FINAL REPORT:
Excitation and Evolution of Structure in Galaxies
(NAG 5-2873)

Martin D. Weinberg

95/2/1—96/8/31

Contents

Chapter 1

Overview

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 now 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). However, all disks are embedded within halos and therefore are not dynamically independent. Are halos susceptible to such disturbances as well? If so, can they affect disks and on what time scales?

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. On all but the largest scales, they look like a nearly homogeneous thermal bath 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 et al. 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. In particular:

1. Because the reaction of the halo to a passing galaxy is long-range, a companion can continuously re-excite structure.

2. Because the halo is inhomogeneous and can sustain long-lived patterns at large scales.

The original proposal planned a three-year investigation of the implications of a reactive halo to the observed properties and long-term evolution of galactic disks. This was to include:

1. Finding the strengths of persistent halo features for two classes of galaxy interactions: 1) satellite companions on both circular and elliptical orbits; and 2) tidal encounters or "flybys" for a range of impact parameters and encounter velocities.

2. Determining the impact of the halo features on disks and their role in facilitating and perhaps seeding spiral structure and bars.

3. Development of a hybrid scheme using both n-body simulation and perturbation theory. The perturbation method, sometimes called *matrix mechanics*, solves the equations of motion as a non-linear eigenvalue problem.

After review, the project was funded for one year, $30K total. It was drastically scaled back and renegotiated with the aim of demonstrating feasibility and with the encouragement to make more contact with n-body computation. To this end, I developed two pilot projects. The first, applied the ideas to the LMC—Milky Way interaction resulting in a publication in ApJL which I will briefly summarize in §2. The second sought to verify some of the perturbation theory results using n-body techniques (ApJ, in press). This lead to a new analysis of fluctuation noise in the self-consistent field Poisson solver, summarized §3. Full references to papers published under this award are listed in §A.

Chapter 2

Pilot Project One: the LMC–Milky Way interaction

Previous attempts at disturbing the galactic disk by the Magellanic Clouds relied on direct tidal forcing. However, by allowing the halo to actively respond rather than remain a rigid contributor to the rotation curve, the Clouds may produce a wake in the halo which then distorts the disk. Recent work reported here suggests that the Magellanic Clouds use this mechanism to produce disk distortions sufficient to account for both the location, position angle and sign of the HI warp and observed anomalies in stellar kinematics towards the galactic anticenter and LSR motion. More generally, the observed response depends on the gravitational potential and therefore provides a diagnostic for the structure of the halo and its extent.

2.1 Introduction

All components of spiral galaxies are coupled by their mutual gravitational field. The fact that we only directly observe the luminous components focuses attention on the disk alone in attempts to understand their striking structure. For instance, "grand design" is often correlated with a companion or interloping galaxy and understood to be a result of differential or *tidal* acceleration of the disk. Similarly, researchers have tried to implicate our own fairly massive pair of companions, the Magellanic Clouds, in producing the observed warp in the Milky Way's disk. However the tidal force, which scales as the inverse distance cubed, is nearly an order of magnitude too small to do the job (see Binney 1992 for a review).

The dark halo was postulated more than 20 years ago to account for extended flat rotation

curves and data continue to imply the existence of this unseen mass. True to the original intent, most dynamical theories regard the halo as inert, only providing gravitational support for the luminous components. However the estimated mass of the halo is comparable to that of the disk inside of the Sun's galactocentric radius and dominates at larger radii. Therefore, any structure in the halo must surely affect the disk, and in this letter, I implicate the halo as a co-conspirator for exciting disk structure.

In short, the long-range gravitational response of the halo response can carry the external disturbance by the Magellanic Clouds to sufficiently small radii that the luminous disk can be affected (§2.2) and the disk response is predicted using perturbation theory (§2.3). Previous work has demonstrated that interactions between galaxies and between components of galaxies can be studied perturbatively with good results. Here, the theory is extended to include the combined effects of a disk embedded in a combined luminous and dark-matter halo. The implications and observational comparisons are discussed in §2.4.

2.2   Effect of the Clouds of the halo/spheroid component

The dark halo causes the Clouds orbit to decay by *dynamical friction*: our galaxy will eventual absorb the Clouds (Murai & Fujimoto 1980, Lin & Lynden-Bell 1982, Lin et al. 1995). The response of the 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 results but can include the self-gravity (Weinberg 1989) and most important for our purposes here, gives us a way to estimate the distortion to the galaxy itself.

Specifically, the response can be calculated by solving the collisionless Boltzmann equation (CBE) after all initial transients have decayed (time asymptotic limit). Since the satellite is small compared to the main galaxy, a perturbation expansion is a fair predictor of the true disturbance. The general approach has been tested and compared with n-body simulations in a variety of contexts with good agreement (e.g. Hernquist & Weinberg 1989, Weinberg 1993) and will be used here to consider both the effect of the satellite on the halo and luminous disk.

The strongest response results from the confluence of two conditions: 1) the existence of low-order and therefore strongly responding commensurabilities between the LMC and halo orbit frequencies; and 2) significant mass in phase space near these resonances. In the present case of a flat outer rotation curve, low-order resonances occur at radii roughly between 2

and n times smaller than the satellite orbital radius for a $n : 1$ resonance. In addition, resonances will cluster where frequencies are changing quickly and commensurability is easier to achieve. These regions occur near the characteristic radii where the profile is in transition to its asymptotic form. *Most important for this inquiry, the orbiting satellite can influence the inner galaxy including the luminous disk via the global gravitational response of the halo!*

2.3 Combined effect of the Clouds and halo on the galactic disk

To estimate the amplitude of the disk response to the halo and LMC together, I extended the perturbative solution of the CBE to include a flat disk embedded in the halo. For now, the disk disturbance is confined to the plane (see §2.4 for a discussion of warping). Both the disk and halo react gravitationally to each other, themselves, and the external satellite. The initial disk has the exponential density profile typical of spirals; the distribution function is constructed self-consistently by a quadratic programming technique (Dejonghe 1989) penalized to prefer a cold tangential distribution with an appropriate velocity dispersion. The gravitational field of each component is represented by a truncated series of orthogonal functions (e.g. Clutton-Brock 1972, 1973, Hernquist & Ostriker 1992). With this discretization, the response is a solution to a matrix equation. The individual galaxian components are gravitationally coupled by matrices which transform one set of orthogonal functions to all others. Since an orthogonal polar disk component, proportional to $\exp(im\phi)$, couples to all spherical components Y_{lm} with $l \geq m$, all desired values l for a single m must be considered together. Once the perturbed distribution function is known, all observable quantities such as density and line-of-sight velocity are immediate consequences [1].

2.4 Implications for the galactic warp

Although the inferred LMC orbital plane intersects the galactic disk near the observed warp maxima, quantitative analysis of the tidal distortion predicts amplitudes smaller by almost an order of magnitude smaller than observed (Hunter & Toomre 1969). Unfortunately, the non-local resonant coupling described in §2.3 is very unlikely to be important here because the vertical stellar motions have high frequencies compared to LMC orbital frequencies. Nevertheless, the halo response does reinforce the vertical force on the disk in two ways: 1)

---

[1]See http://donald.phast.umass.edu/Preprints/martin/martin1/lmc_online.html for an MPEG animation showing the disk distortion along the LMC orbit.

the wake increases the total mass in the perturbation by roughly a factor of two; and 2) the response tends to be at smaller radii which further increases its effect. Overall, we may expect a greater than factor of two increase in bending amplitude near the edge of the stellar disk.

To estimate the forced vertical response of the galactic disk, I coupled the Hunter-Toomre machinery to the combined vertical force from the LMC tide and halo wake. To demonstrate the largest effect, the Schommer et al. (1992) mass estimate is taken rather than the smaller Meatheringham et al. (1988) value used for all of the previously quoted results. The peak height of the induced warp occurs near $R_g \simeq 25\,\mathrm{kpc}$ and varies from $400\,\mathrm{pc}$ to $600\,\mathrm{pc}$ with the azimuth of the LMC orbit. The wake in the outer halo lags the satellite by $160°$ and dominates the combined vertical force because of its proximity. The current height of the warp is predicted to be $450\,\mathrm{pc}$ with position consistent with its observed location roughly towards galactic longitudes $90°$ and $270°$. Similar to the conclusions of Hunter & Toomre (1969), one still has to stretch the parameters to account for the observed amplitude. However, the warp calculation does not include self-gravity or the possible resonant interaction with the global bending modes both of which could sustain or amplify the warp. It is encouraging therefore that the predicted position and amplitude at maximum are comparable to observations. To summarize, this work suggests that an *active* halo distorted by the Clouds *can* drive the warp, given the uncertainty in the LMC parameters and the limitations of the estimate.

2.4.1   Future work

The details of the predictions made here depend on the properties of both the LMC and the potential model for the Milky Way, especially the halo. For example, increasing the core radius with fixed mass and tidal radius or increasing the tidal radius with fixed mass and core radius leads to to weaker disk responses probably because of the smaller halo mass inside of the solar circle. Conversely, the dark matter halo is invoked to account for flat rotation curves. The possibility that the halo has dynamically observable consequences in addition to producing the rotation curve provides an independent way to investigate the properties of dark halos in general. Clearly much remains to be done before the full implications of this mechanism is understood. In addition to the shortcomings already discussed, the following three broad topics are ripe for rapid progress:

- *Effect of the three-dimensional structure of disk.* This is a straightforward application for n-body simulation but will require very large numbers of particles. Alternatively, one could use a hybrid approach with an analytic model for the halo distortion together with an n-body disk.

- *Gas response and comparison with observed HI.* It is likely that the time-dependent effects are crucial to understanding the stellar and gas kinematics together. Both periodic orbit analyses and direct n-body simulations with gas could be performed and compared.

- *Non-linear development.* The results presented here describe a linear response only. The slow retrograde distortions might be amplified by the standard mechanisms to drive spiral structure and perhaps inner bars.

Chapter 3

Pilot Project Two: a minimum relaxation n-body algorithm

This report describes a modification of the orthogonal function Poisson solver for n-body simulations that minimizes relaxation caused by small particle number fluctuations. With the standard algorithm, the noise leading to relaxation can be reduced by making the expansion basis similar to the particle distribution and by carefully choosing the maximum order in the expansion. The proposed algorithm accomplishes both tasks simultaneously while the simulation is running. This procedure is asymptotically equivalent to expanding in an orthogonal series which is matched to the distribution to start and truncating at low order. Because the modified algorithm adapts to a time-evolving distribution, it has advantage over a fixed basis.

The required changes to the standard algorithm are minor and do not affect its overall structure or scalability. Tests show that the overhead in CPU time is small in practical applications. The increase in accuracy and decrease in relaxation rate is demonstrated for both axisymmetric and non-axisymmetric systems and the robustness of the algorithm is demonstrated by following the evolution of unstable generalized polytropes. Finally, the empirically based moment analysis which leads to the uncorrelated basis is an ideal tool for investigating structure and modes in n-body simulations and an example is provided.

3.1 Introduction

No n-body stellar dynamical system has zero relaxation; relaxation results from Poisson fluctuations in the coarse-grained mass density and is an intrinsic feature of the underlying

physics. Because n-body galaxy simulations have orders of magnitude fewer bodies than in reality, the relaxation rates are artificially high. A variety of techniques have been devised to help ameliorate this problem, and most strategies trade-off spatial resolution for decreased relaxation.

Accurate representation of the gravitational field based on a sample of discrete points is an example of the more general statistical problem of density estimation and SCF is a special case of an *orthogonal series density estimator*. Tarter and Kronmal (1970, hereafter TK) derive several theorems for density estimation by Fourier expansion and suggest series termination or stopping criteria based on the mean integrated squared error (MISE). The essence of this technique is intuitively clear: the estimated values of series coefficients are compared to their variances and only included in the sum when the signal-to-noise ratio is larger than unity. These criteria have been further studied and improved by Hall (1981, see Izenman 1991 for a review). The improved approach extremizes the MISE to find optimal trade-off between errors in bias and variance. The end result is a set of weights that are unity for high signal-to-noise and zero for low signal-to-noise coefficients. These results are straightforwardly generalized to complete but non-Fourier basis sets. Merritt & Tremblay (1994) recently discussed another rigorous approach to estimating a smooth density and potential from a particle distribution.

The algorithm derived in the attached paper (see §A for detail) further improves the series estimator by empirically determining a best-fitting, uncorrelated basis before applying Hall's method. This prevents loss of globally correlated signal which finds itself distributed among multiple components with low signal-to-noise due to a poor choice of basis. Features due to noise fluctuations are efficiently suppressed. In addition, the analysis determines the statistical significance of apparent features in an n-body simulation, which was the original motivation behind this study.

3.2 Numerical tests

3.2.1 Equilibrium models

Spherical

HO presented several relaxation measures and here we will adopt the r.m.s. change in relative orbital energy per particle, $\langle(\Delta E/E)^2\rangle$. The standard algorithm with no smoothing was compared with optimal smoothing and oversmoothing using the new scheme. In all cases, the quantity $\langle(\Delta E/E)^2\rangle \propto t$ as expected for relaxation. The optimally smoothed relaxation rate is lower than the unsmoothed case but larger than the over-smoothed case. The relaxation

rate in the over-smoothed case is nearly a factor of 3 smaller than the unsmoothed case and largely results from the amplitude shifts in the central value of the potential; this is physical and impossible to eliminate. The rates for a fixed scheme with increasing values of particle number, N, scale slightly faster ($N^{-1.25}$) than the naively expected N^{-1}; the non-Poisson behavior is due to the dependence of the smoothing algorithm on N.

Similar tests were performed for a Plummer sphere using both Clutton-Brock and Hernquist bases which match the outer boundary conditions for the infinite extent, finite mass Plummer model. The relative improvement from the modified algorithm is much larger than for the King model.). The lowest rate obtains for $l_{max} = 0$[1], optimally smoothed, but is similar in rate to the over-smoothed cases. The over-smoothed algorithm with $l_{max} = 4$ results in a relaxation rate that is a factor of 5 smaller than the direct Clutton-Brock expansion.

For $l_{max} = 4$, the unsmoothed algorithm results in nearly identical relaxation with both bases, consistent with relaxation dominated by non-axisymmetric disturbances. The lowest order basis function in the Clutton-Brock set is coincident with the unperturbed Plummer model. Note that both the smoothed Hernquist set and Clutton-Brock set with give very similar results, suggested that the modified algorithm recovers full benefit of the tailored basis set.

Because fluctuation-driven relaxation is part of the true solution to the n-body problem, over smoothing may modify the underlying dynamics. Although oversmoothing improves the long-term coherence of individual orbits, the over smoothing could artificially suppress, for example, a weakly unstable disturbance and therefore should be used with caution.

Axisymmetric disk

Simulations of exponential disks embedded in rigid equal mass King model $W_0 = 3$ halos for stability, verified that a similar decrease in relaxation rate also obtains in two-dimensions. The magnitude of the decrease using optimal smoothing was similar although slightly smaller, factors of 2—3 rather than 3—4 found in three-dimensions. Unlike the spherical cases, over smoothing made little further improvement. Because of the proposed algorithm adaptively constructs a new orthogononal functions, a fiducial Bessel function basis should suffice for finite extent two-dimensional disk models. This should be easily extendible to a three-dimensional thick disk geometry, although this has not been tested here.

---

[1]l_{max} is the maximum expansion order in spherical harmonics for the potential field.

Non-spherical

The same diagnostics were applied to prolate systems to make sure that the behavior in the previous section holds for non-axisymmetric cases. Distribution functions for the perfect prolate models de Zeeuw (1985) with $a = 1, b = c = 0.2$ were generated with atlas technique following Statler (1987). All in all, the performance of the modified algorithm for the non-axisymmetric prolate model is similar to that for the spherical models.

3.2.2 Unstable models

Because the optimal smoothing procedure diminishes the amplitude of low signal-to-noise features, one needs confidence that otherwise growing modes will not be suppressed. For tests, we chose the generalized polytropes investigated by Barnes et al. (1986) and examined the evolution of a sequence with fixed $N = 0.75$ and $-0.8 < M < 0$. The $M = 0$ model is isotropic and the model is more radially anisotropic with decreasing M. For $M \leq -0.25$, the velocity distribution becomes bimodal which is the stability boundary proposed by Barnes et al(1986).

The instability develops rapidly in the most unstable model tested, $M = -0.8$, for both the smoothed and unsmoothed algorithm with $N = 20000$. The growing modes and final states are in good agreement in both cases, although the structure develops in different directions because the noise spectrum is different. As $M \rightarrow -0.25$, the growth of the mode is delayed in the smoothed relative to the unsmoothed algorithm due to the lower-amplitude noise spectrum. For $M = -0.3$, there is no growing instability evident for after 40 half-mass crossing times although a time-series analysis of the $m = 2$ coefficients suggests that the mode is appearing and disappearing transiently. This might be expected very close to the boundary.

A simulation with $N = 0.75, M = -0.4$ using the over-smoothed ($\alpha = 12$) algorithm develops the $m = 2$ instability within 25 crossing times which the same as the measured rate for the optimally smoothed ($\alpha = 1$) case. This suggests that moderate over smoothing does not adversely affecting the modal structure, but that weak modes most likely *will* be modified.

3.3 Summary

Orthogonal series force computation (the SCF method) reduces relaxation by limiting the range of spatial variation to large scales. Techniques from statistical density estimation are used to derive an optimal smoothing algorithm for SCF which in essence selects the

minimum statistically significant length scale. This is combined with empirically determined orthogonal functions that best represent the particle distribution and separate any correlated global patterns. The overall approach may have application to the more general problem of orthogonal series density estimation.

The procedure provides nearly all of the benefit of an analytically derived basis set which matches the underlying profile at lowest order with the advantage of adaptively following the subsequent evolution. We found that optimal smoothing decreases the relaxation rate by at least a factor of three in most cases. Further smoothing reduces relaxation rate even farther in three-dimensional simulations with only minor loss of resolution. Perhaps more importantly, the algorithm improves the accuracy of the simulatino by suppressing statistically insignificant features by orders of magnitude.

The algorithm adaptively constructs a statistically uncorrelated orthogonal basis from a fiducial orthogonal basis. The fiducial basis is only required to span enough of the function space to be capable of representing the particle distribution. The lowest order member in the new basis represents the underlying profile and successive members represent the dominant disturbance. As a by product, therefore, the new basis is a diagnostic tool for the evolution of structure in simulations and will be useful for identifying multiple growing and damped modes and characterizing forced responses.

Bibliography

Barnes, J., Goodman, J., and Hut, P. 1986, ApJ, 300, 112.

Binney, J. 1992, ARA&A, 30, 51.

Clutton-Brock, M. 1972, Astrophys. Space. Sci., 16(1), 101.

Clutton-Brock, M. 1973, Astrophys. Space. Sci., 23(1), 55.

de Zeeuw, T. 1985, MNRAS, 216(2), 273.

Dejonghe, H. 1989, ApJ, 343, 113.

Hall, P. 1981, Ann. Stat., 9, 683.

Hernquist, L. and Ostriker, J. P. 1992, ApJ, 386, 375.

Hernquist, L. and Weinberg, M. D. 1989, MNRAS, 238, 407.

Ikeuchi, S., Nakamura, T., and Takahara, F. 1974, Prog. Theor. Phys., 52.

Izenman, A. J. 1991, J. Am. Stat. Assoc., 86, 205.

Lin, D. N. C., Jones, B. F., and Klemola, A. R. 1995, ApJ, 439, 652.

Lin, D. N. C. and Lynden-Bell, D. 1982, MNRAS, 198, 707.

Meatheringham, S. J., Dopita, M. A., Ford, H. C., and Webster, B. L. 1988, ApJ, 327, 651.

Merritt, D. and Tremblay, B. 1994, AJ, 108, 514.

Mulder, W. A. 1983, A&A, 117, 9.

Murai, T. and Fujimoto, M. 1980, PASJ, 32, 581.

Schommer, R. A., Olszewski, E. W., Suntzeff, N. B., and Harris, H. C. 1992, AJ, 103, 447.

Statler, T. S. 1987, ApJ, 321, 113.

Tarter, M. and Kronmal, R. 1970, Ann. Math. Stat., 41, 718.

Weinberg, M. D. 1989, MNRAS, 239, 549.

Weinberg, M. D. 1993, ApJ, 410, 543.

Appendix A

Papers published

Weinberg, M. D. (1996), *Production of Milky Way Structure by the Magellanic Cloud*, ApJL, 455, L31.

Weinberg, M. D. (1996), *High-Accuracy Minimum Relaxation N-Body Simulations Using Orthogonal Series Force Computation* , ApJ, 470, 715.