1987). The radial scaling is one unit for each 7 kpc. The satellite orbit is defined by its energy and angular momentum in the halo model. For each halo model, the orbit is assigned a pericenter at 50 kpc and apocenter at 100 kpc. The mass of the LMC in these units is 0.25 (0.1) for $M_{LMC} = 1.5 \times 10^{10}$ M$_\odot$ (6×10^9 M$_\odot$) (Meatheringham 1988, Schommer et al. 1992). For this study, we lump the Small Cloud together with the LMC and do not consider the possibility of a distinct SMC orbit. In all that follows, we will scale results to the Milky Way.

Given the disk profile, the King radius and mass are constrained by the demand for a flat rotation curve, at least for $R \lesssim 50$ kpc. The choice of a $W_0 = 3$ model with truncation radius at 200 kpc and mass of 10 disk masses results in a Milky Way mass of 3.3×10^{11} inside of 50 kpc. This is a bit smaller than but comparable to current values; e.g. Kochanek (1996) estimates 4.9×10^{11} M$_\odot$. We also will consider halo mass ratios of 15 and 20 disk masses. The rotation curves remain approximately flat for both cases and give masses within 50 kpc of 4.7×10^{11} M$_\odot$ and 6.0×10^{11} M$_\odot$, respectively. Finally, the timing argument and its recent generalizations (Peebles 1995 and references therein) estimate a Milky Way mass of 2×10^{12} M$_\odot$. This is a *constraint* not a target for our models since one can add mass beyond 100 kpc without affecting any of the dynamical arguments considered here. Our most massive model is below this limit with a mass of 1.3×10^{12} M$_\odot$.

The perturbing satellite orbit is chosen to match the LMC orbit inferred by Lin et al. (1995) and assigned an orbit consistent with the chosen halo profile and scaling given in the previous section. The clouds are assumed to be near pericenter with $R_{LMC} = 49.5$ kpc. The space velocity

inferred from proper motion, radial velocity measurement, and a distance estimate immediately determines the orbital plane as described in Paper I. In a Galactocentric coordinate system with the LSR along the x-axis ($x = -8.5$ kpc) and moving towards positive y, one may straightforwardly derive the following instantaneous orbit: $-76 \pm 13°$ inclination, $-82 \pm 10°$ longitude of ascending node, $-36 \pm 3°$ argument of perigalacticon. However, considered in these Galactic coordinates, the Milky Way disk rotates clockwise and has negative angular momentum. In the development below, disks are assigned positive angular momentum with counterclockwise rotation. Our models may be transformed to the Milky Way system by reflection through the y–z plane.

We will adopt a standard galaxy model with a King model halo with tidal radius $R_t = 28$ (200 kpc), $\log c = 0.67$, and a mass of 10 times the disk mass; this is "maximal" disk model. The bulge is Hernquist model with scale length 0.2 (1.4 kpc) and mass of 20% of the disk (1.2×10^{10} M$_\odot$). The disk is the Hunter-Toomre 16X model with $R_{max} = 4$ (28 kpc). In this system, one velocity unit is 350 km s^{-1}. The rotation curve for this model rises from 0.6 at $R = 0.5$ to 0.7 at $R = 1.8$, drops slowly to 0.6 at $R = 10$ (70 kpc) and drops off more rapidly beyond this point. Although better fits to the observed Milky Way rotation curve are available, our goal of understanding the underlying mechanism and the computational simplicity of these components supports our choice. Finally, the standard model includes a satellite with an LMC orbit. The magnitude of the response scales with satellite mass and need not be chosen a priori.

3 A FORMALISM FOR MULTI-SCALE INTERACTIONS

A full treatment requires the dynamical coupling of the multiple time scales and multiple length scales of the external

disturbance and the galaxian components discussed above. Relevant characteristic length and time scales may differ by an order of magnitude between satellite and halo or disk orbits. In addition, we will see that the halo disturbance may be relatively weak and a significant perturbation of the outer disk at the same time. These multiple-scale weak regimes are a challenging task for an n-body computation. However, this class of problems is ideally suited to linear techniques and the work here will use the expansion technique known as the *matrix method*. Although the matrix method is computationally intensive, it is no more so than n-body methods and is practical on current workstations. In this section, I will give a brief overview of the general method with details on posing and implementing the coupled response solutions in the references cited below and in the Appendix.

In short, the matrix method represents the response of a galaxy to an external perturbation by a truncated series of orthogonal functions, similar to those one would use to solve an electrostatics problem. The perturbation is also represented by this series and the temporal dependence of each coefficient is Fourier transformed to a (complex) frequency distribution. The response of the galaxy to one of the orthogonal functions at a particular forcing frequency is then computed in the continuum limit using the collisionless Boltzmann (Vlasov) equation. The entire procedure is analogous to signal processing in Fourier space. Pursuing the analogy, we now do the inverse transform. The response to any perturbation, the weighted superposition of the response to each basis function, is then a matrix equation. Finally, to get the full time dependence of the response, one resums the solutions to the matrix equation at each frequency weighted by the Fourier coefficients.

This method assumes that the perturbation is small enough that the overall change to the structure of the galaxy is small. In this limit, the method has the advantage of accuracy and sensitivity to the large scale structures of interest. For contrast, the n-body simulation determines the response of a galaxy to a perturbation by solving the equations of motion for a representative set of orbits. The orbit is advanced in a fixed potential for a short time interval and the gravitational potential or force is then recomputed. The simulation works well for large perturbations but because the simulation uses a finite number of particles, fluctuation noise limits the sensitivity to small amplitude distortions. The matrix method nicely complements the n-body simulations, excelling in the regimes where the n-body simulation are suspect.

Historically, the approach is related to the treatment of general eigenvalue problems described in the mathematical physics literature (e.g. Courant & Hilbert 1953, Chap V). The matrix method in stellar dynamics had varied applications beginning with Kalnajs (1977) who investigated the unstable modes of stellar disks. Polyachenko & Shukhman (1981) adapted the method to study a spherical system (see also Fridman & Polyachenko 1984, Appendix) and it was later employed by both Palmer & Papaloizou (1987) in the study of the radial orbit instability and by Bertin & Pegoraro (1989) to study the instability of a family of models proposed by Bertin & Stiavelli (1984). In addition to Paper I, Weinberg (1989, Paper II) used the matrix formulation to study the response of a spherical galaxy to an encounter with

a dwarf companion and Saha (1991) and Weinberg (1991) investigated the stability of anisotropic galaxian models.

3.1 Mathematical overview

The response of a galaxy initially in equilibrium to a gravitational interaction with a companion is described by the simultaneous solution of the Boltzmann and Poisson equations. The simultaneous system is a set of coupled partial integro-differential equations. However if the orbits in each component are regular, any phase-space quantity—such as density and gravitational potential—may be expanded in a Fourier series in the orbital frequencies. Truncating this expansion, the quantity may be represented as a vector of Fourier coefficients; this is standard practice in filtering and approximation theory and canonical perturbation theory (e.g. Lichtenberg & Lieberman 1983). In Fourier space, the Boltzmann PDE becomes an algebraic integral equation. The system is further simplified if the basis functions are chosen to satisfy the Poisson equation explicitly. After a Laplace transform in time, the remaining solution of the Boltzmann equation becomes the solution of a matrix equation, each column describing the response to a particular basis function. See references cited abouve for mathematical detail.

To give the flavor of its use, consider two interacting galaxies. Denote the expansion of the perturbation potential caused by the companion galaxy as vector $\mathbf{b}$. Then the direct response of the galaxy, vector $\mathbf{a}$ may be written:

$$\mathbf{a} = \mathbf{R} \cdot \mathbf{b}. \tag{1}$$

The matrix $\mathbf{R}$, the *response operator*, implicitly contains the time-dependence of the perturbation as well as the dynamics of the Boltzmann equation. If we are interested in the self-gravitating response to the perturbation, we need to solve:

$$\mathbf{a} = \mathbf{R} \cdot (\mathbf{a} + \mathbf{b}). \tag{2}$$

In words, equation (2) states that the self-consistent reaction of a galaxy to a perturbation is the gravitational response to both the perturbing force and the force of the response itself. Equations (1) and (2) result from the Laplace transform of the Boltzmann equation and therefore represent a particular frequency component, $\mathbf{a} = \mathbf{a}(s)$. Therefore, the solution of equations (1) or (2) requires an inverse Laplace transform to recover its explicit time dependence (see Paper II for details). See Nelson & Tremaine (1997) for a general discussion of the response operator.

3.2 Combining multiple components

This approach is easily extended to find the simultaneous self-consistent response of several galaxian components. For example, let us consider a halo and a disk; any number of components may be combined similarly. Each component's distribution function solves a Boltzmann equation and is coupled to the others only through the total gravitational potential. If the interaction between the components is artificially suppressed, the simultaneous solution is that of the following augmented matrix equation:

$$\begin{pmatrix} \mathbf{a}_h \\ \mathbf{a}_d \end{pmatrix} = \left(\begin{array}{c|c} \mathbf{R}_h & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{R}_d \end{array} \right) \cdot \left[\begin{pmatrix} \mathbf{a}_h \\ \mathbf{a}_d \end{pmatrix} + \begin{pmatrix} \mathbf{b}_h \\ \mathbf{b}_d \end{pmatrix} \right], \tag{3}$$

where the subscripts h and d stand for the halo and disk, $\mathbf{b}$ is the external perturbation, and $\mathbf{0}$ is defined to be the matrix with the same rank as $\mathbf{R}$ and all elements zero. Equation (3) is simply two stacked versions of equation (2). The off-diagonal partitions in the augmented matrix, now $\mathbf{0}$, describe the mutual interaction between components.

To allow the components to interact, we project the halo response $\mathbf{a}_h$ onto the disk basis and then add this to the right-hand-side of the disk response equation and vice versa for the response of the halo to the disk. Letting the matrices that perform these projections be $\mathbf{P}_{hd}$ and $\mathbf{P}_{dh}$, we may write the fully coupled set as

$$
\begin{pmatrix} \mathbf{a}_h \\ \mathbf{a}_d \end{pmatrix} = \left(\begin{array}{c|c} \mathbf{R}_h & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{R}_d \end{array} \right) \cdot \left(\begin{array}{c|c} \mathbf{1} & \mathbf{P}_{hd} \\ \hline \mathbf{P}_{dh} & \mathbf{1} \end{array} \right) \cdot \begin{pmatrix} \mathbf{a}_h \\ \mathbf{a}_d \end{pmatrix} +
$$
$$
\left(\begin{array}{c|c} \mathbf{R}_h & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{R}_d \end{array} \right) \cdot \begin{pmatrix} \mathbf{b}_h \\ \mathbf{b}_d \end{pmatrix},
$$
$$
= \left(\begin{array}{c|c} \mathbf{R}_h & \mathbf{R}_h \mathbf{P}_{hd} \\ \hline \mathbf{R}_d \mathbf{P}_{dh} & \mathbf{R}_d \end{array} \right) \cdot \begin{pmatrix} \mathbf{a}_h \\ \mathbf{a}_d \end{pmatrix} +
$$
$$
\left(\begin{array}{c|c} \mathbf{R}_h & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{R}_d \end{array} \right) \cdot \begin{pmatrix} \mathbf{b}_h \\ \mathbf{b}_d \end{pmatrix}. \tag{4}
$$

The first term on the right-hand-side describes the mutual self-gravitating response due to each component and the second term describes the response of each component to the external perturbation.

We may straightforwardly isolate effects of interest by coupling or uncoupling components or by including or suppressing self-gravity. For example, we may consider the response of the disk to a halo wake, but without the back reaction of the halo to the disk by setting the upper right term in the augmented response matrix to zero. Or, if we want to limit the disk response to forcing by the halo alone, the direct response of the disk to the perturbation may be left out by setting $\mathbf{b}_d$ to zero, yielding

$$
\begin{pmatrix} \mathbf{a}_h \\ \mathbf{a}_d \end{pmatrix} = \left(\begin{array}{c|c} \mathbf{R}_h & \mathbf{0} \\ \hline \mathbf{R}_d \mathbf{P}_{dh} & \mathbf{R}_d \end{array} \right) \cdot \begin{pmatrix} \mathbf{a}_h \\ \mathbf{a}_d \end{pmatrix} + \left(\begin{array}{c|c} \mathbf{R}_h & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{R}_d \end{array} \right) \cdot \begin{pmatrix} \mathbf{b}_h \\ \mathbf{0} \end{pmatrix}. \tag{5}
$$

This is equivalent to first solving $\mathbf{a}_h = \mathbf{R}_h \cdot (\mathbf{a}_h + \mathbf{b}_h)$ and then $\mathbf{a}_d = \mathbf{R}_d \cdot (\mathbf{a}_d + \mathbf{P}_{dh} \mathbf{a}_h)$. This formalism couples the components by their perturbation from equilibrium, not the fully gravitational attraction. For this reason, any mild deviation from perfect self consistency produced by the prescription described in §2 does not cause any problems and is unlikely to be significant. Most of the computational work is in producing the matrices $\mathbf{R}$. Afterward, all of the variants may be studied with little additional effort.

3.3 Satellite perturbation

In order to apply the technique described in §3.2 to a perturbation by an orbiting satellite, we need to expand its potential in the chosen basis to get the perturbation vector $\mathbf{b}$. This expansion is described in §3.3.1 and applied in §3.3.2. Perturbation by an interloping galaxy may be treated similarly but is not described here.

3.3.1 Fourier expansion

Our complete set of basis functions are pairs of functions, (p_i, d_i), which solve Poisson's equation, $\nabla^2 d_i = 4\pi G p_i$, and are biorthogonal:

$$
\frac{1}{4\pi G} \int d\mathbf{r} \, p_i^*(\mathbf{r}) d_j(\mathbf{r}) = \delta_{ij}. \tag{6}
$$

The potential for the arbitrary point mass may be expanded directly in a biorthogonal harmonic series:

$$
\Phi(\mathbf{r}) = \sum_{lm} Y_{lm}(\theta, \phi) \sum_i b_k^{lm}(t) p_i^{lm}(r) \tag{7}
$$

where

$$
\begin{aligned}
b_i^{lm}(t) &= \int d^3 r \, Y_{lm}^*(\theta, \phi) p_i^{lm\,*}(r) \delta^3 \left(\mathbf{r} - \mathbf{r}(t) \right) \\
&= Y_{lm}^*(\theta(t), \phi(t)) p_i^{lm\,*}(r(t)) \tag{8}
\end{aligned}
$$

where $\mathbf{r(t)}$ describes the orbit of the satellite. Alternatively, since b_i^{lm} is an implicit function of time through $\mathbf{r}$, we may expand in an action-angle series which makes the time dependence explicit. This gives

$$
b_i^{lm}(t) = \sum_{l_1, l_2 = -\infty}^{\infty} e^{im\gamma} e^{il_2\alpha} Y_{lm}(\pi/2, 0) W_{l\,l_2\,m}^{l_1\,i\,*} e^{i(l_1\Omega_1 + l_2\Omega_2)t} \tag{9}
$$

where the coefficient $W_{l\,l_2\,m}^{l_1\,i\,*}$ depends only on the energy and angular momentum of the satellite orbit. Derivation of equation (9) is given in the Appendix.

The perturbation vectors b in equation (2) are given by the Laplace transform of $b_i^{lm}(t)$ from equation (7) for each term in equation (9). The Laplace transform of $b_i^{lm}(t)$ in this form is trivial. Because the physical measurable must be real, we can simplify the analytic computation for $m \neq 0$ by using only $m > 0$ terms and adding the complex conjugate in the end.

3.3.2 Application to response calculation

Finally, recovery of the response, $\mathbf{a}$, in the time domain requires the inverse Laplace transform of the following solution for $\mathbf{a}$ from equation (4):

$$
\begin{pmatrix} \mathbf{a}_h \\ \mathbf{a}_d \end{pmatrix} = \left[\left(\begin{array}{c|c} \mathbf{1} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{1} \end{array} \right) - \left(\begin{array}{c|c} \mathbf{R}_h & \mathbf{R}_h \mathbf{P}_{hd} \\ \hline \mathbf{R}_d \mathbf{P}_{dh} & \mathbf{R}_d \end{array} \right) \right]^{-1} \cdot
$$
$$
\left(\begin{array}{c|c} \mathbf{R}_h & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{R}_d \end{array} \right) \cdot \begin{pmatrix} \mathbf{b}_h \\ \mathbf{b}_d \end{pmatrix}
$$

or, in compact form,
$$
\hat{\mathbf{a}}^{lm}(s) \equiv \mathcal{D}^{-1}(s) \cdot \mathcal{R}(s) \cdot \hat{\mathbf{b}}^{lm}(s). \tag{10}
$$

The matrix in large square brackets in this equation is the dispersion relation; its determinant vanishes at eigenmodes. The assumption that our multi-component galaxy is stable ensures that the inverse of this matrix has no poles in the complex half plane with $\Re(s) > 0$; poles on the half plane with $\Re(s) < 0$ correspond to damped modes. Elements of the second term matrix have at least one pole on the imaginary s-axis due to the harmonic forcing by the satellite. In the end, the inverse transform may be simply evaluated by deforming the integration contour through the imaginary axis and taking the time-asymptotic limit (see §C for details). After many satellite orbits, the harmonic-forcing contribution dominates all but a very weakly damped mode.

3.4 Disk bending

A differential vertical force applied by the external perturbation and the halo wake can warp the disk plane. HT describe a linearized solution for the dynamical evolution of the plane for an isolated disk. The bending analysis here uses the formalism developed by HT with several modifications. First, because our equilibrium disk models are embedded in an external halo, we must retain their equation (12) rather than simplify using relationships based on the specific form of the background model. Second, we solve the linearized equations of motion under a forced disturbance (their eq. 19) by Laplace transform for consistency with the approach in §3.2. The vertical force follows directly from the expansion coefficients describing the external perturbation, equation (9) and the halo wake, equation (10). The back-reaction to the in-plane distortion is included and is a relatively minor contribution to the total response. This calculation does not consider the back-reaction of the halo to the vertical distortion (Nelson & Tremaine 1995).

For reasons described in HT, their polynomial disk models are well suited to numerical analysis and adopted here as described in §2. I also tried exponential disks with different basis sets but could not find an alternative which allowed an accurate computation of the height alone, rather than combining height times the surface density.

4 WAKES AND WARPS

4.1 Wake in halo

There are two contributions to the disk disturbance: the direct tidal force of the satellite on the disk and the force of disturbance excited by the satellite in the halo—the halo wake—on the disk. The strength of the halo wake is proportional to the mass of the satellite and mass density of the halo. However it depends critically on the orbital structure of the halo because a particle's orbit will respond most strongly near resonances between the satellite's and halo particle's orbital frequencies. Co-orbiting trajectories will have the strongest response but the total mass involved is small for the standard model. Higher-order resonances will be weaker but occur at smaller galactocentric radii where the mass density is high. The wake is the product of these competing effects and, generally, the wake peaks far inside the satellite orbit.

As an example, Figures 2 and 3 show the space density distortions induced by the LMC orbit in the standard King model halo in the orbital plane. The satellite has pericenter at $R = 7$ and apocenter at $R = 14$ and here, orbits in the counter-clockwise direction. Pericenter is $X = 7, Y = 0$. If the satellite were completely outside the halo, the dipole response would be a linear displacement representing the new center of mass position. The wake would be proportional to $-d\rho(r)/d\mathbf{r} \cdot \mathbf{e}$ where $\mathbf{e}$ is the unit vector from the halo to the satellite center (Weinberg 1989). In Figure 2, the satellite is inside the halo, and the wake deviates from the pure displacement. The amplitude of the quadrupole (Fig. 3) is a factor of roughly two smaller than the dipole. The dominant wake is near the satellite pericenter as expected, but note the inner lobe of the wake at roughly half the pericenter distance. This is due primarily to the 2:1 resonance between
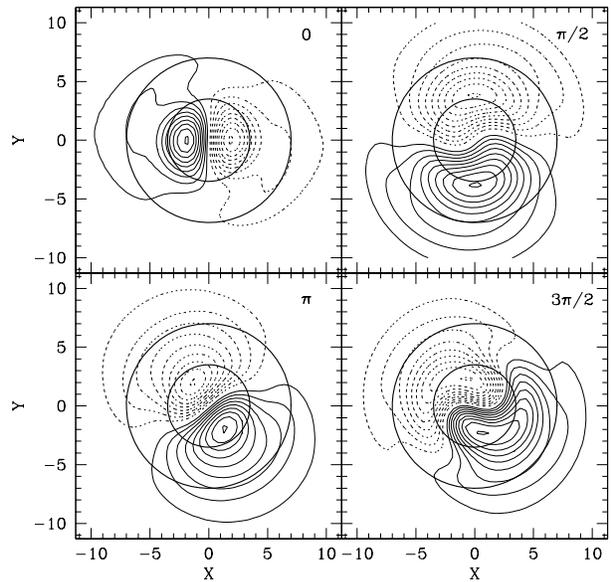

Figure 2. Wake in the fiducial halo due to orbiting satellite for $l = m = 1$ (dipole). Nine linear spaced contours of overdensity (solid) and underdensity (dotted) shown. Each successive panel shows the wake radial phase as labeled, with pericenter at phase 0 and apocenter at phase π. Satellite locations at the four phases are $(X, Y) = (7.0, 0.0), (0.2, 11.6), (-8.5, 11.2), (-11.3, 2.9)$. The outer and inner circles indicate the pericentric and half-pericentric radii, respectively.

satellite and halo orbital azimuthal frequencies. Although the relative density of this inner lobe is smaller than the primary outer one at pericenter (phase 0), both the proximity and spatial structure causes the force from the inner wake to dominate over direct force from the satellite. As the satellite approaches pericenter (phase $3\pi/2$), these inner lobes become relatively stronger and can dominate the response. The inner galaxy wake is weaker past pericenter (phase $\pi/2$).

4.2 Vertical force on disk due to wake

In the absence of the halo, the first multipole contributing to the differential or *tidal* acceleration of the disk is at quadrupole ($l = 2$) order. It is straightforward to convince oneself of this fact: the $l = 0$ term is constant yielding no force, the $l = 1$ term is proportional to r yielding a spatially constant force, and therefore the $l = 2$ term provides the lowest order differential force. Because the warp has $m = 1$ symmetry, the dominant warp inducing term will be $l = 2$, $|m| = 1$. Similar symmetries apply for the action of the perturbed halo on the disk. Including the halo warp, the dipole still can not produce the classic odd integral-sign warp but causes an even distortion. The lowest order halo wake that can excite an integral-sign warp is also the $l = 2, |m| = 1$ component. To illustrate the three-dimensional structure, Figure 4 renders the isosurface corresponding to 75% of peak amplitude wake contoured in Figure 3. The wake is symmetric about the satellite's orbital plane and this plane is easily visualized. A wake must asymmetric about the z axis to
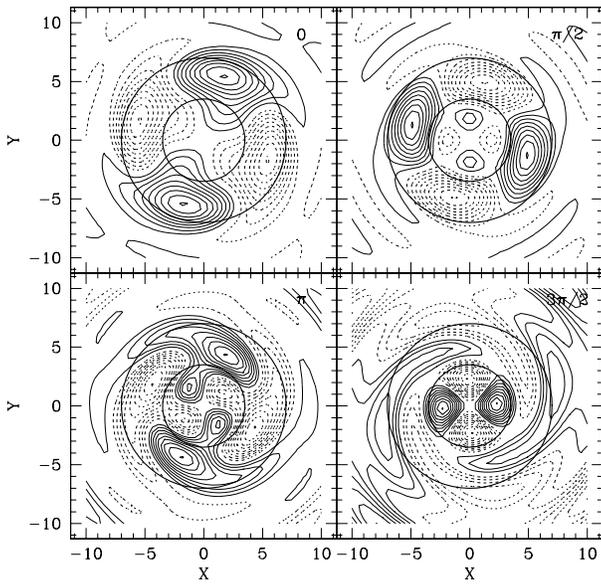
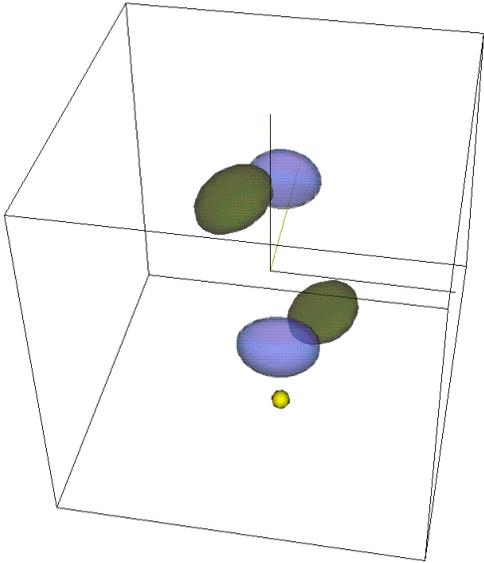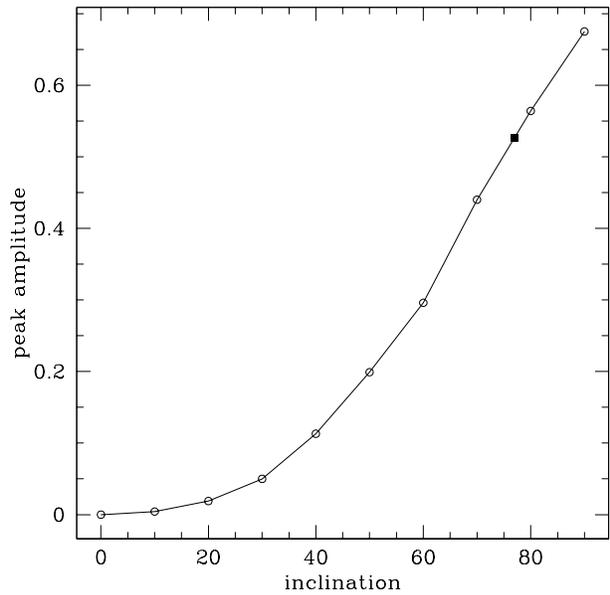**Figure 3.** As in Fig. 2 but for $l = m = 2$ (quadrupole)halo

Figure 5. Warp amplitude as a function of orbital plane inclination. A polar orbit produces the maximum warp with heights 473 pc (1181 pc) for LMC mass of $6 \times 10^9 \, M_\odot$ ($1.5 \times 10^{10} \, M_\odot$). The filled square shows the elevation of inferred LMC orbit.

Figure 4. A three dimensional rendering of the $l = |m| = 2$ halo wake shown in Fig. 3). The isolevel shown at 75% of the peak amplitude. Overdensity and underdensity are shaded light grey and dark grey, respectively. The wire-frame outline extends ± 10 units in x, y, and z and the x–y–z axes are shown with the z-direction along the vertical. The satellite pericenter is a bit over 7 units and shown as a small sphere.

cause a differential vertical acceleration of the disk; in other words, a satellite in the disk plane produces no warp. The maximum vertical force occurs when the pattern shown in Figures 3 and 4 is oriented perpendicular to the disk plane (see Fig. 5). The orientation of the LMC disk plane is nearly ideal for producing a disk warp.
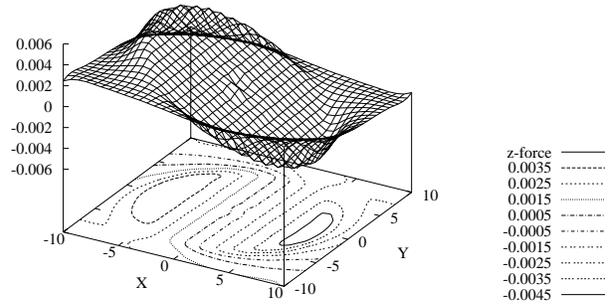
Figure 6. Vertical force on disk plane due to orbiting satellite only for harmonics $l = 2, |m| = 1$. Shown midway between pericenter and apocenter, at radial phase $\pi/2$ (cf. Fig. 2).

Figures 6–7 describes the vertical force on the disk plane due to both the satellite alone and the halo response to the satellite for the standard model. Although the net force from the satellite is similar in magnitude to that from the halo, the satellite force is varies linearly with distance and only gives rise to a uniforming tilting of the disk. However, the spatial structure in the halo force causes a differential bending of the disk with vertical $m = 1$ symmetry: a warp.

4.3 Vertical response of the disk

One can think of the vertical response as a superposition or packet of modes. Bending modes in the absence of a halo
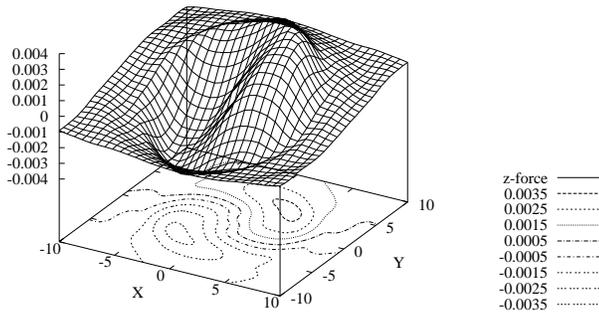
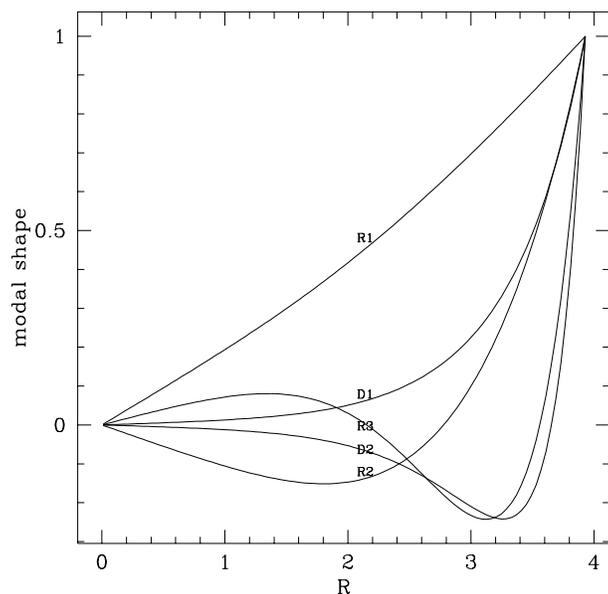**Figure 7.** As in Fig. 6 but for the halo response to the satellite.

Figure 8. Tipping modes for the Hunter $N = 16$ disk with edge at $R = 4$ (approximately 28 kpc) embedded in a $W_0 = 3$ King model with $R_t = 28$ (approximately 200 kpc scaled to the Milky Way). Modes are labeled in the notation of HT.

Figure 9. Evolution of the tipping modes for the Hunter-Toomre 16X disk embedded in a $W_0 = 3$ King model with $R_t = 28$ (approximately 200 kpc scaled to the Milky Way) as a function of halo mass. Curves are labeled by halo-to-disk mass ratio. At zero halo mass, the mode is bodily tipping of the disk at zero frequency. Each curve parallel to the y axis describes the mode with an increasing halo mass fraction shown on the x axis, up to a possible total of $M = 30$ times the disk mass.

Figure 10. Vertical response of the Hunter-Toomre 16X disk due to the combined vertical force from the standard satellite and halo. Scaled to the LMC and Milky Way following §2, the peak height is 420 pc (1050 pc) for the lower and higher LMC mass estimates, respectively.

have been described by HT. The addition of the halo shifts and slightly modifies the shape of these modes although they are qualitatively similar to those described in Figure 3 of HT. The corresponding modes in the standard halo model ($W_0 = 3$ King model halo with truncation radii at $R_t = 28$ [200 kpc]) are shown in Figure 8. In the absence of the halo, the zero frequency $m = 1$ mode is a bodily tipping of the disk. As the halo mass increase relative to the disk mass, the shape of the modes change. The tipping mode, for example, evolves into a distortion with warp-like structure. For the standard (Hunter-Toomre 16X) disk, the tipping mode is the only discrete mode and appears to be a prominent feature of the response discussed below. Its shape for increasing halo to disk mass ratio is shown in Figure 9. The tipping mode is distorted from linear to an integral-sign shape.
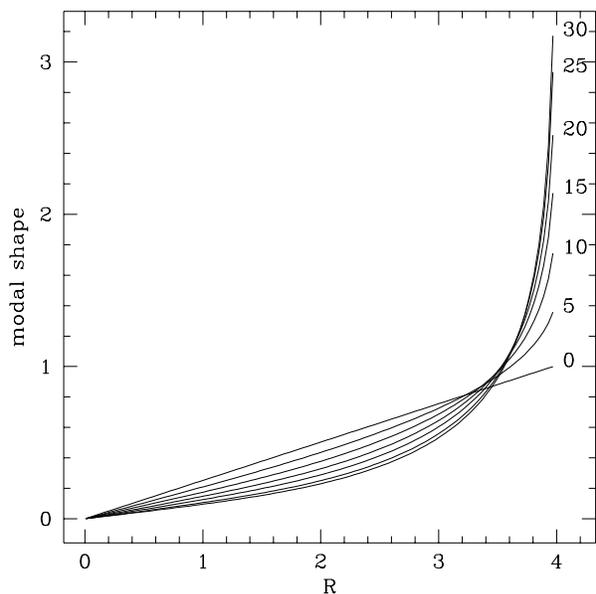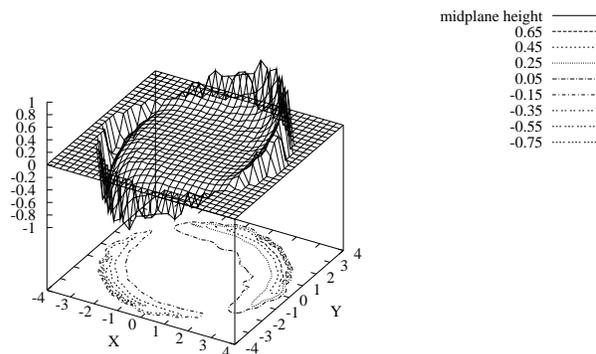
The combination of the force exerted by the satellite and the satellite-induced halo wake excites a vertical response in the disk. A strong vertical disk response obtains for near commensurable halo-wake pattern speeds and disk bending mode frequencies. These are *accidental* resonances in the sense that their existence is circumstantial and not the result of a tuning mechanism. Figure 10 shows the response to the combined force from Figures 6–7. The vertical
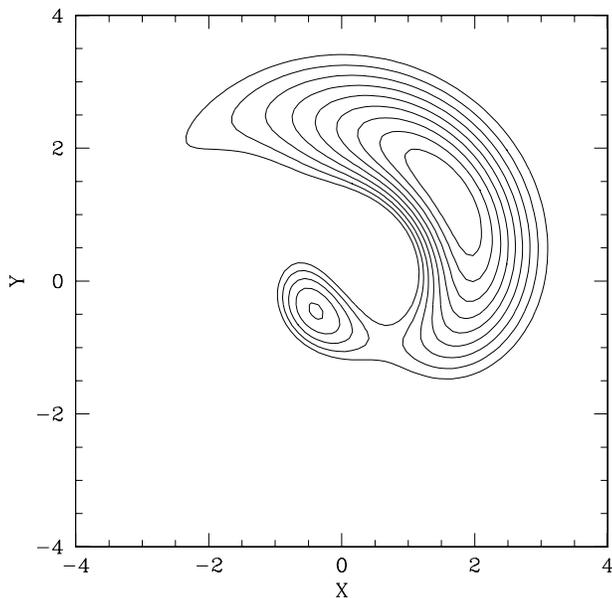
Figure 11. $m = 1$ density distortion in the disk plane due to the satellite and halo perturbation. Contours are evenly spaced from 10% to 90% of maximum.

Figure 12. As in Fig. 11 but combined with the background for the larger LMC mass estimate (right).

distortion of the midplane within a galactocentric radius of roughly 10 kpc is less than 100 pc, however, the warp reaches a peak height of about 400 pc (1.0 kpc) at $R_g \approx 20$ kpc for LMC mass of 6×10^9 M$_\odot$ (1.5×10^{10} M$_\odot$). In many cases, the warp has a local maximum in the outer disk, corresponding to the location maximum curvature in the integral sign. At larger radii the mode may turn over and reach a global maximum amplitude at the very edge of the disk, corresponding to the ends points of the integral sign. We will refer to this first local maximum as the *peak* when describing warps. The sharp edge has been truncated in Figure 10 and others below to best show this first maximum. We will see below, that depending on the halo structure, both larger and smaller warp distortions are possible with the same satellite perturbation.

4.4 In-plane response of the disk

To complete the example, we describe the concurrent in-plane distortion for $m = 1$ and $m = 2$ harmonics. The overall distortion is small in the inner disk, $R_g \lesssim R_o = 8.5$ kpc, and dominated by the $m = 1$ term with a relative density amplitude of roughly 1.6% (4%) for the small (large) LMC mass estimate. The $m = 1$ distortion appreciable in the outer disk, reaching 16% (40%) near $R_g \gtrsim 16$ kpc and increasing beyond that. This distortion will produce an observably lopsided disk outer disk (see Figs. 11–12). Unfortunately, it is difficult to determine precise distances to gas in the outer galaxy and this signature will be difficult to affirm.

The quadrupole leads to a measurable oval distortion only for $R_g \gtrsim 20$ kpc of 6% (16%) and is at the percent level or smaller near the solar circle (see Figs. 13–14). This mild oval distortion is much smaller than and will be swamped by the predicted $m = 1$ signature.
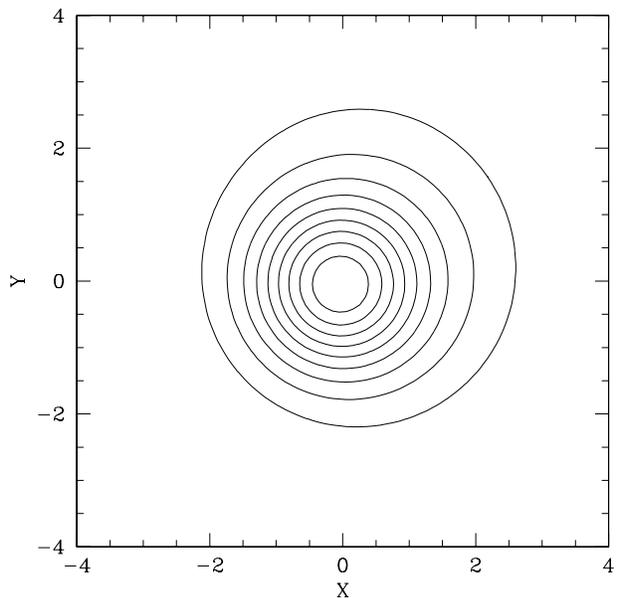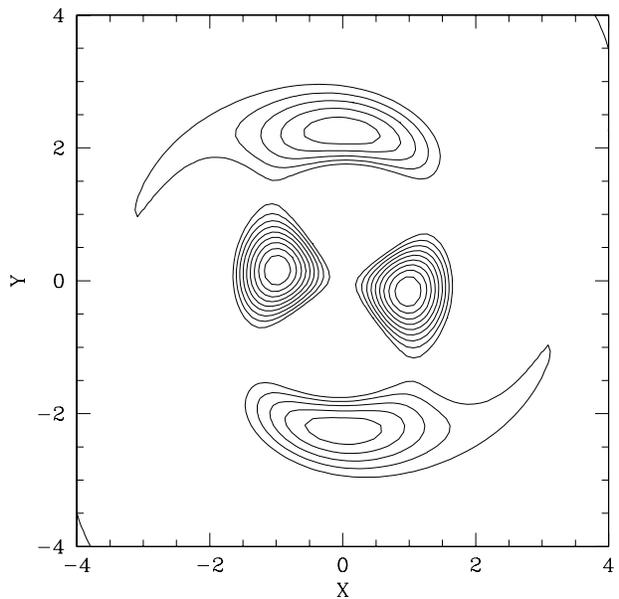

Figure 13. As in Fig. 12 by for the $m = 2$ density distortion.

5 DISCUSSION

There are no simple formulae describing the warp amplitudes in general because of the complexity of the interaction. Qualitative guidelines are as follows. Within the range of scenarios explored here, the most important condition is the coincidence of wake pattern speed and the disk bending-mode frequency. Exploration suggests the 2:1 resonance between the satellite and orbital azimuthal frequencies and the ILR-like resonance are most important. A secondary consideration is the location and amplitude of the wake itself which
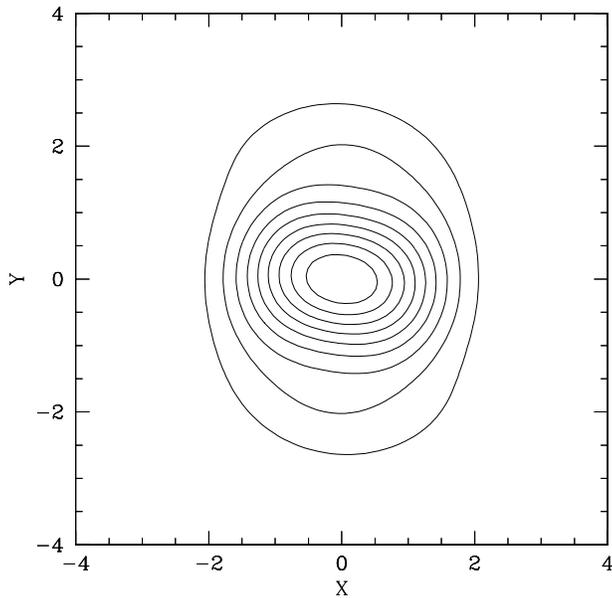
Figure 14. As in Fig. 11 by for the $m = 2$ density distortion. The mass of the LMC is exaggerated by a factor of 16 to illustrate the shape of the feature.

depends on the halo profile and the existence of low-order resonances in the vicinity of the disk. These two features are not independent. However, if one *could* fix the orbital frequencies of halo stars, the wake amplitude would be proportional to the halo density and if one *could* fix the density, the wake location would scale with the orbital frequencies. The halos considered here are chosen to have flat rotation curves between the outer disk and satellite pericenter. Because the wake is dominated by the 2:1 resonance, the wake peaks at roughly half the pericenter distance. Therefore, increasing the mass of the halo will tend to increase the wake amplitude but can decrease the disk warp if the frequency match with the global disk modes is less favorable.

Because of this complicated interplay and sensitivity to the actual disk and halo profiles, I will illustrate the range of possibilities with some examples rather than give an exhaustive set of models.

5.1 Disk model dependence

For a standard halo and satellite interaction, the disk warp in a Hunter $N = 16$ and Hunter-Toomre 16X provide a telling comparison. The Hunter disk has a shallower less centrally concentrated profile than the Hunter-Toomre disk, naively suggesting that the Hunter disk will be more susceptible to an outer disturbance. But, because the lower-frequency mode in the Hunter-Toomre disk better couples to forcing by the halo wake, its response is larger. Figure 15 shows the response for the Hunter disk for comparison with Figure 10. Figures 17 and 18 compare the same scenario for a more centrally concentrated halo model (King $W_0 = 7$, cf. §5.3). As above, the Hunter $N = 16$ is less warped than the Hunter-Toomre 16X disk.

Figure 19 contrasts the the Hunter-Toomre 16X and the Hunter $N = 16$ profiles. The latter is scaled to $R_{disk} = 3$
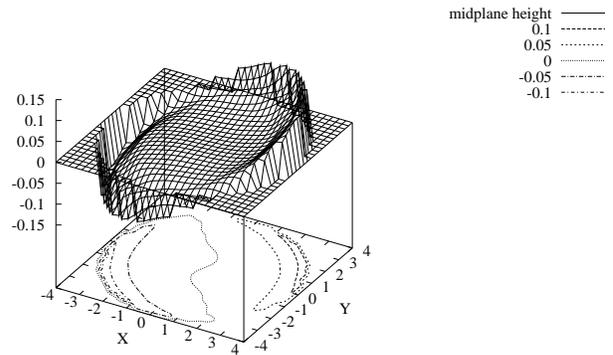

Figure 15. Warp height for the Hunter $N = 16$ disk with outer radius $R_{disk} = 4$ and halo model $W_0 = 3, R = 28, M = 10$ (cf. Fig. 10).
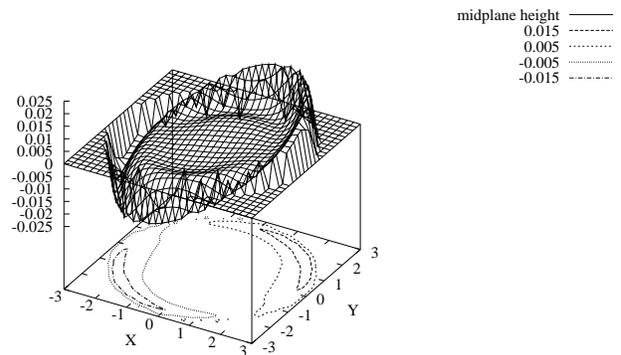

Figure 16. Warp height for the Hunter $N = 16$ disk with outer radius $R_{disk} = 3$ and halo model $W_0 = 3, R = 28, M = 10$ (cf. Fig. 10).
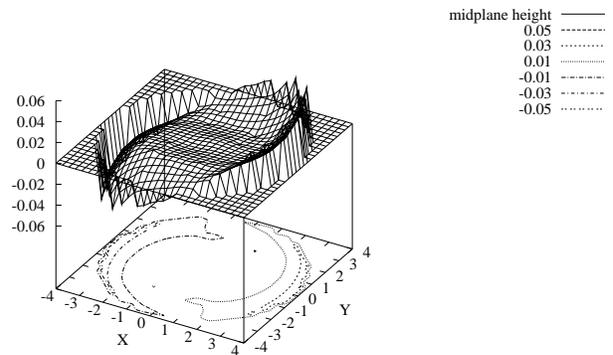

Figure 17. Warp height for the Hunter $N = 16$ disk with outer radius $R_{disk} = 4$ and halo model $W_0 = 7, R = 28, M = 10$.
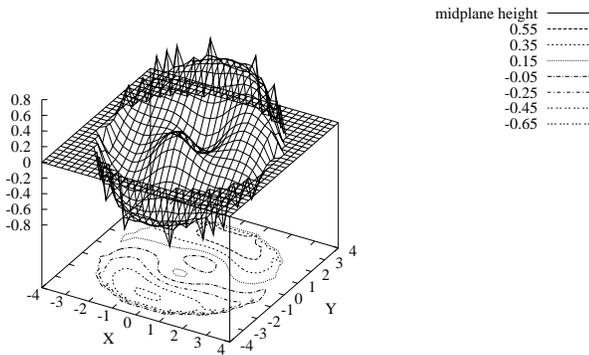
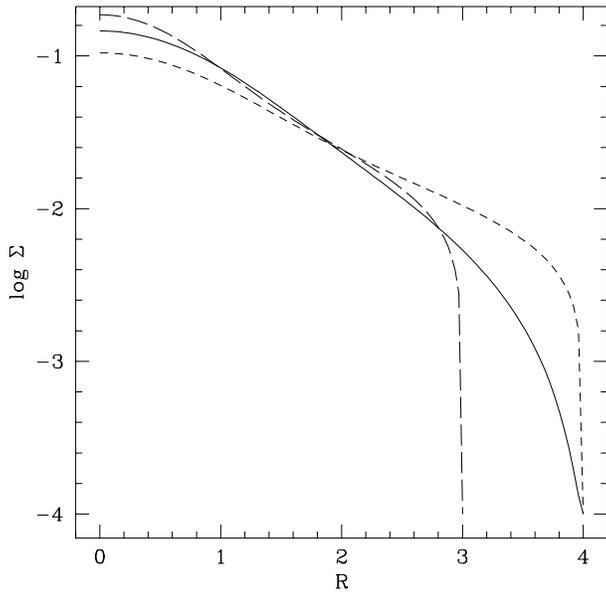**Figure 18.** Same as Fig. 17 but for the Hunter-Toomre 16X disk.

Figure 19. Comparison of the 16X disk (solid) and the Hunter $N = 16$ disks for $R_{disk} = 3, 4$ (long dash, short dash).

and 4. Note that the Hunter $N = 16$ for $R_{disk} = 3$ is quite similar to the Hunter-Toomre 16X with $R_{disk} = 4$ inside of $R \approx 2.8$. Figure 16 shows that the warp in a Hunter $N = 16$ disk with the same standard halo satellite perturbation is much smaller! Although the disk less extended, the mode location and morphology are to blame. The overall weaker halo support means that the tipping mode is closer to a bodily tip and has very low frequency. Therefore, it poorly couples to the wake. The first retrograde and prograde modes, which are warp-like, have relatively high frequencies which also couple poorly.

5.2 Effect of halo mass and satellite orbit

Figures 10, 20 and 21 show the warp in the Hunter-Toomre 16X disk for increasing halo mass, $M = 10, 15$ and 20 respectively, for the standard model. As the halo mass in-

creases, the amplitude of warp amplitude decreases due to a more poorly matched halo-wake pattern speed. In a related scenario, the amplitude of the wake increases with decreasing satellite energy for a fixed halo and disk as expected. However, the warp shape changes with energy because the increasing orbital frequencies couple differently to the continuum disk response.

5.3 Dependence on halo profile

Here, we explore a more concentrated halo, a $W_0 = 7$ King model, and therefore a sub-maximal disk model. The $W_0 = 7$ halo profile is too centrally concentrated to be an acceptable match to standard rotation curves: the rotation curve rises too steeply and over-supports the disk in the inner galaxy. The response for both the 16X and $N = 16$ modes are shown in Figures 17 and 18 for both disks. The warp has a feature inside of of $R = 1$ (inside the solar circle in the Milky Way) due to the larger effect of the halo at smaller radii. The outer warp profile is more gradual and relatively larger at smaller radii. Similar to the standard model, the warp in Hunter-Toomre 16X disk is stronger than in the Hunter $N = 16$ disk.

6 N-BODY SIMULATIONS

The semi-analytic method of §3 is ideally suited to N-body simulation. The same biorthogonal bases can be used to represent the potential and force fields of a particle distribution (e.g. Clutton-Brock 1972, 1973, Hernquist & Ostriker 1992). Because the Poisson equation is linear, each component can be represented by an expansion suited to its geometry. Similarly, one can easily tailor the inter-component forces to isolate any particular interaction. In particular, the n-body tests are not hampered by the difficultly in producing a perfect time-dependent equilibrium disk-halo model; the disk feels the background halo but halo does not react to the background disk, as described in §3.2.

There are two differences between the simulation and the perturbation calculation. First, the disk must be thickened to keep it stable against local instability (Toomre 1964). We use a simple isothermal disk with $\sigma_z = 20\,\mathrm{km\,s^{-1}}$. This disk would also be bar unstable without a bulge component. A rigid Hernquist (1990) model with scale length of 1.4 kpc and mass of 0.2 M_{disk} (roughly $1.2 \times 10^{10}\,\mathrm{M_\odot}$) suppressed bar growth. Second, the direct acceleration by the satellite differentially accelerates the disk causing the expansion center to drift from the center of mass. This technical difficultly was circumvented for the purpose of these tests by dividing the original satellite in two and placing half of of the mass an mirror in the same orbit but 180° out of phase.

Simulations with 100,000 particles with several different partitions between the disk and halo components were performed on a network parallel n-body code using the LAM implementation (Burns, Daoud & Vaigl 1994) of MPI (e.g. Gropp, Lusk & Skjelum 1994). This particular force evaluation scheme lends itself to workstation clusters because the communication overhead is very low (only coefficients need to be passed, e.g. Hernquist, Sigurdsson & Bryan 1995) as long as the load per node remains balanced. To suppress transients, the satellite is slowly turned on over an orbital
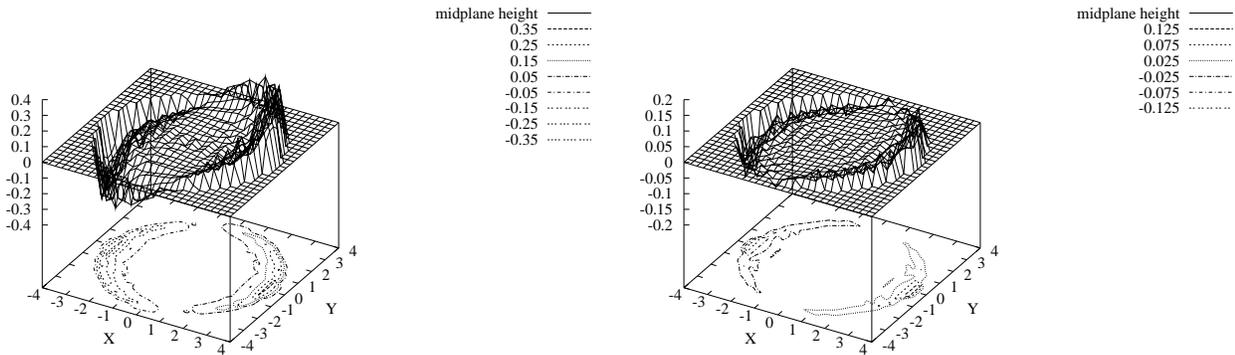
Figure 20. Warp height for 16X disk with outer radius $R_{disk} = 4$ and halo model $W_0 = 3, R = 28, M = 15$. Scaled to Milky Way, 0.1 units corresponds to 175 pc for the large LMC mass estimate (cf. Fig. 10 for $M = 10$).

Figure 21. Same as Fig. Fig. 20 but for $M = 20$.

period. In general, the predicted features were observed: vertical wakes with midplane heights of roughly 500 pc in the outer disk in the predicted orientation.

Unfortunately even with 100,000 particles and the approximations above designed to clearly isolate the $m = 1$ wake, the simulation results are difficult to interpret due to discreteness noise. Simulations without any satellite reveal that the noise has two effects. First, the disk origin random walks in the fluctuating $l = m = 0$ halo component. This causes $m = 0$ vertical distortions which lead to scale height thickening. One should expect these low-order fluctuations are amplified above the Poisson amplitude by the global gravitational response as described in Weinberg (1993, 1997). Secondly, one finds $m = 1$ height distortion from the $l = 2$ noise-excited halo component similar in scale to that expected from the satellite excitation. In retrospect, this might have been expected. Any response to a perturbation will feature any discrete modes. For the halo, the strongest of these will be weakly damped. The halo wake, which is now the self-gravitating response to the disk distortion, will be contain the same weakly damped mode found in the response to the satellite. Therefore, particle fluctuation noise can excite a similar with modes and a similar disk response. The amplitude of the noise excited distortions was as much as 30% of the amplitude observed with a satellite present. However, different Monte Carlo realizations of the same initial model gave vertical warps of arbitrary orientation whereas in the presence of the satellite, the predicted orientation is obtained.

A proper n-body investigation of these effects will require simulations with many more particles and will be left for the future. This investigation, however, inadvertently raises the interesting possibility that a satellite may interact with noise-excited features to produce an overall response which is larger than the time-asymptotic predictions described above. Because the modes are already excited by noise, they may be entrained and strengthed by the external perturbation. This is similar to other noise amplification mechanisms found in Nature, e.g. *stochastic resonance* mechanism observed in neurobiology (e.g. Collins, Chow & Imhoff

1995). Moreover, there are many sources of inhomogeneity in real galaxies such as star clusters, gas clouds, unmixed streams from dwarf dissolution, and other sources of perturbations such as tidal distortions from interactions with a host galaxy cluster and near-neighbor interactions. In fact, the noise in the simulation presented here corresponds to that produced by a halo of $10^6 M_\odot$ black holes (e.g. Lacey & Ostriker 1985). Altogether, it seems likely that observable effects on disks due to halo interactions will be stronger than the predictions in §4.

7 SUMMARY AND FUTURE WORK

This paper summarizes the effect of a satellite companion on a galaxian disk embedded in a responsive (or *live*) halo. Because the mechanisms described here involve multiple time and length scales, the interaction is complex and depends on the details of each dynamical system: satellite orbit, halo and disk structure. The general conclusions and expectations are enumerated below:

(i) The halo can sustain a significant wake in the presence of an LMC-like satellite. This wake is most often caused by a 2:1 orbital resonance and therefore peaks roughly half way between the satellite orbit in the halo center.

(ii) The halo wake, because it has structure at smaller galactocentric radii, can excite warp modes in the disk more efficiently than direct tidal perturbation.

(iii) The strongest warps obtain for a near commensurability between a disk mode and the pattern speed of the halo wake. This makes robust predictions for warp amplitudes more difficult, but we find that warps with observable amplitude can easily result by the mechanism described here.

(iv) For halos with roughly flat rotation curves out to roughly 50 kpc, the satellite orbit (pericenter, plane orientation, and eccentricity) which determines the forcing frequencies and disk profile which determines the bending frequencies are most important in determining the induced warp.

(v) A polar satellite orbit will produce the largest warp.

The inferred LMC orbit is nearly optimal for maximum warp production.

The halo wake (e.g. Figs. 2 and 3) plays a critical role in producing a warp and is an observable consequence of the massive dark halo. A warp survey combined with recent satellite surveys (Zaritsky et al. 1993) may provide additional statistical evidence for the massive halo hypothesis.

Several possibly important interactions have been ignored in this study and present topics for further research. First, the disk is model is infinitely thin; including the three-dimensional structure will damp the warp producing modes but this damping is likely to be small at large scales (Weinberg 1991). Second, the disk respond to the halo wake and the halo responds to the two-dimensional disk distortion but not the three-dimensional one. Therefore the dynamical friction against the halo explored by Nelson & Tremaine (1995) is not included. The calculational method used here is tractable because of the assumption that all transients have mixed away. As they mix, they produce fluctuations on many scales. Intrinsic halo inhomogeneities such as clouds, clusters and massive black holes are also a source of noise. Together, this noise may seed interesting observable features, such as inner bars and arms, which are not part of the long-term wake. In addition, as described in §6, these fluctuations will drive the same modes which produce the wakes at largest scales. N-body simulations suggest that noise-excited structure can have a effect on the disk similar to the satellite companion. This leads to the possibility that the amplitude of the large-scale response to an external disturbance may be amplified by entraining the pre-existing noise-excited features.

ACKNOWLEDGEMENTS

REFERENCES

Bertin G., Pegoraro F., 1989, unpublished
Bertin G., Stiavelli M., 1984, A&A, 137, 26
Binney J., Tremaine S., 1987, Galactic Dynamics, Princeton University Press, Princeton
Burns G. D., Daoud R. B., Vaigl J. R. 1994, in Supercomputing Symposium '94
Clutton-Brock M., 1972, Astrophys. Space. Sci., 16, 101
Clutton-Brock M., 1973, Astrophys. Space. Sci., 23, 55
Collins J. J., Chow C. C., Imhoff T. T., 1995, Nature, 376, 236
Courant R., Hilbert D., 1953, Methods of Mathematical Physics, Vol. 1, Interscience, New York
Dejonghe H., 1989, ApJ, 343, 113
Digel S. W., 1991, Ph.D. thesis, Harvard Univ., Cambridge, MA.
Diplas A., Savage B. D., 1991, ApJ, 377, 126
Edmonds A. R., 1960, Angular Momentum in Quantum Mechanics, Princeton University Press, Princeton, New Jersey
Fridman A. M., Polyachenko V. L., 1984, Physics of Gravitating Systems II, Springer-Verlag, New York.
Gropp W., Lusk E., Skjelum A., 1994, Using MPI: Portable Parallel Programming with the Message-Passing Interface, Scientific and Engineering and Computation Series, MPI Press

Hernquist L., 1990, ApJ, 356, 359
Hernquist L., Ostriker J. P., 1992, ApJ, 386, 375
Hernquist L., Sigurdsson S., Bryan G. L., 1995, ApJ, 446, 717
Hernquist L., Weinberg, M. D., 1989, MNRAS, 238, 407
Hunter C., 1963, MNRAS, 126, 299
Hunter C., Toomre A., 1969, ApJ, 155, 747
Ikeuchi S., Nakamura T., Takahara F., 1974, Prog. Theor. Phys., 52
Kalnajs A. J., 1977, AJ, 212, 637
Kochanek C. S., 1996, ApJ, 457, 228
Krall N. A., Trivelpiece A. W., 1973, Principles of Plasma Physics, McGraw-Hill, New York
Kulkarni S. R., Heiles C., Blitz L., 1982, ApJ, 259, L63
Lacey C. G., Ostriker, J. P., 1985, ApJ, 299, 633
Lawrence C. T., Zhou J. L., Tits A. L., 1994, User's Guide for CFSQP Version 2.5: A C Code for Solving (Large Scale) Constrained Nonlinear (Minimax) Optimization Problems, Generating Iterates Satisfying All Inequality Constraints, Institute for Systems Research, University of Maryland, College Park, MD 20742, Technical Report TR-94-16
Lichtenberg A. J., Lieberman M. A., 1983, Regular and Stochastic Motion, Springer
Lin D. N. C., Jones B. F., Klemola A. R., 1995, ApJ, 439, 652
Meatheringham S. J., Dopita M. A., Ford H. C., Webster B. L., 1988, ApJ, 327, 651
Moore B., Katz N., Lake G., Dressler A., Oemler A., 1996, Nature, 379, 613
Mulder W. A., 1983, A&A, 117, 9
Nelson R. W., Tremaine S., 1995, MNRAS, 275, 897
Nelson R. W., Tremaine S. 1997, Linear response, dynamical friction and the fluctuation-dissipation theorem in stellar dynamics, preprint
Palmer P. L., Papaloizou J., 1987, MNRAS, 224, 1043
Peebles P. J. E., 1995, ApJ, 449, 52
Polyachenko V. L., Shukhman I., 1981, Sov. Astron., 25, 533
Prugniel P., Combes F., 1992, A&A, 259, 25
Robin A. C., Crézé M., Mohan V., 1993, ApJL, 400, L25
Saha P., 1991, MNRAS, 248, 494
Schommer R. A., Olszewski E. W., Suntzeff N. B., Harris H. C., 1992, AJ, 103, 447
Schwarzschild M., 1979, ApJ, 232, 236
Sellwood, J. A., Carlberg R. G., 1984, ApJ, 282, 61
Toomre A., 1964, ApJ, 139, 1217
Toomre A., 1981, in The structure and evolution of normal galaxies, Proceedings of the Advanced Study Institute, pp 111–136, Cambridge University Press
Weinberg M. D., 1986, ApJ, 300, 93
Weinberg M. D., 1989, MNRAS, 239, 549
Weinberg M. D., 1991a, ApJ, 368, 66
Weinberg M. D., 1991b, ApJ, 373, 391
Weinberg M. D., 1993, ApJ, 410, 543
Weinberg M. D., 1994a, AJ, 108, 1398
Weinberg M. D., 1994b, ApJ, 421, 481
Weinberg M. D., 1995, ApJL, 455, L31
Weinberg M. D., 1997, Fluctuations in finite N equilibrium stellar systems, preprint
Wouterloot J. G. A., Brand J., Burton W. B., Kwee K. K., 1990, A&A, 230, 21
Zaritsky D., Smith R., Frenk C., White S. D. M., 1993, ApJ, 405, 464

APPENDIX A: COMPUTING THE DISK DISTRIBUTION FUNCTION

The distribution function is the solution to the following integral equation:

$$\Sigma(R) = \int d^2 v f(E, J), \tag{A1}$$

where E and J are the orbital energy and angular momentum, respectively. For a given energy, the maximum angular momentum, that of a circular orbit with energy E, is denoted by $J_{max}(E)$. The potential of the halo (Φ_{halo}) and disk profile ($\Sigma(R)$, Φ_{disk}) are fixed and assumed to be known to start. We allow the distribution function to be represented by a Gaussian basis in the variables E and $\kappa \equiv J/J_{max}(E)$:

$$\tilde{f}(E, \kappa) = \sum_{i,j} a_{ij} \exp\left[-(E - E_i)^2/2\sigma_E^2 - (\kappa - \kappa_j)^2/2\sigma_\kappa^2\right]. \tag{A2}$$

At any point R_k, the distribution function is related to $\Sigma(R_k)$ through equation (A1)

$$\tilde{\Sigma}(R_k) = 2\int_0^{v_{max}(R_k)} dv_r \int_0^{\sqrt{v_{max}^2(R_k) - v_r^2}} dv_t \tilde{f}(E, \kappa) \tag{A3}$$

where $v_{max}(R_k) = \sqrt{2(E_{max} - \Phi_{disk} - \Phi_{halo})}$, $E = (v_r^2 + v_t^2)/2 + \Phi_{disk} + \Phi_{halo}$, and $\kappa = R_k v_t/J_{max}(E)$. An exact solution requires that the a_{ij} to satisfy equation (A1) for all R. We can state this demand as the set a_{ij} which minimizes the square of the difference of $\Sigma(R_k)$ and $\tilde{\Sigma}(R_k)$ at all R. This demand may be discretized to the set a_{ij} which minimizes

$$\chi^2 = \sum_k w_k \left[\Sigma(R_k) - \tilde{\Sigma}(R_k)\right]^2,$$
$$= \sum_k w_k \left[\sum_{i,j} a_{ij}\Sigma_{ij}(R_k) - \tilde{\Sigma}(R_k)\right]^2 \tag{A4}$$

with the condition

$$\tilde{f}(E, \kappa) \geq 0 \tag{A5}$$

where w_k is a weighting, which may be $w_k = 1$. This defines a standard quadratic programming problem for the a_{ij} which we solve using Powell's algorithm (1982) as implemented in the QLD code provided by Andre Tits (Lawrence, Zhou & Tits 1994).

Distribution functions used for models described here used a grid of 20×20 in E and κ and penalized the expression in equation (A4) to construct a tangentially anisotropic distribution,

$$\chi_p^2 = \sum_k w_k \left\{\left[\sum_{i,j} a_{ij}\Sigma_{ij}(R_k) - \tilde{\Sigma}(R_k)\right]^2 + \lambda \sum_{i,j}\sum_{r,s} a_{ij}a_{rs}\kappa_j\kappa_s^{-\alpha}\right\}, \tag{A6}$$

with $w_k = 1$, $\lambda = 10^{-3}$ and $\alpha = 6$.

APPENDIX B: PERTURBATION COEFFICIENTS FOR AN ORBITING SATELLITE

The biorthogonal expansion coefficients (cf. eq. 7) for a satellite orbiting in a spherical halo are conveniently derived by expanding its gravitational potential in an action-angle series. Assuming a point-mass perturber, the coefficients are

$$b_i^{lm}(t) = Y_{lm}^*(\theta(t), \phi(t))p_i^{lm*}(r(t)). \tag{B1}$$

Because $\mathbf{r}(t)$ is quasiperiodic in two frequencies for a spherical halo, we may expand equation (B1) (or eq. 8 in the main text) in a Fourier series in time.

This is straightforwardly done in the perturber's orbital plane and then rotated to the desired orientation using the rotational properties of spherical harmonics (e.g. Edmonds 1960):

$$Y_{lk}(\pi/2, \psi) = \sum_m D_{km}^{l*}(\alpha, \beta, \gamma)Y_{lm}(\theta, \phi) \tag{B2}$$

where

$$D_{km}^l(\alpha, \beta, \gamma) = e^{ik\alpha}r_{km}^l(\beta)e^{im\gamma}. \tag{B3}$$

The $r_{mk}^l(\beta)$ are the rotation matrices and α, β, γ are the Euler angles describing the orientation to the orbital plane.

Using this and the inverse of equation (B2), we can now expand equation (B1) in action-angle variables. For spherical systems, the third angle variable describes the line of nodes; it has zero frequency and has been suppressed. The action-angle coefficients are then defined by

$$b_{i1}^{lm} = \frac{1}{(2\pi)^2}\sum_k \int d^3w\, e^{-i(l_1w_1 + l_2w_2)}e^{im\gamma}r_{mk}^l(\beta) \times$$
$$e^{ik\alpha}Y_{lk}(\pi/2, 0)e^{ik\psi}p_i^{lm*}(r(w_1))$$
$$= \frac{1}{2\pi}\sum_k Y_{lk}(\pi/2, 0)r_{mk}^l(\beta)\delta_{kl_2}\delta_{ml_3}e^{im\gamma}r_{mk}^l(\beta)e^{ik\alpha} \times$$
$$\int dw_1 e^{-i(l_1w_1 - l_2(\psi - w_2))}p_i^{lm*}(r(w_1)). \tag{B4}$$

The angles w_1 and w_2 describe the equal-time motion from pericenter to pericenter and the mean azimuthal motion respectively. The angle α describes the rotation of the orbital plane; $\alpha = 0$ corresponds to the y axis in the orbital plane coincident with the line of nodes. The angle γ describes the orientation of the line of nodes in the original azimuthal coordinate; $\gamma = 0$ places the line of nodes coincident with the y axis in the original system. Because the orbital radius and the deviation of the true azimuth from the mean azimuth $(\psi - w_2)$ is even in w_1, the integral over w_1 may be simplified and one finds:

$$b_{i1}^{lm} = e^{im\gamma}e^{il_2\alpha}Y_{ll_2}(\pi/2, 0)r_{l_2m}^l(\beta)W_{l\,l_2\,m}^{l_1\,i*}(\mathbf{I}), \tag{B5}$$

where W is defined by

$$W_{l\,l_2\,m}^{l_1\,i}(\mathbf{I}) = \frac{1}{\pi}\int_0^\pi dw_1 \cos(l_1w_1 + l_2f(w_1))p_i^{lm}(r(w_1)) \tag{B6}$$

with

$$f(w_1) = \oint dr\left[2(E - \Phi_0) - J^2/r^2\right]^{-1/2}\left(\Omega_2 - J/r^2\right). \tag{B7}$$

Each term in the expansion in equation (7) is then

$$b_i^{lm}(t) = \sum_{l_1, l_2 = -\infty}^{\infty} e^{im\gamma}e^{il_2\alpha}Y_{lm}(\pi/2, 0)W_{l\,l_2\,m}^{l_1\,i*}e^{i(l_1\Omega_1 + l_2\Omega_2)t} \tag{B8}$$

for $w_{10} = w_{20} = 0$. We will focus on the $m = 1, 2$ responses for $l \leq 4$. Several tests with $l \leq 6$ suggest that $l \leq 4$ dominate the large-scale response.

APPENDIX C: NUMERICAL COMPUTATION OF RESPONSE MATRICES

The elements of the response matrix have been derived in several places (Paper II and references) and are given by:

$$\mathcal{R}_{ij}^{lm}(s) = -4\pi(2\pi)^3 \frac{2}{2l+1} \sum_j \sum_1 \int \frac{dE\,dJ\,J}{\Omega_1(E,J)} i\mathbf{l} \cdot \frac{\partial f_o}{\partial \mathbf{I}} \times$$

$$\frac{1}{s+i\mathbf{l}\cdot\mathbf{\Omega}} |Y_{ll_2}(\pi/2,0)|^2 W_{ll_2 m}^{l_1 i*}(\mathbf{I}) W_{ll_2 m}^{l_1 j}(\mathbf{I}). \text{ (C1)}$$

The inverse Laplace transform as described in §3 requires evaluation of matrix elements of this form in two contexts. The Laplace transform of the coefficients in equation (9), $\hat{b}_i^{lm}(s)$, contribute a simple pole on the real axis is easily performed by deforming the contour to $\Im(s) \to -\infty$. The inverse transform then has the following form

$$\mathbf{a}^{lm}(t) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} ds \mathcal{D}^{-1}(s)\mathcal{R}(s) \cdot \hat{\mathbf{b}}^{lm}$$

$$= \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} ds e^{st} \mathcal{D}^{-1}(s)\mathcal{R}(s) \cdot \mathbf{b}^{lm} \frac{1}{s-i\omega}. \text{(C2)}$$

Although the stability assumption ensures that $\mathcal{D}^{-1}$ has no poles in the $\Re(s) > 0$ half plane, there are poles in the $\Re(s) < 0$ half plane corresponding to damped modes (e.g. Weinberg 1994). Of course, weakly damped modes will be subdominant to an oscillatory mode after sufficiently long time and we assume this limit here.

The elements of $\mathcal{D}(s)$ for the deformed integration path must be analytically continued to $\Re(s) \le 0$ from $\Re(s) > 0$ where the transform is defined. This leads to a Cauchy integral with two simple poles on the real axis for each matrix element (cf. eq. C1): $s = i\omega$ and $s = i\mathbf{l} \cdot \mathbf{\Omega}$. Numerically, this may be straightforwardly evaluated by subtracting the singularity from the integrand and evaluating it separately. Consider the following Cauchy integral:

$$\int dz \frac{f(z)}{z-z_o} = \int dz \frac{f(z)-f(z_o)}{z-z_o} + f(z_o) \int dz \frac{1}{z-z_o}. \text{ (C3)}$$

The first term on the right hand side is non-singular and can be evaluated by simple quadrature. The second term can be performed analytically once the contour is specified. In this case, analytic continuation requires the standard Landau contour (cf. Krall & Trivelpiece 1973). Unlike the case of damped modes, the evaluation of the forced response requires no extrapolation or explicit complex integration.

This paper has been produced using the Royal Astronomical Society/Blackwell Science LaTeX style file.