

DYNAMICAL MODELS OF EARLY-TYPE GALAXIES

DYNAMICAL MODELS OF EARLY-TYPE GALAXIES

PROEFSCHRIFT

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van de Rector Magnificus Dr. W.A. Wagenaar,
hoogleraar in de faculteit der Sociale Wetenschappen,
volgens besluit van het College voor Promoties
te verdedigen op donderdag 9 september 1999
te klokke 14.15 uur

door

NICOLAS CRETTON

geboren te Sion, Zwitserland
in 1968

PROMOTIE COMMISSIE

Promotores:	Prof. dr. P. T. de Zeeuw Prof. dr. H.-W. Rix (MPIA, Heidelberg)
Co-promotor:	Dr. Roeland P. van der Marel (STScI, Baltimore)
Overige Leden:	Dr. K. Kuijken (Rijksuniversiteit Groningen) Dr. W. Jaffe Prof. Dr. M. Franx Prof. Dr. G. K. Miley

*Aux trois femmes de ma vie:
Mike, Anne et Babi.*

Front cover:

The large inclined galaxy on the cover is the nearby spiral galaxy M31 in the constellation of Andromeda. The bright object close to the central bulge of M31 is the compact elliptical galaxy M32.

The dynamical models presented in this thesis show that M32 contains a super-massive central black hole of 3.4×10^6 solar masses (chapter 4).

Contents

1	Introduction	1
1.1	Understanding galaxies: structure, formation and subsequent evolution	1
1.2	Massive Central Black Holes	2
1.2.1	How to detect Black Holes ?	2
1.2.2	The problems with the data	2
1.2.3	The problems with the models	3
1.3	Dark Halos	3
1.4	Aim of this thesis	3
1.5	Galaxy Dynamics	4
1.5.1	Prelude	4
1.5.2	Velocity Profiles	5
1.5.3	The tools of the galactic dynamicist	5
1.5.4	Schwarzschild models: a historical perspective	7
1.5.5	Improvements of the Schwarzschild method	8
1.6	Outline and Summary	9
1.6.1	The dark halo of NGC 2434	9
1.6.2	Axisymmetric three-integral dynamical models	9
1.6.3	Black holes in M32 and NGC 4342	10
1.6.4	The internal dynamics of NGC 2320	10
2	The Dark Halo Around the Elliptical Galaxy NGC 2434	13
2.1	Modeling technique	16
2.1.1	Orbit Library	16
2.1.2	Observables	17
2.1.3	Comparison with the observational constraints	22
2.1.4	Regularization	24
2.2	An illustration of the method	26
2.3	Improved mass modeling for elliptical galaxies	29
2.3.1	Choice of potentials	29
2.3.2	An application: The dark matter halo around NGC 2434	29
2.4	Conclusions	35
2.A	The differential mass density	39
2.A.1	Inferring the distribution function from the orbital weights	39
2.A.2	Construction of $f(E)$ -components	40
2.B	Cosmologically motivated star+halo models	41

3	Axisymmetric Three-Integral Models for Galaxies	43
3.1	Construction of Dynamical Models	45
3.1.1	Mass Model	45
3.1.2	Choice of Orbits	46
3.1.3	Orbit Calculation	48
3.1.4	Storing the Orbital Properties	50
3.1.5	Modeling Observed Kinematical Quantities	51
3.1.6	Fitting the Constraints	52
3.1.7	Regularization	54
3.2	Two-integral Components and Isotropic Components	54
3.2.1	Definition	54
3.2.2	Velocity Profiles	56
3.3	Tests	57
3.3.1	The Test Model	57
3.3.2	Reproducing the Test Model with Two-integral Components	58
3.3.3	Reproducing the Test Model with Regular orbits	60
3.4	Concluding Remarks	64
3.A	Construction of $f(E, L_z)$ - and $f(E)$ -components	68
3.A.1	Normalization	68
3.A.2	Velocity profiles	70
3.B	Relation Between Orbital Weights and the DF	70
3.B.1	$dM / dE dL_z$ for an $f(E, L_z)$ model	70
3.B.2	Scale-free Density in a Kepler Potential	71
4	Improved evidence for a black hole in M32	73
4.1	Mass density and potential	75
4.2	Construction of two-integral models	77
4.3	Predictions of two-integral models	78
4.3.1	Data-model comparison	78
4.3.2	Is the M32 distribution function of the form $f(E, L_z)$?	80
4.4	Construction of three-integral models	81
4.5	Predictions of three-integral models	83
4.5.1	Implementation	83
4.5.2	Data-model comparison	84
4.5.3	Smooth solutions	87
4.5.4	Dynamical Structure	88
4.6	Models with an extended dark nuclear object	92
4.7	Conclusions and discussion	94
4.7.1	Summary of results	94
4.7.2	Dynamical stability	96
4.7.3	Dynamical relaxation	96
4.7.4	Triaxiality	96
4.7.5	Adiabatic black hole growth	97
4.7.6	Tidal disruption of stars	97
4.7.7	Accretion onto the black hole	98
4.7.8	Forthcoming observations	98
4.A	χ^2 topology for orbit-superposition models	98

5	A super-massive black hole in the S0 galaxy NGC 4342	103
5.1	The data	106
5.1.1	Photometric data	106
5.1.2	Kinematic data	106
5.2	The mass model	107
5.3	Jeans modeling	109
5.3.1	Formalism	109
5.3.2	Application to NGC 4342	110
5.4	Three-integral modeling	113
5.4.1	The method	113
5.4.2	The orbit library	113
5.4.3	The observational constraints	114
5.4.4	Non-negative least squares fitting	115
5.4.5	Regularization	116
5.5	Shortcoming of the Gauss-Hermite parameterization	116
5.6	Results and discussion	119
5.6.1	The black hole and mass-to-light ratio	119
5.6.2	Alternatives to a black hole	124
5.6.3	The dynamical structure of NGC 4342	127
5.6.4	The influence of central radial anisotropy	130
5.6.5	The density distribution in the outer region of NGC 4342	132
5.6.6	Comparison with other BH detections	132
5.7	Conclusions	133
6	The Distribution of Stellar Orbits in NGC 2320	139
6.1	The data	141
6.1.1	Photometric data	141
6.1.2	Stellar kinematic data	142
6.1.3	Kinematics of the ionized gas	143
6.2	The models	144
6.2.1	The mass model	144
6.2.2	The dynamical models	146
6.3	Results	147
6.3.1	Fits to the data	147
6.3.2	Goodness of fit	148
6.3.3	Intrinsic velocity dispersions	151
6.4	Conclusions and Discussion	152
	Nederlandse Samenvatting	155
	English Summary	161
	Curriculum Vitae	165
	Acknowledgments	167

Chapter 1

Introduction

1.1 Understanding galaxies: structure, formation and subsequent evolution

GALAXIES occupy a central place in astronomy since they are the constituent bricks of the Universe. They are intermediate in size between stars and super-structures (clusters and super-clusters of galaxies). Understanding the formation and evolution of galaxies involves many different areas of astrophysics.

Galaxies are believed to form hierarchically in a universe dominated by dark matter: small scale objects collapse first and successively merge into larger objects. Cosmological N-body simulations have dramatically improved over the last few years and now have the ability to resolve galaxies (Evrard, Summers & Davis 1994; Navarro, Frenk, & White 1996; Moore *et al.* 1999 and references therein). Density profiles and shapes (Tissera & Dominguez-Tenreiro 1998) of dark matter halos can be predicted and compared to “observations”. Dynamical modeling of real galaxies is an essential tool in that respect, since quantities of interest are generally not readily observable.

Better observations and larger surveys have opened new questions. For instance, the superb optical capability of the Hubble Space Telescope (HST) has revolutionized (amongst other things) our ideas about the central regions of elliptical galaxies (Crane *et al.* 1993; Jaffe *et al.* 1994; Lauer *et al.* 1995; Carollo *et al.* 1997). These observations found that essentially all ellipticals have central density cusps (i.e., the luminosity density approaches a power-law form $\rho(r) \propto r^{-\gamma}$ at small radii) instead of a flat constant density core. More important, it seems that there is a *dichotomy* between giant ellipticals ($M_B < -20.5$) with shallow cusps (mean $\gamma \sim 0.8$), and low-luminosity ellipticals ($M_B > -22.0$) with steep cusps (mean $\gamma \sim 1.9$) (see Gebhardt *et al.* 1996). These nuclear properties also appear to correlate with *global* parameters of the parent galaxy: mean rotation and apparent shape. These correlations hint to some fundamental underlying process. The origin of these density cusps is actively investigated by several groups. The dynamical effect of a central massive black hole (hereafter BHs) is certainly a key ingredient here (Gerhard & Binney 1985; Quinlan, Hernquist & Sigurdsson 1995; Merritt & Quinlan 1998; Nakano & Makino 1999).

Several other fundamental questions regarding structure and formation of galaxies can be summarized as follows :

- What is the distribution of intrinsic shapes in elliptical galaxies and do they correlate with other properties such as e.g., luminosities or cusp slopes? Twenty years ago, the answer to that question was that elliptical galaxies were triaxial as a class. Today, it is still believed that they are triaxial systems, although probably not strongly triaxial, i.e., near-oblate or near-prolate (Franx *et al.* 1991; Merritt & Tremblay 1996; Statler, Smecker-Hane & Cecil 1996).

- What is the origin of scaling relations (Tully–Fisher or Fundamental Plane)? The Fundamental Plane relations for elliptical galaxies deviate slightly from what can be deduced from the virial theorem (Pahre, Djorgovski & de Carvalho 1995). Is this tilt due to a dependence of the mass–to–light ratio M/L on mass, to a difference in the dark matter content, or to different anisotropies amongst galaxies? But whatever the origin of this deviation, the scatter in these relations is small despite the large differences between galaxies (in luminosity and dispersion profiles, rotational support) and needs to be explained.
- What are the effects of environment? Minor mergers or weak tidal interactions (galaxy harassment, Moore *et al.* 1996) probably play a crucial role in modifying the morphology of cluster galaxies. This may be the reason why rapid rotation is observed in the outer regions of Cen A and NGC 1399 (a redistribution of angular momentum outwards can occur in mergers, see e.g., Hernquist, 1993).

1.2 Massive Central Black Holes

Massive BHs have been proposed in the 60's as the most probable source of energy for quasars (Zel'dovich & Novikov 1964, Lynden–Bell 1969). From simple considerations on quasar number counts and accretion efficiency, one can estimate the mean density of BHs in the universe. Chokshi & Turner (1992) find that $\rho_{\text{BH}} \simeq 2 \times 10^5 M_{\odot} \text{Mpc}^{-3}$ for a radiative efficiency of 0.1. This result compared to the galaxy luminosity density of $1.4 \times 10^8 L_{\odot} \text{Mpc}^{-3}$ (with $H_0 = 100 \text{ km/s/Mpc}$) implies that the mean BH mass per solar luminosity is $\sim 1.4 \times 10^{-3} M_{\odot} / L_{\odot}$. Therefore one should expect that a typical bright ($L_* = 10^{10} L_{\odot}$) galaxy harbors a $10^7 M_{\odot}$ BH.

1.2.1 How to detect Black Holes ?

The most common method to “prove” the existence of a BH is to measure the amount of mass and light inside a sphere and then argue that no stellar population could explain it (i.e., the required M/L would be too large for any reasonable stellar population). The mass is usually derived dynamically: one measures velocities at a given distance from the center and essentially relies on Kepler's law to derive the enclosed mass. One can measure the kinematics of the stars (absorption lines), or of the gas (emission lines). In practice, a number of difficulties (observational and in the models) makes this approach not as straightforward as it first seems. Alternatively the volume of the accreting material can be estimated from variability measurements (together with light travel time arguments) and its mass can be derived from the measured luminosity, assuming an accretion rate close to the Eddington limit.

Another method to detect BHs is to model the 6.4 keV Fe $K\alpha$ fluorescence emission line of the central accretion disk. The distorted shape of this line is explained by the gravitational redshift of the photons produced very close to the BH (Tanaka *et al.* 1995, Nandra *et al.* 1997).

1.2.2 The problems with the data

The first problem when hunting for BHs is due to the small size of the region in which the BH dominates the gravity. This can be illustrated with the concept of the “radius of influence”, defined as $r_{\text{inf}} = GM_{\text{BH}}/\sigma^2$ (Peebles 1972), where G is the constant of gravitation, M_{BH} the mass of the BH and σ the central velocity dispersion of the stars. For typical values ($M_{\text{BH}} = 2 \cdot 10^8 M_{\odot}$, $\sigma = 200 \text{ km/s}$) and a distance of 5 Mpc, one finds $r_{\text{inf}} = 1''$, which is close to the resolution achievable from the ground, so only the heaviest BHs in nearby galaxies could be clearly detected with ground-based telescopes. In this respect, the Hubble Space Telescope (HST) has made a critical contribution by improving the spatial resolution by a factor ~ 10 . Very high resolution radio observations of water masers can be obtained with VLBI techniques

and have demonstrated the existence of a $3 \times 10^7 M_{\odot}$ BH in the center of NGC 4258 (Miyoshi *et al.* 1995). Unfortunately, the number of galaxies with known water masers is very small.

Furthermore, only one component of the velocity can be measured (the line-of-sight velocity) so with no a priori ideas about the orbit, it is difficult to derive the full three-dimensional motion. It can be done in the case of a “tracer” population, for which we know the dynamics, e.g., thin cold disks are mainly made up of circular orbits.

1.2.3 The problems with the models

The second type of problem in the quest of BHs is more subtle and deals with the dynamical modeling. It basically originates from the mass–anisotropy degeneracy: to a certain extent, one can obtain the same results (in the velocities of the dynamical model) by changing the amount of mass inside a given volume *or* by varying the anisotropy in the orbital distribution.

If a dynamical model (based solely on the luminous mass) under predicts the kinematic data, it can be taken as an indication that some extra (dark) mass is required. But if, due to simplifying assumptions, this model has not enough freedom in its orbital structure to fit the data, the evidence for an additional dark mass is inconclusive, since it remains to be seen if a change in the anisotropy (forbidden by the models) could have improved the poor fit (see an early illustration of this problem in Binney & Mamon 1982). There are some uncertainties in the luminous mass profile as well since it involves a deprojection (see Gerhard & Binney 1996), but it seems to have negligible consequences for the BH search (van den Bosch 1997).

1.3 Dark Halos

Dark halos (DHs) around spiral galaxies are well established, thanks to the relative dynamical simplicity of their extended HI disk (see e.g., Burstein & Rubin 1985). Although their existence and, to some extent, their radial profile is well known, the situation concerning their shape is much more sketchy (see e.g., the review by Sackett 1999). The nature of dark matter is still unknown, though progress “by a process of elimination” seems to favor cold weakly interacting particles over faint stars (because of their rarity in the deep HST images, see Bahcall *et al.* 1994; Flynn *et al.* 1996) and over sub-stellar objects (because of the absence of short timescale events in the MACHO data, see e.g., Freeman 1996).

Comparison with spiral galaxies leads to a number of questions concerning the DH of ellipticals: what is the size, mass, shape and extent of DHs around ellipticals, and how do they compare to their spiral “cousins” of similar luminosity? In trying to answer these questions and modeling these systems, essentially the same problems described in Section 1.2 occur. The absence of a tracer population and the limited radial extent of spectroscopic data are the major difficulties on the observational side. Usually kinematic constraints can be obtained only within one effective radius (hereafter R_{eff}), thus constraining at most half of the luminous matter (see e.g., Bender, Saglia & Gerhard 1994, but see Carollo *et al.* 1995). In the models, the mass–anisotropy degeneracy is the main limitation to the secure detection of dark mass.

1.4 Aim of this thesis

The aim of this thesis is the study of early-type galaxies through the use of realistic dynamical models. A substantial part of this thesis is devoted to the construction of such models, i.e., how to use the theoretical knowledge of galaxy dynamics (theory of orbits, potential, integrals of motion, etc) to build dynamical models of real galaxies with the least number of simplifying restrictions.

These models are customized such as to reproduce as well as possible real galaxies: their luminous mass distribution fits the observed surface brightness, after projection and seeing convolution. Pixel binning and seeing convolution are taken into account when computing the orbit library, such as to mimic the observational setup and to reproduce the observed kinematics.

The models allow to test for dark matter using the following procedure: As a first step, one builds a model in which light is assumed to follow the mass. If no good fit to the data can be achieved, one adds dark matter to the mass model, until a reasonable agreement with the data is obtained. From a grid of models, we can measure quantitatively the amount of dark mass needed. Furthermore, such models provide a (kinematically unbiased) view of the internal structure of galaxies. These equilibrium models can also be used as starting point for an N-body simulation to study, e.g., the stability of the model.

1.5 Galaxy Dynamics

1.5.1 Prelude

The tools and theories of galactic dynamics are used in this thesis to build dynamical models. All the information contained in a dynamical model can be summarized into one quantity: the distribution function (DF). It describes the density of stars over position, velocity and time. For a physical model, this function has to be positive.

Galaxies can be considered collisionless since their relaxation time is much larger than the age of the Universe (see e.g., Binney & Tremaine 1987). It means we can safely ignore the granularity of the potential of the galaxy due to individual stars and compute orbits in a smooth force field. In such a system, the distribution function $f = f(\mathbf{x}, \mathbf{v}, t)$ satisfies a continuity equation, called the Boltzmann equation

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla f - \nabla \Phi \cdot \frac{\partial f}{\partial \mathbf{v}} = 0. \quad (1.1)$$

If the DF does not depend explicitly on time ($\partial f / \partial t = 0$), then the model is said to be steady-state or stationary. From Newton's law of gravitation, one can deduce the so-called Poisson equation that relates the density to the potential (see page 31–32 of Binney & Tremaine 1987),

$$4\pi G \rho(\mathbf{x}) = \nabla^2 \Phi(\mathbf{x}), \quad (1.2)$$

where G is the constant of gravitation.

By definition, the density is the integrated DF over all velocities:

$$\rho(\mathbf{x}) = \int \int \int f(\mathbf{x}, \mathbf{v}) \, d\mathbf{v}. \quad (1.3)$$

If the stellar orbits in the gravitational field $\Phi(\mathbf{x})$ can be populated such as to reproduce the corresponding density (obtained by solving Poisson equation), then the model is self-consistent. Models with dark matter do not satisfy equation (1.2) and are not self-consistent. The relevance of self-consistent models to real galaxies should not be overemphasized, since probably most of the mass of the universe is dark. When applied to real galaxies, such models are constructed primarily to test if the luminous mass can account for the observed kinematics with a reasonable mass-to-light ratio Υ . If such luminous models fail to do so, we can rule them out. Subsequently, stellar & dark models are built to study the amount of dark mass needed to match the data.

Solving for the DF with the Poisson and Boltzmann equations is a difficult mathematical (and computational) problem, since in general phase-space has six dimensions. The Jeans' theorem helps in reducing the dimensionality, because it states that the DF depends on the six phase-space coordinates \mathbf{x} and \mathbf{v} only through the isolating integrals of motion (Lynden-Bell 1962): there are at most three such integrals, but in general they are not all known analytically.

For mathematical convenience, dynamical models with simple choices of DF have been constructed. For instance in the axisymmetric case, two-integral DFs have become popular, since they can now be constructed for *any* type of density profile (Hunter & Qian 1993; Dehnen & Gerhard 1994; Qian *et al.* 1995; Merritt 1996).

If one specifies only the mass distribution, many different DFs (corresponding to different kinematic behaviors) can be found that self-consistently generate the input density profile (see e.g., Pfenniger 1984). Including kinematic constraints greatly helps in reducing the solution space.

1.5.2 Velocity Profiles

The distribution of line-of-sight velocities of the stars can be derived from the absorption lines of the integrated light spectrum. These lines are broadened by the Doppler shift due to the motion of stars. This velocity distribution is called Velocity Profile (VP) or Line-Of-Sight Velocity Distribution (LOSVD). The VP is of great importance, since it is the *observable part* of the DF, i.e., the DF integrated over the quantities that are not accessible to observers: velocities $v_{x'}$ and $v_{y'}$ in the plane of the sky (x', y'), and position along the line-of-sight z' . At a given position (x', y'), the distribution of stars as function of the line-of-sight velocity $v_{z'}$ is given by

$$\text{VP}(x', y', v_{z'}) = \int \int \int \text{DF}(\mathbf{x}, \mathbf{v}) dv_{x'} dv_{y'} dz'. \quad (1.4)$$

In the past, VPs were assumed to follow a gaussian shape and only the first two moments of the distribution were measured. More efficient detectors and better algorithms (see e.g., Rix & White 1992; van der Marel & Franx 1993) allow to derive the full shape of the VP. Even if in practice the VPs are close to gaussians, there are no theoretical reasons to believe they are *exactly* gaussian. In fact, these deviations from perfect “gaussianity” contain important dynamical information about the galaxy. Because of their close relation to the DF (equation 1.4), VPs reflect in their shape the degree of anisotropy of the galaxy (Dejonghe 1987; Winsall & Freeman 1993). In the spherical case, Gerhard (1993) has shown that the knowledge of high-order moments of the VP could break the so-called mass-anisotropy degeneracy: increasing the mass inside a sphere of radius r , and/or varying the amount of anisotropy in such a sphere leaves specific signatures on the VP shapes, therefore suggesting the possibility to disentangle the two processes.

1.5.3 The tools of the galactic dynamicist

To construct a dynamical model, one can resort to various analytic and numerical methods. In this Section, we briefly describe the weak and strong points of the most relevant ones. Below, we have attempted to organize these tools in increasing complexity/capability order, although they are best understood as complementary. Orbit-based models “à la Schwarzschild” are described separately in Section 1.5.4.

1. *Theory of orbits*: In a galaxy model, it is worth studying the family of stable periodic orbits since they provide a first-order approximation of the dynamics. These orbits constitute the backbone of the galaxy because stars are trapped on quasi-periodic orbits around them. Chaotic orbits tend to cluster around instable periodic orbits (for periodic orbits

studies, see e.g., Contopoulos & Papayannopoulos 1980; Martinet & de Zeeuw 1988; Martinet & Udry 1990; Pfenniger & Friedli 1991). Non-intersecting closed periodic orbits are interesting because they describe gas motions.

However, the periodic orbit approach ignores *collective* effects since each periodic orbit (and in particular, its stability) is studied individually in the gravitational field of the model. One example of such collective mode is the buckling of (N-body) bars: thin bars can bend into a V shape (Raha *et al.* 1991; for further references on collective instabilities, see Section 6 of Merritt, 1999).

2. *Separable models*: For these models, all integrals of motion are known analytically, which greatly helps in the construction of dynamical models since the orbital densities can be calculated explicitly (de Zeeuw 1985; Statler 1987). These special potentials support only regular motion. They display a variety of shapes (de Zeeuw, Peletier & Franx 1986) but can not model accurately the central regions of low-luminosity elliptical galaxies (with steep density cusps).
3. *Jeans models*: A partial description of the kinematics of a galaxy can be obtained by solving the Jeans equations. In practice, low order moments of the velocity ellipsoid can be derived only when some symmetry is imposed (spherical or axisymmetric) and when special restriction are made on the kinematics (such as assuming that the DF depends on two integrals only). Fillmore, Boroson & Dressler (1986) and Kent (1989, 1992) modeled the bulges of spirals using this Jeans technique. With the same type of models, Binney, Davies & Illingworth (1990), van der Marel, Binney & Davies (1990) and van der Marel *et al.* (1994) concentrated on elliptical galaxies. The major limitation of this approach is that the Jeans solutions are not guaranteed to be physical, i.e., the corresponding DF may not be positive everywhere.
4. *Basis function DF*: Here, the DF is expressed as a weighted sum of suitable basis functions. The required analytic knowledge of all the integrals of motion limits the applicability of this approach to separable potentials (Dejonghe 1989, Robijn & de Zeeuw 1996), or to potentials for which an approximate integral exists (Matthias & Gerhard 1999, see also Emsellem, Dejonghe & Bacon 1999). If all the integrals of motion are not known analytically, this approach can still be used to build simpler (less general) models, where one assumes that the DF depends only on one or two integrals (as in the Jeans approach). The axisymmetric two-integral models presented in Chapter 3 of this thesis assume that the DF of each component is proportional to delta functions in the two known integrals of motion, energy E and vertical angular momentum L_z . The proportionality factors are then computed numerically through a least-square minimization.
5. *N-body models*: In such models, the force exerted on each particle is computed numerically from the contribution of all the other particles (i.e., the models are self-consistent by construction). These models are fully general (3D and no restriction placed on the velocity ellipsoid) and can be used to study the secular evolution of isolated galaxies (Friedli 1999) or galaxies in interaction (Barnes 1996). The properties of N-body models are directly accessible (e.g., density profile, ellipticity, velocity anisotropy). The number of particles is limited by the available computer resources, so in practice it is still small ($\lesssim 10^6$) compared with the number of stars in galaxies (10^{11}). Furthermore N-body models are difficult to customize to represent a particular galaxy, i.e., with a given density profile or shape: usually, these models relax to a stable configuration that is different from the initial

one (see e.g., Udry 1993). Syer & Tremaine (1996) devised a hybrid method that merges the N-body and the Schwarzschild approaches (Section 1.5.4) to overcome these difficulties.

Observations can hardly be considered as a “tool of the dynamicist”, but nevertheless they constitute essential constraints for analytical and numerical dynamical models.

1.5.4 Schwarzschild models: a historical perspective

In the so-called Schwarzschild method, a large library of orbits is computed in a fixed potential (corresponding to the deprojected luminous density through Poisson’s equation). During the numerical integration of the equations of motion, we store the orbital properties on various grids: projected and intrinsic density, VPs and lowest order velocity moments of each orbit. The gravitational contribution of some additional dark mass can be included in the total mass model. The orbits are combined with weights, such as to fit the input model density and a set of kinematic constraints. The Schwarzschild models allow for arbitrary geometry and put no a priori condition on the degree of anisotropy. Furthermore they do not require an analytic knowledge of the integrals of motion. They can be made-to-measure for the study of individual objects, such as the various applications developed in this thesis. By construction, this type of model is unable to address secular evolution.

This method for constructing dynamical models is relatively young, so it is possible to make an exhaustive list of the various Schwarzschild models built in the last twenty years. In this section, we will focus on axisymmetric and triaxial models only.

Axisymmetric models

Richstone (1980, 1982, 1984) constructed scale-free oblate models in a self-consistent logarithmic potential ($\rho(r) \propto r^{-2}$). Levison & Richstone (1985a, 1985b) explored more realistic luminous densities ($\rho(r) \propto r^{-3}$) in the logarithmic potential (non self-consistent models). Bishop (1986) constructed three-integral axisymmetric models in a separable Kuzmin potential (1956). Fillmore & Levison (1989) did a survey of highly flattened oblate models following a de Vaucouleurs’ surface brightness profile, using self-consistent and logarithmic potentials.

In this thesis, we present several applications of an axisymmetric three-integral implementation of the Schwarzschild method (see Section 1.4). An important feature of our method is that we calculate and fit the full VP shapes, represented by a Gauss-Hermite series. In a parallel effort, Gebhardt and collaborators have developed a similar machinery and have applied it to NGC 3379 (Gebhardt *et al.* 1999), NGC 3377, NGC 4473 and NGC 5845.

Triaxial models

Most of the triaxial models have been constructed for self-consistency studies (with the exception of the Galactic bar models). Schwarzschild built his original models (1979, 1982) to establish the existence of triaxial equilibrium figures. Statler (1987) computed self-consistent models of the “perfect” (triaxial) ellipsoid (for a full account on these integrable models, see de Zeeuw 1985). Statler’s models are composed of four orbit families described in Kuzmin (1973) and de Zeeuw (1985): they are precisely the same four major orbit families constituting the non-rotating triaxial Schwarzschild (1979) models. Levison & Richstone (1987) used a logarithmic potential with a finite density core and found a large set of possible solutions.

Kuijken (1993) studied scale-free non-axisymmetric *disks* and found that only models with an axis ratio larger than 0.7 could be self-consistently reproduced. When the model isodensity contours are made more boxy, it is even harder to reach self-consistency. Syer & Zhao (1998)

did a similar study using the separable scale-free disks models of Sridhar & Touma (1997). They found that no choice of the surface density slope would permit the construction of a self-consistent model. Schwarzschild (1993) extended the Kuijken study to three dimensions. When he excluded the stochastic orbits, he could only make self-consistent the nearly spherical models ($c/a \gtrsim 0.5$, where a and c are the principal axis of the density, see his equation [2.5]). Merritt & Fridman (1996) explored non scale-free triaxial models with weak and strong cusps. Both cases could be made self-consistent if the stochastic orbits were treated like the regular ones. If only regular orbits are considered, no solution was found for either cusp slope. Merritt (1997) extended his previous study (with strong cusp) to a whole grid of axis ratios. He found that only nearly axisymmetric models could be reproduced when only the regular orbits were included (see his Figure 1). Siopis (1999) confirmed the weak cusp results of Merritt & Fridman (1996), i.e., no models could be made self-consistent if the fully mixed chaotic orbits are included. He showed further that a self-consistent solution with chaotic orbits could be found only in the inner regions (containing at most 65 per cent of the mass).

Zhao (1996) and Haefner *et al.* (1999) have constructed triaxial orbit models for the Galactic bar. Zhao found that as much as 40 per cent of the mass could be attributed to “collective orbits” or “components”, representing fully mixed stochastic orbits (see Section 1.5.5).

One main goal of the Schwarzschild models adapted to real galaxies has been to infer the mass distribution (luminous and dark) from a set of photometric and kinematic data. In this thesis, two “distributions” of dark matter have been considered: super-massive central BHs and extended dark halos. But in principle, the study and construction of full DFs is possible (Haefner *et al.* 1999).

1.5.5 Improvements of the Schwarzschild method

Chaotic orbits and components

The treatment of stochastic orbits in the Schwarzschild models poses a numerical and conceptual problem. Some non-regular orbits may need a very long time to reach a well-mixed state. Thus one should numerically integrate such orbits until the level of secular fluctuations has decreased below some (small) threshold, which requires a large amount of CPU time. If we include such orbits before they have converged, we may expect the final model to evolve with time, in conflict with the steady-state hypothesis. Merritt & Valluri (1996) propose to average different stochastic orbits at the same energy, to obtain what they call “fully mixed solutions”. Merritt (1997) simply chooses to exclude entirely these orbits to guarantee the stationarity of the model. Nevertheless, it is not clear quantitatively how much the model will evolve (and on what time-scale) due to these stochastic orbits (but see Schwarzschild 1993 for a first step in that direction).

Another way of dealing with “difficult offsprings” in the Schwarzschild method is to compute analytically the kinematic properties of *components* (instead of orbits), for which the DFs are delta-functions of the *analytically* known integrals of motion. For instance, Zhao (1996) constructed $f(E)$ -components depending only on the energy E in a triaxial potential (see also Haefner *et al.* 1999). These components are useful since they are time-independent (by construction), they include all the orbits (stochastic and regular) having that energy E and are smoother than conventional orbits. In chapter 3 we describe axisymmetric $f(E, L_z)$ -components. In a self-consistent model, the components can be used in conjunction with regular (numerically integrated) orbits.

In the case where all integrals are known analytically (separable model), orbital properties can be calculated using the same ideas, i.e., the orbital DF can be expressed as a triple delta-

function over the integrals E, I_2, I_3 (see e.g., eq. [A15] of Lake & Norman, 1983). In that case, these $f(E, I_2, I_3)$ -components are equivalent to numerically integrated orbits. As already mentioned in Section 1.5.4, this property facilitates the construction of self-consistent models for this kind of potential (see e.g., Bishop 1986; Teuben 1987; Statler 1987).

Spectral dynamics

One can avoid the discretization of space with the help of spectral dynamics. Each regular orbit is fully characterized by a few numbers (amplitudes and frequencies) and the orbits are expressed as truncated Fourier series (Papaphilippou & Laskar 1996, Carpintero & Aguilar 1998). No grid is necessary to store the individual orbital quantities, because they can be quickly computed from the series expansion. These ideas can be advantageously incorporated into Schwarzschild models since orbital VPs can also be written down explicitly (Copin 1998).

Made-to-measure N-body models

As we mentioned earlier, the Syer and Tremaine (1996) algorithm, combining advantages of the Schwarzschild and the N-body approach seems a very promising avenue.

1.6 Outline and Summary

1.6.1 The dark halo of NGC 2434

In chapter 2 of this thesis, we describe the spherical implementation of our dynamical modeling technique, based on Schwarzschild's orbit superposition method. A representative library of orbits is calculated and the non-negative superposition of these orbits is determined that best fits a given set of observational constraints. We calculate and fit the full velocity profile shapes, represented by a Gauss-Hermite series. This allows us to constrain the orbital anisotropy in the fit. Using χ^2 statistics, we assess the relative likelihood of different orbit combinations in a given potential, and of models with different potentials. Aperture binning and seeing convolution of the data are properly taken into account, and smoothness of the models in phase-space can be enforced through regularization.

As an application, we modeled the E0 galaxy NGC 2434, for which we have extended VP measurements (up to $2.5 R_{\text{eff}}$, Carollo *et al.* 1995). First, models with no dark mass are clearly ruled out by the data: no value of the stellar mass-to-light ratio can be found to fit simultaneously all the kinematic data. Using cosmologically motivated dark halos, we show that roughly half the mass inside one R_{eff} must be dark and the rotation curve of this best fit model is flat for radii in $0.3 < R/R_{\text{eff}} < 3$, in particular there is no special feature at the transition radius from light-dominated to dark-dominated region. One could have expected a more centrally peaked rotation curve, since L_* -elliptical galaxies are much denser than L_* -spirals. We find a best fit model for $v_c \simeq 300$ km/s, which means that the DH around NGC 2434 is more massive than for a typical L_* -spiral.

1.6.2 Axisymmetric three-integral dynamical models

In chapter 3, we focus on the *axisymmetric* implementation of our modeling technique. As in the spherical case (chapter 2), our method originates from Schwarzschild's orbit superposition technique, with the inclusion in the models of the full VPs. We first show how to build galaxy models from individual orbits. This provides a method to build models with fully general DFs, without the need for analytic integrals of motion, which is the main strength of this technique.

We then discuss a set of alternative building blocks, the two-integral and the isotropic components, for which the observable properties can be computed analytically. Models built entirely from two-integral components yield DFs of the form $f(E, L_z)$, which depend only on E and L_z .

We have tested our method, by using it to reconstruct the properties of a two-integral model built with independent software. The test model is reproduced satisfactorily, either with the regular orbits, or with the two-integral components. This chapter mainly deals with the technical aspects of the method, while applications to the galaxies M32, NGC 4342 and NGC 2320 are described in chapter 4, 5 and 6.

1.6.3 Black holes in M32 and NGC 4342

In chapter 4 of this thesis, we model the compact elliptical M32 and show that it must contain a $3.4 \times 10^6 M_\odot$ central dark mass. The central density exceeding $10^8 M_\odot \text{pc}^{-3}$ rules out clusters of stellar remnants or a variation in the stellar population as alternative explanations to a BH for the central object. The models can be fully anisotropic, but it appears that even maximally radially anisotropic models can not fit the data without a BH (in particular, the central velocity dispersion). The best fitting model (with a BH of $3.4 \times 10^6 M_\odot$) has a velocity structure reasonably similar to a two-integral model. This is probably why such simplified models could predict roughly correctly the BH mass (Qian *et al.* 1995). The peculiar velocity structure of M32 may have been caused by the central BH (see e.g., Merritt & Quinlan 1998). We find furthermore that such a BH is consistent with the predictions of models in which the BH grows adiabatically into a preexisting core.

In chapter 5, we apply our modeling technique to the S0 galaxy NGC 4342 and find that it contains a $3 \times 10^8 M_\odot$ BH. Again radially anisotropic models can be ruled out and the best fit dynamical structure resembles a two-integral model. The ratio of BH mass to bulge mass is 2.6 percent, one of the highest known amongst galaxies with suspected BHs (Magorrian *et al.* 1998, van der Marel 1999).

1.6.4 The internal dynamics of NGC 2320

In chapter 6, we present direct observational constraints on the internal velocity distribution of the giant elliptical NGC 2320. Spectroscopic data within one R_{eff} along multiple position angles is used to derive the line-of-sight velocity distribution, quantified by the Gauss-Hermite moments. In addition, the gas rotation curve and dispersion profile are measured from the [OIII] emission lines. After correcting for the asymmetric drift, we calculate the circular velocity of the gas, which provides an independent constraint on the normalization of the gravitational potential. As described in chapter 3, we construct dynamical models to interpret the stellar motions. We consider models in which the mass follows the light (i.e., no dark matter) and models with a logarithmic gravitational potential. Using χ^2 -statistics, we combine the stellar and gas data to constrain the value of the mass-to-light ratio Y_V (in the V-band). We find $Y_V = 15.0 \pm 0.6 h_{75}$ for the mass-follows-light models and $Y_V = 17.05 \pm 0.7 h_{75}$ for the logarithmic models. For the latter, Y_V is defined as the mass enclosed within $15''$ divided by the total amount of light in the same volume. Radially constant Y_V models (without dark matter) and logarithmic models (with dark matter) provide comparable good fits to the data and possess similar dynamical structure. For the full range of Y_V permitted by the observational constraints, the models are radially anisotropic along the major axis in the interval meaningfully constrained by the kinematical data ($1'' \lesssim r \lesssim 40''$). Along the symmetry axis, they are more nearly isotropic. We explore the uncertainties associated with the anisotropy profiles by examining the formal 3σ -range of models around the best fitting Y_V .

References

- Bahcall, J. N., Flynn, C., Gould, A., & Kirhakos, S. 1994, *ApJ*, 435, L51
- Barnes, J. E., 1996 in "Interactions and Induced Star Formation", Saas-Fee Advanced Course 26. Lecture Notes. Swiss Society for Astrophysics and Astronomy, XIV, Springer-Verlag Berlin/Heidelberg
- Bender, R., Saglia, R. P., Gerhard, O. E. 1994, *MNRAS*, 269, 785
- Binney, J., Mamon, G. A. 1982, *MNRAS*, 200, 361
- Binney, J., & Tremaine, S. 1987, *Galactic Dynamics* (Princeton: Princeton University Press)
- Binney, J., Davies, R. L., & Illingworth, G. D., 1990, *ApJ*, 361, 78
- Bishop, J. L. 1986, *ApJ*, 305, 14
- Burstein, D., & Rubin, V. C. 1985, *ApJ*, 297, 423
- Carollo, C. M., de Zeeuw, P. T., van der Marel, R. P., Danziger, I. J., & Qian, E. E. 1995, *ApJ*, 441, L25
- Carollo, C. M., Franx, M., Illingworth G. D., Forbes D. 1997, *ApJ*, 481, 710
- Carpintero, D. D., & Aguilar, L. A. 1998, *MNRAS* 298, 1
- Chokshi, A., & Turner, E. L. 1992, *MNRAS*, 259, 421
- Contopoulos, G., & Papayannopoulos, T., 1980, *A&A*, 92, 33
- Copin, Y. 1998, in *Galactic Dynamics*, eds. D. Merritt, J. A. Sellwood, & M. Valluri, Provo: ASP
- Crane, P., *et al.* 1993, *AJ*, 106, 1371
- Dehnen, W. D., & Gerhard, O. E. 1994, *MNRAS*, 268, 1019
- de Zeeuw, P. T. 1985, *MNRAS*, 216, 273
- de Zeeuw, P. T., Peletier, R. F. & Franx, M. 1986, *MNRAS*, 235, 949
- Dejonghe, H. 1987, *MNRAS*, 224, 13
- Dejonghe, H. 1989, *ApJ*, 343, 113
- Emsellem, E., Dejonghe, H., Bacon, R. 1999, *MNRAS*, 303, 495
- Evrard, A. E., Summers, F. J., & Davis, M. 1994, *ApJ* 422, 11
- Fillmore, J. A., Boroson, T. A., & Dressler, A. 1986, *ApJ*, 302, 208
- Fillmore, J. A., & Levison, H. F. 1989, *AJ*, 97, 57
- Flynn, C., Gould, A., & Bahcall, J. N. 1996, *ApJ*, 466, L55
- Franx, M., Illingworth, G., de Zeeuw, P. T. 1991, *ApJ*, 383, 112
- Freeman, K. C. 1996, in "Dark and Visible Matter in Galaxies", ASP Conference Series, Vol. 117, 1997, Massimo Persic and Paolo Salucci (eds.), p. 242
- Friedli, D. 1999, in "The Evolution of Galaxies on Cosmological Timescales", eds J. E. Beckman & T. J. Mahoney, ASP Conf. Ser., (astro-ph/9903143)
- Gebhardt, K., *et al.* 1996, *AJ*, 112, 105
- Gebhardt, K., *et al.* 1999, *AJ*, submitted
- Gerhard, O. E. 1993, *MNRAS*, 265, 213
- Gerhard, O. E., Binney, J. J. 1985, *MNRAS*, 216, 467
- Gerhard, O. E., Binney, J. J. 1996, *MNRAS*, 279, 993
- Haefner, R. M., *et al.* 1999, submitted to *MNRAS*, (astro-ph/9905086)
- Hernquist, L. 1993, *ApJ*, 409, 548
- Hunter, C., & Qian, E. E. 1993, *MNRAS*, 262, 401
- Jaffe, W., Ford, H. C., O'Connell, R. W., van den Bosch, F. C., Ferrarese, L. 1994, *AJ*, 108, 1567
- Kent, S. M. 1989, *AJ*, 97, 1614
- Kent, S. M. 1992, *ApJ*, 387, 181
- Kuijken, K. 1993, *ApJ*, 409, 68
- Kuzmin, G. G. 1956, *Tartu Astr. Obs.*, 33, 27
- Kuzmin, G. G. 1973, translated in *IAU Symposium 127, Structure and Dynamics of Elliptical Galaxies*, ed. P.T. de Zeeuw (Dordrecht: Reidel), p. 553
- Lake, G., & Norman, C. 1983, *ApJ*, 270, 51
- Lauer, T. R., *et al.* 1995, *AJ*, 110, 2622
- Levison, H. F., & Richstone, D. O. 1985a, *ApJ*, 295, 340
- Levison, H. F., & Richstone, D. O. 1985b, *ApJ*, 295, 349
- Levison, H. F., & Richstone, D. O. 1987, *ApJ*, 314, 476
- Lynden-Bell, D. 1962, *MNRAS*, 124, 1
- Lynden-Bell, D. 1969, *Nature*, 223, 690

- Maggorian, J., *et al.* 1998, AJ, 115, 2285
- Martinet, L., & de Zeeuw, P. T. 1988, A&A, 206, 269
- Martinet, L., & Udry, S. 1990, A&A, 235, 69
- Matthias, M., & Gerhard, O. E. 1999, MNRAS, submitted, (astro-ph/9901036)
- Merritt, D. 1996, AJ, 112, 1085
- Merritt, D., Tremblay, B. 1996, AJ, 111, 2243
- Merritt, D., & Fridman, T. 1996, ApJ, 460, 136
- Merritt, D., & Valluri, M. 1996, ApJ, 471, 82
- Merritt, D. 1997, ApJ, 486, 102
- Merritt, D., & Quinlan, G. D. 1998, ApJ, 498, 625
- Merritt, D. 1999, Elliptical Galaxy Dynamics (review), PASP, 111, 756
- Miyoshi, M., *et al.* 1995, Nature, 373, 127
- Moore, B., *et al.* 1996, Nature, 379, 613
- Moore, B., Quinn, T., Governato, F., Stadel, J., & Lake, G. 1999, submitted to ApJ, (astro-ph/9903164)
- Nakano, T., & Makino, J., 1999, ApJ, 510, 155
- Nandra, K., George, I. M., Mushotzky, R. F., Turner, T. J., & Yaqoob, T. 1997, ApJ, 477, 602
- Navarro, J., Frenk, C., & White, S. D. M. 1996, ApJ, 462, 563
- Pahre, M., Djorgovski, G., de Carvalho, R. 1995, ApJ, 453, L17
- Papaphilippou, Y., & Laskar, J. 1996, A&A, 307, 472
- Peebles, P. J. E. 1972, ApJ, 178, 371
- Pfenniger, D. 1984, A&A, 141, 171
- Pfenniger, D., & Friedli, D., 1991, A&A, 252, 75
- Raha, N., Sellwood, J. A., James, R. A., & Kahn, F. D. 1991, Nature, 352, 411
- Richstone, D. O. 1980, ApJ, 238, 103
- Richstone, D. O. 1982, ApJ, 252, 496
- Richstone, D. O. 1984, ApJ, 281, 100
- Rix, H.-W., & White, S. D. M. 1992, MNRAS, 254, 389
- Robijn, F. H. A., & de Zeeuw, P. T. 1996, MNRAS, 279, 673
- Qian, E. E., de Zeeuw, P. T., van der Marel, R. P., & Hunter, C. 1995, MNRAS, 274, 602
- Quinlan, G. D., Hernquist, L., Sigurdsson, S. 1995, ApJ, 440, 554
- Sackett, P. 1999, in Galactic Dynamics, eds. D. Merritt, J. A. Sellwood, & M. Valluri, Provo: ASP (astro-ph/9903420)
- Schwarzschild, M. 1979, ApJ, 232, 236
- Schwarzschild, M. 1982, ApJ, 263, 599
- Schwarzschild, M. 1993, ApJ, 409, 563
- Siopis, C. 1999, PhD Thesis, University of Florida (astro-ph/9902089)
- Sridhar, S., & Touma, J. 1997, MNRAS, 292, 657
- Statler, T. S. 1987, ApJ, 321, 113
- Statler, T. S., Smecker-Hane, T., & Cecil, G. 1996, AJ, 111, 1512
- Syer, D., & Tremaine, S. 1996, MNRAS, 282, 223
- Syer, D., & Zhao, H. S. 1998, MNRAS, 296, 407
- Tanaka, Y. *et al.* 1995, Nature 375, 659
- Teuben, P. 1987, MNRAS, 227, 815
- Tissera, P. B., & Dominguez-Tenreiro, R. 1998, MNRAS, 297, 177
- Udry, S. 1993, A&A, 268, 35
- Van den Bosch, F. C. 1997, MNRAS, 287, 543
- Van der Marel, R. P., Binney, J., & Davies, R. L. 1990, MNRAS, 245, 582
- Van der Marel, R. P., & Franx, M. 1993, ApJ, 407, 525
- Van der Marel, R. P., Evans, N. W., Rix, H.-W., White, S. D. M., de Zeeuw, P. T. 1994, MNRAS, 271, 99
- Van der Marel, R. P. 1999, AJ, 117, 744
- Winsall, M. L., Freeman, K. C. 1993, A&A, 268, 443
- Zel'dovich, Ya. B., & Novikov, I. D. 1964, Sov. Phys. Dokl., 158, 811
- Zhao, H. S. 1996, MNRAS, 283, 149

Chapter 2

The Dark Halo Around the Elliptical Galaxy NGC 2434

Rix, H.-W., de Zeeuw, P. T., Cretton, N., van der Marel, R. P., Carollo, C. M.
1997, *ApJ*, 488, 702

We describe a powerful technique to model and interpret the stellar line-of-sight velocity profiles of galaxies. It is based on Schwarzschild's approach to build fully general dynamical models. A representative library of orbits is calculated in a given potential, and the non-negative superposition of these orbits is determined that best fits a given set of observational constraints. The most significant new feature of our implementation is that we calculate and fit the full velocity profile shapes, represented by a Gauss-Hermite series. This allows us to constrain the orbital anisotropy in the fit. We also use an objective χ^2 measure for the quality-of-fit, taking into account the error on each observational constraint. Given χ^2 from the observational constraints, the technique assesses the relative likelihood of different orbit combinations in a given potential, and of models with different potentials. In our implementation only projected, observable quantities are included in the fit, aperture binning and seeing convolution of the data are properly taken into account, and smoothness of the models in phase-space can be enforced through regularization. This scheme is valid for any geometry.

In a first application of this method, we focus here on spherical geometry; axisymmetric modeling is described in companion papers by Cretton *et al.* and van der Marel *et al.* We test the scheme on pseudo-data drawn from an isotropic Hernquist model, and then apply it to the issue of dark halos around elliptical galaxies. We model radially extended stellar kinematical data for the E0 galaxy NGC 2434, obtained by Carollo *et al.* This galaxy was chosen because it may be nearly round, in which case the present spherical modeling is applicable. Models with constant mass-to-light ratio are clearly ruled out, regardless of the orbital anisotropy. To study the amount of dark matter needed to match the data, we considered a sequence of cosmologically motivated 'star+halo' potentials. These potentials are based on the CDM simulations by Navarro *et al.*, but also account for the accumulation of baryonic matter; they are specified by the stellar mass-to-light ratio $\Upsilon_{*,B}$ and the characteristic halo velocity, V_{200} . The star+halo models provide an excellent fit to the data, with $\Upsilon_{*,B} = 4.35 \pm 0.35$ (in B-band solar units) and $V_{200} = 450 \pm 100$ km/s. The best-fitting potential has a circular velocity V_c that is constant to within $\sim 10\%$ between 0.2–3 effective radii and is very similar to the best-fitting logarithmic potential, which has $V_c = 300 \pm 15$ km s $^{-1}$. In NGC 2434 roughly half of the mass within an effective radius is dark. In comparison, our models without a dark halo estimate a mass-to-light ratio for the stellar population which is twice as large. If NGC 2434 is a significantly flattened system seen nearly face-on, it would be considerably more difficult to limit the gravitational potential without further observational constraints.

MAPPING how the ratio of luminous to dark matter in a galaxy changes as a function of radius provides an important test for galaxy formation scenarios. Numerical simulations of halo formation in a cosmological context have reached a level where they can predict the radial profile of an isolated dark halo (Navarro *et al.* 1996, hereafter NFW; Cole & Lacey 1996). This profile is altered by the presence of dissipative, baryonic matter, which collects at the center and contracts the dark matter profile. This contraction may provide a natural explanation for the

observed fact that the circular velocity is approximately constant with radius in spiral galaxies (Blumenthal *et al.* 1986; NFW). Elliptical galaxies are more centrally concentrated than spiral galaxies of the same mass, suggesting that they may have circular velocities that are higher in the inner parts than in the outer parts.

Observational studies of the dark halos in spiral galaxies (e.g., van Albada *et al.* 1985) are comparably straightforward. HI gas provides an excellent tracer to large radii. To interpret the kinematics it is justifiable to assume a nearly co-planar distribution with nearly circular orbits, upon which the gravitational potential can be constrained from the observables. By contrast, the majority of elliptical galaxies do not have HI disks that are in equilibrium, and the transition from where the stellar mass dominates to where the dark halo dominates has remained poorly constrained. This has been due to the difficulties in obtaining unambiguous results from the stellar kinematics, as caused by two main problems. First, modeling the stellar dynamics for ellipticals is much more complex than for the cold disks of spirals (e.g., de Zeeuw & Franx 1991; Bertin & Stiavelli 1993; de Zeeuw 1996). Not only is the deprojection of the stellar surface brightness not unique, but also random motions dominate the kinematics, and the stars can occupy a host of qualitatively different orbits in any given potential. Hence, the dynamical modeling must solve for *both* the potential *and* the orbital distribution of the stars, given the observed projected positions and velocities of stars. In practice, most existing studies have not done this. Instead, the orbital structure has often been assumed a priori, by requiring that the distribution function (hereafter, DF; i.e., the number of stars per unit volume of the phase space of stellar positions and velocities) has a certain simple form. The second reason for the poor understanding of the star-halo connection in ellipticals has been the fact that until recently good stellar kinematic data were available only out to approximately one effective radius R_{eff} . This left more than half the stellar mass kinematically unconstrained. In addition, the data were generally restricted to measurements of the two lowest order velocity moments, i.e., the mean streaming velocity V and the line-of-sight velocity dispersion σ . These quantities contain no or little independent information on the intrinsic velocity dispersion anisotropy of the system (Binney & Mamon 1982), which provides the main indeterminacy in the modeling. As a consequence, stellar dynamical indications for the presence of dark halos around elliptical galaxies have remained ambiguous. There have been indications for dark halos in studies that employed restricted classes of dynamical models (e.g., van der Marel 1991; Saglia *et al.* 1992). However, the few models in the literature that had the full freedom of rearranging orbits (e.g., Richstone & Tremaine 1984; Dejonghe 1989) have been able to fit the data for most ellipticals without requiring any dark matter (Katz & Richstone 1985; Saglia *et al.* 1993; Bertin *et al.* 1994). Discrete kinematical tracers in elliptical galaxies (i.e., planetary nebulae and globular clusters) have the advantage that they can be observed to larger radii, but they have the disadvantage of small-number statistics. Dynamical modeling of their observed radial velocities is beset by the same degeneracies that plague the interpretation of the integrated light measurements. Not surprisingly, these studies have produced similarly ambiguous results (e.g., Ciardullo, Jacoby & Dejonghe 1993; Tremblay, Merritt & Williams 1995). Yet, as summarized in various reviews (e.g., Ashman 1992; de Zeeuw 1995; Saglia 1996), there is independent evidence for extended dark matter from X-ray measurements of luminous galaxies (e.g., Forman, Jones & Tucker 1985; Awaki *et al.* 1994), from HI kinematics (Franx *et al.* 1994), and from gravitational lensing (e.g., Maoz & Rix 1993; Kochanek 1995). It is likely that this apparent discrepancy is attributable to the shortcomings of the stellar dynamical tests, i.e., to the uncertainty about the orbital distributions resulting from insufficiently constraining data.

There are several reasons for being optimistic that this situation can be improved and that the luminous to dark matter distribution in elliptical galaxies can now be investigated in some

detail through stellar dynamical studies. First, improved CCD technology (larger detector size, higher QE, lower read-out noise and dark current), combined with improved strategies to ensure accurate sky subtraction and better stellar template matching, have made it possible to obtain stellar kinematics to $2\text{--}4 R_{\text{eff}}$ (Carollo *et al.* 1995, hereafter C95; Statler *et al.* 1996). With data reaching out to these radii, kinematic constraints can be obtained over most of the stellar body. Second, improved analysis techniques (e.g., Rix & White 1992; van der Marel & Franx 1993, hereafter vdMF; Kuijken & Merrifield 1993) make it possible now to extract the entire line-of-sight velocity profile (VP) from the absorption-line spectra, rather than only its lowest order moments V and σ . The higher moments of the VP, which can now be measured, contain essential information about the anisotropy of the velocity distribution (e.g., Dejonghe 1987; Gerhard 1993; C95).

The main purpose of this paper is to describe the development and application of theoretical tools that permit to model and fit VP shape data in a flexible and objective way. In its most general form the modeling technique presented here can answer the question: given a variety of gravitational potentials and given a set of observational constraints (photometry and kinematics, including VPs), what is the relative likelihood of the different potentials? For each potential the orbital distribution is determined that best fits the data, and the likelihood follows from the quality of the fit. Thus the method allows one to determine which are the best-fitting potentials, and which potentials are excluded by the data.

Our technique is based on the numerical calculation of a representative library of orbits in a chosen potential, and the subsequent determination of the non-negative superposition of these orbits that best fits the data. This approach was pioneered by Schwarzschild (1979), who required the orbit superposition to reproduce the galaxy density, and so built triaxial galaxy models. Richstone (1980; 1984) built scale-free axisymmetric models with this technique. In the past decade, the approach has been used to build a variety of spherical, axisymmetric and triaxial galaxy models, which also include the observed radial velocities and/or velocity dispersions as constraints (e.g., Pfenniger 1984; Richstone & Tremaine 1984, 1985, 1988; Levison & Richstone 1985; Zhao 1996).

We have built on this previous work by calculating and comparing an arbitrary number of moments of the Gauss-Hermite series expansion of the VP, and by showing how they can be used as linear constraints on the model. We also use the error on each observational constraint in the superposition procedure to obtain an objective measure for the quality-of-fit, which allows us to compare the relative likelihoods of different models. Only projected, observable quantities are included in the fit. We include a proper seeing convolution of each orbit, so that the observational setup of the data (including aperture binning) can be accurately taken into account. The addition of VP modeling not only allows us to interpret the VP shape data that is now becoming available, but it also removes the need for additional simplifications and assumptions. We do not have to assume that the true lowest order velocity moments of the models can be compared without bias to the best Gaussian fits V and σ obtained from the data. This was done in previous implementations, and is also implicit in modeling based on the Jeans equations (e.g., Merritt & Oh 1997). Avoiding this assumption removes the possibility of systematic errors (easily 10–20%; vdMF) in the interpretation of V and σ .

Our extension of Schwarzschild's method is valid for any geometry. To illustrate and test the new elements in the most straightforward way, we restrict ourselves here to the spherical case, which simplifies the calculation of the stellar orbits. The treatment of the VP shapes and of the orbital superposition is fully general. We test the method on analytic models, and then apply it to the kinematic measurements for the E0 galaxy NGC 2434 obtained by C95; preliminary results from this analysis are published in Rix (1996a,b). The extension of our scheme to

axisymmetric systems is described in Cretton *et al.* (1999, chapter 3 of this thesis) and applied in van der Marel *et al.* (1997, 1998, chapter 4 of this thesis).

The paper is organized as follows. In Section 2.1 we describe the modeling procedure. In Section 2.2 we test the method on isotropic Hernquist models, and in Section 2.3 we use it to study the presence and properties of a dark halo around the E0 galaxy NGC 2434. In the analysis we use a family of cosmologically motivated galaxy potentials that is discussed in Appendix 2.B. We summarize our results in Section 2.4.

2.1 Modeling technique

Our extension of Schwarzschild’s method tests in three steps whether a given potential is compatible with all observational data:

1. A representative library of orbits is calculated in the chosen potential.
2. Each orbit is projected onto the space of observables.
3. The combination of orbits with non-negative occupation numbers is found that best matches the data, taking into account the observational errors.

For computational convenience, we here carry out steps 1 and 2 for spherical geometry. Our description of step 3 is fully general. Other techniques to constrain the potentials of spherical stellar systems through modeling of observed VP shapes are available (e.g., Dejonghe & Merritt 1992; Merritt 1993a; Merritt & Saha 1993), but these do require the availability of analytic integrals of motion. Our approach has the advantage that it can be generalized to more complicated geometries in a straightforward manner.

2.1.1 Orbit Library

In a spherical gravitational potential $\Phi(r)$, all orbits are planar rosettes, characterized by four isolating integrals of motion: the energy E , and the three components of the angular momentum \vec{L} . For each energy, $L = |\vec{L}|$ lies in the interval $[0, L_{\max}]$, where L_{\max} is the angular momentum for the circular orbit at energy E . This circular orbit has a radius r_c , given by the implicit equation

$$\Phi(r_c) + \frac{1}{2}r_c \left[\frac{\partial \Phi}{\partial r} \right]_{r=r_c} = E. \quad (2.1)$$

To cover phase space, we choose a grid in the (E, L) plane (the Lindblad diagram) in the following way. We specify a set of radii $r_{c,i}$, $i = 1, \dots, N_E$, taken to be $\log r_{c,i} = \log r_{\min} + \sqrt{i-1} \log \Delta$. This functional form proved useful because it covers a large logarithmic range for small radii, while (for a suitable step Δ) still providing a grid at large radii that is comparable to the observed resolution. In practice, the radial spacing of the sub-orbits we describe below is $\sim 0.5''$ at the effective radius. Each radius $r_{c,i}$ defines an energy grid point E_i , with an associated $L_{\max,i}$. For each of the N_E grid points E_i , we choose N_L angular momenta $L_{ij} \equiv \beta_j L_{\max,i}$, $j = 1, \dots, N_L$. The numbers β_j are distributed uniformly in the interval $[\epsilon, 1 - \epsilon]$, with $\epsilon = 0.02$. This choice of ϵ excludes the exactly radial and circular orbits, which avoids numerical difficulties in the radial orbit calculation for singular potentials, and avoids the sharp edges in the projected density and VP of a circular orbit. Typically we set $N_E \sim 50$ and $N_L \sim 10$. As we described below, each orbit is in practice constructed by adding ~ 25 sub-orbits of slightly different (E, L) , in order to reduce discreteness effects.

The orbits in a spherical model need only be integrated in the radial dimension, which provides the instantaneous radius r and radial velocity v_r (e.g., Binney & Tremaine 1987). The

tangential velocity v_t follows from $v_t = L/r$. Each orbit is started at its apocenter, and is calculated for an integral number of periods. For a spherical potential one half of a period is in principle sufficient, but calculating a small number of complete periods (~ 5) reduces discreteness effects in the result, arising from the finite number of time-steps. We chose approximately 300 time steps per radial period. During the integration the time steps were adjusted to conserve energy equally well at each step. For every orbit the energy fractional conservation was at least 10^{-5} over the whole integration. The forces and the potential were obtained through power-law interpolation from previously tabulated arrays; the values were tabulated for $10^{-6}R_{\text{eff}} < r < 10^4 R_{\text{eff}}$ in logarithmic steps of 0.001. We checked that all errors from the force interpolation were negligible.

Our orbit integrations are carried out with a simple predictor-corrector integrator, because the orbits need not be computed with great accuracy. If, for example, the orbit were to drift in energy by a small amount during the integration, the orbit would no longer represent a δ -function in phase space, but a short line or small area in the (E, L) plane. But since the (presumably) continuous DF is represented by a finite number of discrete orbits (or ‘basis vectors’), there is no reason to prefer true δ -functions over ‘fuzzy’ ones. Also, physical meaning cannot be attached to rapid fluctuations of the phase space density, given realistically available constraints. Indeed, we have found it advantageous to assemble each orbit (E_i, L_{ij}) from a number of ‘sub-orbits’, typically about 25, whose integrals of motion are drawn at random from a small phase space cube $(\delta E, \delta L)$ around (E_i, L_{ij}) (see also Zhao 1996). This reduces sharp edges in the projected density and VP (especially for nearly circular orbits). It also relieves memory requirements by reducing the number of orbits that must be stored.

Many models can be most fully described analytically if their DFs are isotropic. Hence, such isotropic models are suitable benchmarks for testing, or merely debugging, numerical techniques. It is not entirely trivial how to construct isotropic models with Schwarzschild’s technique. One way to do this is to use smoother building blocks, $f(E)$ -components, in addition to the (E, L) -orbits. These components can be viewed as a weighted combination of orbits with different L , but the same E . They can be constructed almost analytically for spherical potentials, as described in Appendix 2.A.

2.1.2 Observables

Projected orbits

We adopt a Cartesian coordinate system (x, y, z) , with the z -axis directed towards the observer. The associated cylindrical and polar coordinate systems are denoted (R, z, ϕ) and (r, θ, ϕ) . This makes R the projected radius in the (x, y) plane of the sky. The tangential velocity v_t satisfies $v_t^2 = v_\theta^2 + v_\phi^2$. An angle ξ defines v_θ and v_ϕ in terms of v_t , through $v_\theta = v_t \cos \xi$ and $v_\phi = v_t \sin \xi$.

As an orbit is integrated, it is projected onto the space of the observables (x, y, v_z) . In the following, we will refer to the line-of-sight velocity v_z as v . To store the projected orbital properties we adopt a grid (i.e., a storage cube) in the (x, y, v) -space. The spatial grid spacing should be matched to the resolution of the photometric or spectroscopic observations. The velocity coordinate should cover in principle $[-v_{\text{escape}}, v_{\text{escape}}]$. We found in practice that 30-50 velocity bins covering the range $[-4\sigma_{\text{max}}, 4\sigma_{\text{max}}]$ is sufficient, where σ_{max} is the largest observed velocity dispersion. For the specific case of NGC 2434 we have used a $1''$ spatial grid, extending to $80''$, and a velocity grid covering ± 1000 km/s, at 40 km/s sampling.

The orbit integration yields only the phase-space coordinates (r, v_r, v_t) at each time step. However, for projection onto the space of the observables, all six phase-space coordinates are required. These are obtained by drawing a random viewing angle and a random direction of

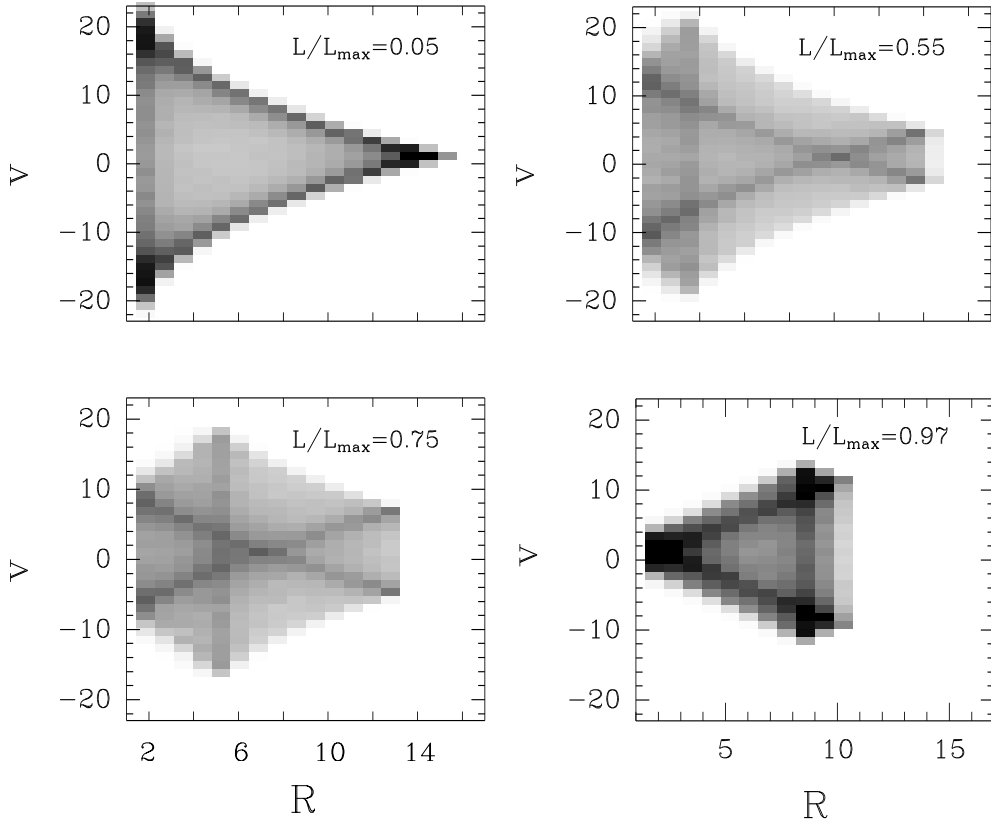


FIGURE 2.1— Grey-scale representations of the occupation weights w_{xyv}^k of four orbits, shown in the (R, v) plane of projected radius vs. line-of-sight velocity. The orbits have the same energy E , but different angular momenta L/L_{\max} . Each orbit is assembled from 25 sub-orbits, as described in Section 2.1.1. The axes are in units of the cell sizes, described in Section 2.1.2. The orbits were calculated in a Hernquist potential at an energy for which $r_c(E) \approx R_{\text{eff}}$. Note the changes in the ‘caustic’ structure from the nearly radial orbit in the top left panel to the nearly circular orbit in the bottom right panel, reflecting the changes in peri- and apocenter distance.

the tangential velocity vector at each time step, i.e., we draw $\cos \theta \in [-1, 1]$, $\phi \in [0, 2\pi]$ and $\xi \in [0, 2\pi]$ from uniform distributions. This yields the following observables at the given time step:

$$x = r \sin \theta \cos \phi, \quad y = r \sin \theta \sin \phi, \quad v = v_r \cos \theta - v_t \cos \xi \sin \theta. \quad (2.2)$$

This procedure properly takes into account the fact that the storage cube should contain the average contribution of *all* trajectories that correspond to the given (E_i, L_{ij}) (i.e., it ensures that the models have DFs that depend only on the modulus of \vec{L} , and not on its direction). The procedure may be repeated numerous times at each time step, to obtain several different (x, y, v) for the same (r, v_r, v_t) . This corresponds to viewing the model from all geometrically equivalent angles at a fixed time, and creates a smoother projection.

We denote the different orbits in the library (each corresponding to a fixed combination (E_i, L_{ij})) with the index k , with $k = 1, \dots, N_o$. The total number of orbits $N_o = N_E \times N_L$. The *occupation weight* of orbit k on the storage cube cell centered on (x, y, v) is denoted as w_{xyv}^k . At the start of the integration of each orbit, the weights are set to zero. As the integration of the trajectory proceeds, and projected coordinates are obtained as described above, a weight is added to the cube cell that contains the projected coordinates. This weight is chosen equal to the

average of the sizes of the previous and the next time-step in the integration. This assignment of weights is effectively a Monte-Carlo integration of the orbit over the grid, because the chance of dropping a weight in a grid cell is proportional to the time the orbit spends in it. Once the orbit calculation (and projections) are finished, the occupation weights for each orbit are normalized to unit mass,

$$\sum_{xyv} w_{xyv}^k = 1, \quad \forall k. \quad (2.3)$$

Sometimes projected quantities extend beyond the boundaries of the storage cube. In this case no weights are added to the grid, but the time spent outside the cube is used in the normalization.

Figure 2.1 illustrates the occupation weights for four orbits in a test model, all with the same energy E but with different angular momenta L/L_{\max} . Because of the spherical symmetry of the system, the weights can be conveniently displayed in the two-dimensional space of projected radius R and line-of-sight velocity v ¹. The difference in the appearance between e.g., the radial orbit at the top and the circular orbit at the bottom is evident in this space of observables.

PSF convolution

It is important for model predictions to incorporate the observational setup and to account for the point-spread-function (PSF) of the data. The final model is a linear superposition of orbits and the PSF convolution of an orbit is also a linear operation. Therefore, the two operations commute and the PSF convolution can be carried out separately for each orbit. The PSF does not correlate velocities² and hence is carried out separately for each velocity slice of the storage cube. Each velocity slice is an ‘iso-velocity’ image of the orbit on a Cartesian grid. Hence, the PSF convolution is most efficiently carried out by a Fast Fourier Transform of these iso-velocity images. As many orbits, e.g., the tightly bound ones, only occupy a small fraction of the spatial grid, the size 2^N of the FFT grid can be adjusted for each orbit and each slice, resulting in a considerable speed-up. The storage-grid must extend a few PSF widths beyond the outermost observational data points. This PSF convolution is most important for studying the dynamics at small radii if steep kinematic gradients are present, e.g., in the application of our technique to the search for massive nuclear black holes (e.g., van der Marel *et al.* 1997, 1998, Cretton *et al.* 1999).

Calculating the velocity moments

After the k -th orbit has been calculated, projected onto the storage cube and convolved with the PSF, we need to extract quantities for direct comparison with the observational constraints. The data contain information on the projected properties of the galaxy at a select number of *constraint positions* on the projected face of the galaxy. Photometry is generally available over the whole face of the galaxy, extending to much larger radii than the kinematic measurements. So in general, there are different constraint positions for the photometric and the kinematic data. In addition, the constraint positions are often extended areas (e.g., the width of a spectroscopic slit multiplied by the number of pixels along the slit that were averaged to obtain spectra of sufficient S/N). Clearly, the storage cube must be chosen sufficiently big to cover all constraint

¹The use of a three-dimensional (x, y, v) storage cube in our technique is motivated by the fact that realistic observational setups generally do not have circular symmetry on the sky. It also makes the generalization to axisymmetric systems simple. In axisymmetric or triaxial potentials the orbit integration proceeds differently, but after projection, the fitting of the observational constraints through orbit superposition is identical.

²The finite spectrograph resolution, i.e., the PSF in the velocity direction, is being accounted for by the kinematic data analysis technique (e.g., Rix & White 1992; vdMF).

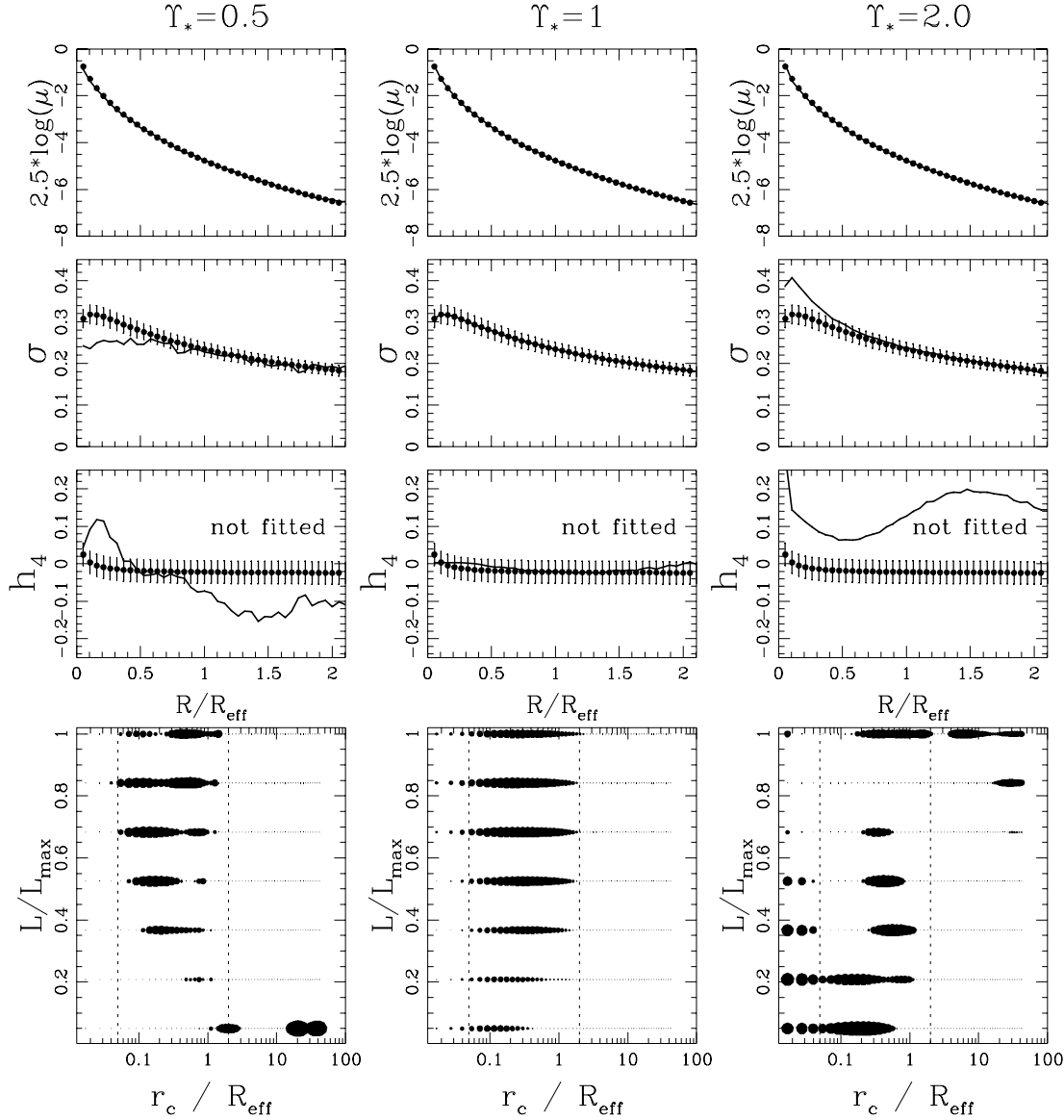


FIGURE 2.2— Fits to pseudo-data (small dots with error bars) drawn from a spherical, isotropic, non-rotating Hernquist model with $\Upsilon_*^{\text{true}} = 1$. In each column we show (from top to bottom) surface brightness μ (in magnitude units), velocity dispersion σ (arbitrary units) and the VP shape coefficient h_4 , all as a function of radius. The solid lines in each panel (in some cases coinciding with the sequence of dots), show the model prediction when only $\mu(R)$ and $\sigma(R)$, but not $h_4(R)$ were fitted. The bottom panel in each column shows the orbital weights of the best fit model in the Lindblad diagram. The area of each dot is proportional to the logarithm of the orbital weight. The axes of the Lindblad diagram are not the conventional (E, L) , but rather $r_c(E)/R_{\text{eff}}$ on the abscissa and $L/L_{\text{max}}(E)$ on the ordinate. For the model in the center column the orbits were calculated in the potential with the correct value of Υ_* . The panels on the left and right show the model predictions if Υ_* is assumed to be 0.5 and 2, respectively.

areas. The constraint positions are labeled by l , with $l = 1, \dots, N_c$. We denote by $f_{xy,l}$ the fraction of the area of the storage cube cell centered on the grid point (x, y) , that is contained within the constraint area l .

Let M_l^k be the fraction of the total mass on orbit k that contributes to constraint area l . This

mass fraction is obtained by summing over the (PSF convolved) storage cube for the given orbit:

$$M_l^k = \sum_{xyv} f_{xy,l} w_{xyv}^k. \quad (2.4)$$

A dynamical model is determined by its *orbital weights* γ_k , which measure the fraction of the total mass of the system that resides on each orbit k (see Appendix 2.A.1 for details). The total mass fraction M_l of the model that contributes to constraint area l is obtained as a sum over all orbits:

$$M_l = \sum_k \gamma_k M_l^k. \quad (2.5)$$

To obtain the observed mass fractions M_l^{obs} at the constraint positions l from the observed surface brightnesses μ_l^{obs} , we assume that the *stellar population* has the same mass-to-light ratio everywhere in the galaxy³. One then has

$$M_l^{\text{obs}} = \mu_l^{\text{obs}} A_l / L_{\text{tot}}, \quad (2.6)$$

where A_l is the area of constraint position l , and L_{tot} is the total observed luminosity. Fitting the predicted mass fractions M_l to the observed mass fractions M_l^{obs} is then a linear superposition problem for the γ_k .

For our technique to work, we must also ensure that the contributions of individual orbits to all kinematic constraints add up linearly. As we show below, this can be achieved in a straightforward manner if we choose the Gauss-Hermite coefficients h_m ($m = 1, \dots, M$) to describe the shape of the VP (vdMF; Gerhard 1993). The normalized VP contributed by orbit k to constraint position l is

$$\text{VP}_{l,v}^k = \frac{1}{M_l^k} \sum_{xy} f_{xy,l} w_{xyv}^k. \quad (2.7)$$

The total normalized VP at constraint position l is obtained as a sum over all orbits:

$$\text{VP}_{l,v} = \frac{1}{M_l} \sum_k \gamma_k M_l^k \text{VP}_{l,v}^k. \quad (2.8)$$

This ‘histogram’, with the velocity v in the subscript on the left-hand-side as the independent variable, is a discrete representation of the underlying continuous profile $\text{VP}_l(v)$. The Gauss-Hermite moment $h_{m,l}$ of order m at constraint position l is defined as an integral over $\text{VP}_l(v)$:

$$h_{m,l} = 2\sqrt{\pi} \int_{-\infty}^{\infty} \text{VP}_l(v) \alpha(w_l) H_m(w_l) dv. \quad (2.9)$$

The function α is a Gaussian weighting function:

$$\alpha(w_l) \equiv \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{1}{2} w_l^2 \right]. \quad (2.10)$$

The quantity w_l is defined as $w_l \equiv (v - V_l)/\sigma_l$, where the velocity V_l and dispersion σ_l are (for the moment) free parameters. The $H_m(w_l)$ are Hermite polynomials (see, e.g., Appendix A of vdMF). One may similarly define the Gauss-Hermite moment $h_{m,l}^k$ of orbit k and order m for

³An independently known radial gradient in the stellar mass-to-light ratio, e.g., from a population analysis, can be included by scaling the ‘photometric constraints’ at the beginning of the analysis.

constraint position l , as an integral over $VP_l^k(v)$ (of which $VP_{l,v}^k$ is the discrete representation). When the free parameters V_l and σ_l are chosen to be the same for each orbit k , it follows that

$$M_l h_{m,l} = \sum_k \gamma_k M_l^k h_{m,l}^k. \quad (2.11)$$

Thus, fitting the observed Gauss-Hermite moments $h_{m,l}^{\text{obs}}$ through the combination $M_l^{\text{obs}} h_{m,l}^{\text{obs}}$ is also a linear superposition problem for the γ_k .

In practice we choose V_l and σ_l equal to the parameters of the best-fitting Gaussian to the observed VP at constraint position l (these are the observationally determined quantities). This implies $h_{1,l}^{\text{obs}} = h_{2,l}^{\text{obs}} = 0$ for the first- and second-order observed Gauss-Hermite moments (vdMF). By requiring the predicted moments $h_{1,l}$ and $h_{2,l}$ to reproduce this, the model VP automatically has the correct mean velocity and velocity dispersion (as determined through a Gaussian fit). Hence, these latter quantities need not be fitted separately. In this procedure we do require knowledge of the errors $\Delta h_{1,l}^{\text{obs}}$ and $\Delta h_{2,l}^{\text{obs}}$ that correspond to the observationally quoted errors ΔV in V_l and $\Delta \sigma$ in σ_l . These can be obtained from the general relations for Gauss-Hermite expansions (vdMF),

$$\Delta h_1 = -\frac{1}{2}\sqrt{2}\Delta V/\sigma, \quad \Delta h_2 = -\frac{1}{2}\sqrt{2}\Delta\sigma/\sigma, \quad (2.12)$$

which are valid to first order in the (small) quantities $(\Delta V/\sigma)$, $(\Delta\sigma/\sigma)$ and h_3, h_4, \dots

The zeroth-order moment h_0 defined by equation (2.9) measures the normalization of the best-fitting Gaussian to the normalized VP. This quantity is *not* included in the fit, because it is observationally inaccessible: it is directly proportional to the unknown difference in line strength between the galaxy spectrum and the template spectrum used to analyze it. In practice one uses the assumption $h_0 = 1$ to estimate the line strength from the observations. The observational estimates for the higher-order Gauss-Hermite moments are also influenced by uncertainties in the line strength, but only to second order. These uncertainties can be safely ignored in all cases of practical interest.

Our scheme uses the Gauss-Hermite *moments* h_m , which are defined as integrals over the VPs. These integrals are well-defined for arbitrary functions, even highly non-Gaussian ones. Our scheme therefore assumes neither that the individual orbital VPs are well described by the lowest order terms of a Gauss-Hermite *series* (which is not generally the case), nor that the observed VPs are well described by the lowest order terms of such a series (which is generally the case). The Gauss-Hermite moments can be viewed as the projections of the VP's onto the set of Gauss-Hermite functions. In this picture, our scheme finds a solution where the weighted sum of projections for the orbital VPs equals the projection of the observed VP within the error. It does not enter anywhere in the scheme that these projections, or moments, have to be small compared to unity for the higher order terms.

2.1.3 Comparison with the observational constraints

Once the properties of all orbits are calculated for all constraint positions, we need to find the non-negative superposition of orbital weights γ_k that best matches the observational constraints within the error bars. When the observational errors are normally distributed, the quality of the fit to the data is determined by the χ^2 statistic:

$$\chi^2 \equiv \sum_{l=1}^{N_p} \left(\frac{M_l^{\text{obs}} - \sum \gamma_k M_l^k}{\Delta M_l^{\text{obs}}} \right)^2 + \sum_{l=N_p+1}^{N_c} \sum_{m=1}^M \left(\frac{M_l^{\text{obs}} h_{m,l}^{\text{obs}} - \sum \gamma_k M_l^k h_{m,l}^k}{\Delta(M_l^{\text{obs}} h_{m,l}^{\text{obs}})} \right)^2, \quad (2.13)$$

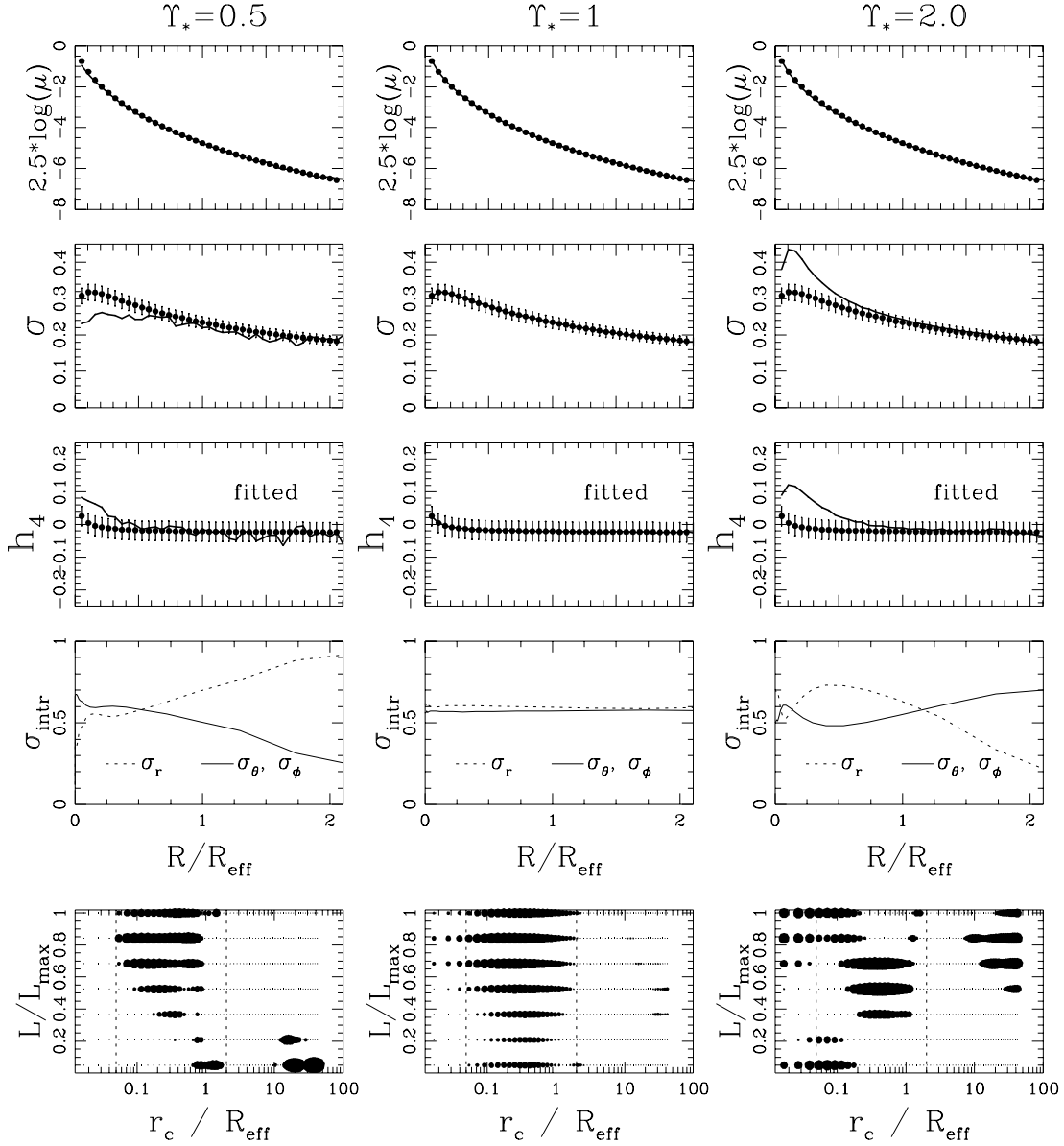


FIGURE 2.3— As Figure 2.2, but now all observational constraints, $\mu(R)$, $\sigma(R)$ and $h_4(R)$ were fitted. We also show the radial variations of the intrinsic velocity dispersions, σ_r and $\sigma_\theta = \sigma_\phi$, in units of the total dispersion $(\sigma_r^2 + \sigma_\theta^2 + \sigma_\phi^2)^{-1/2}$. While the model with the correct Υ_* is very close to isotropic, the other two show anisotropies that vary strongly with radius.

where we assume that there are N_p photometric and $N_k = N_c - N_p$ kinematic constraint positions. For a good model fit to actual data we should expect that the contribution to χ^2 from the photometry and kinematics is $\sim N_p$ and N_k ; in our application to NGC 2434, we find that the contributions to the total χ^2 from the photometry and the kinematics are comparable. Currently, the number M of Gauss-Hermite moments that can be extracted from spectroscopic observations is typically 4. Once even higher order moments can be measured over a wide range of radii, they should be included in the fit. Note that we have chosen to include the photometric errors explicitly, in contrast to other implementations (e.g., Richstone and Tremaine, 1988;

van der Marel *et al.* 1998; Cretton *et al.* 1999). This procedure bypasses any arbitrary relative weighting of the photometry vs. the kinematics. A consequence of this approach is that even models without dark matter components will not be exactly self-consistent for a truly constant mass-to-light ratio. This is not a drawback, as they do represent a self-consistent solution for a model where the mass-to-light varies at the few percent level. Given our scarce knowledge of the *stellar* mass-to-light ratio variations in ellipticals, such models are no less consistent with the observations.

When all quantities are divided by their observational uncertainties, e.g., $M_1^{\text{obs}} \rightarrow M_1^{\text{obs}} / \Delta M_1^{\text{obs}}$, $M_1^1 h_{1,1}^1 \rightarrow M_1^1 h_{1,1}^1 / \Delta(M_1^{\text{obs}} h_{1,1}^{\text{obs}})$, etc., the χ^2 minimization is converted into a least squares problem:

$$\begin{bmatrix} M_1^1 & \dots & \dots & M_1^{N_o} \\ M_2^1 & \dots & \dots & M_2^{N_o} \\ \vdots & \vdots & \vdots & \vdots \\ M_{N_p}^1 & \dots & \dots & M_{N_p}^{N_o} \\ M_{p+1}^1 h_{1,p+1}^1 & \dots & \dots & M_{p+1}^{N_o} h_{1,p+1}^{N_o} \\ \vdots & \vdots & \vdots & \vdots \\ M_{N_c}^1 h_{1,N_c}^1 & \dots & \dots & M_{N_c}^{N_o} h_{1,N_c}^{N_o} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ M_{p+1}^1 h_{M,p+1}^1 & \dots & \dots & M_{p+1}^{N_o} h_{M,p+1}^{N_o} \\ \vdots & \vdots & \vdots & \vdots \\ M_{N_c}^1 h_{M,N_c}^1 & \dots & \dots & M_{N_c}^{N_o} h_{M,N_c}^{N_o} \end{bmatrix} \times \begin{bmatrix} \gamma_1 \\ \vdots \\ \vdots \\ \gamma_{N_o} \end{bmatrix} = \begin{bmatrix} M_1^{\text{obs}} \\ M_2^{\text{obs}} \\ \vdots \\ M_{N_p}^{\text{obs}} \\ M_{p+1}^{\text{obs}} h_{1,p+1}^{\text{obs}} \\ \vdots \\ M_{N_c}^{\text{obs}} h_{1,N_c}^{\text{obs}} \\ \vdots \\ \vdots \\ M_{p+1}^{\text{obs}} h_{M,p+1}^{\text{obs}} \\ \vdots \\ M_{N_c}^{\text{obs}} h_{M,N_c}^{\text{obs}} \end{bmatrix}, \quad (2.14)$$

which must be solved for the occupation vector $(\gamma_1, \dots, \gamma_{N_o})$, with the constraints $\gamma_k \geq 0$, for $k = 1, \dots, N_o$. There are standard algorithms for solving this problem and we use the Non-Negative Least Squares (NNLS) algorithm by Lawson & Hanson (1974; see also Pfenniger 1984; Zhao 1996).

The NNLS fit returns the orbital weights γ_k and the model predictions for all the observed quantities on the right-hand-side of equation (2.14). Among these are the predicted h_1 and h_2 , but not the predicted V and σ . In practice it is often useful to know the latter, for visual comparison to the data. The predicted V and σ can be calculated to first order accuracy from the predicted h_1 and h_2 using the relations (2.12). For higher accuracy one may fit a Gaussian to the actual VPs predicted by the model, which are obtained by substituting the γ_k into equation (2.8).

In practice one considers potentials that depend on a number of parameters. After the (set of) best-fitting parameter combination(s) has been determined, the confidence regions on the model parameters can be estimated from the relative likelihood statistic $\Delta\chi^2 \equiv \chi^2 - \chi_{\min}^2$. If the observational errors are normally distributed, then $\Delta\chi^2$ follows a χ^2 probability distribution, with the number of degrees of freedom equal to the number of parameters in the potential (Press *et al.* 1992). Errors on parameter values quoted in the modeling below correspond to the 68% confidence level, unless mentioned otherwise.

2.1.4 Regularization

If one uses fewer orbits than constraints, the NNLS fit will always have a formally unique solution, even when the underlying physical problem allows a wide range of solutions. An example is the case in which the observations constrain only the projected mass distribution

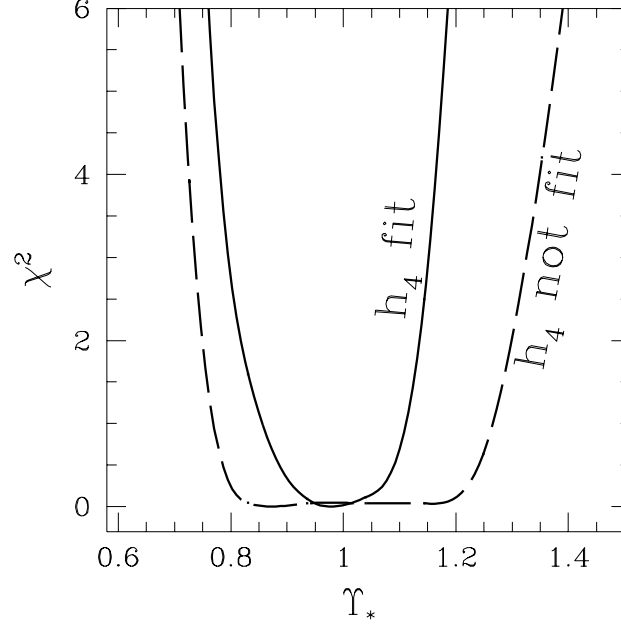


FIGURE 2.4— Illustration of the additional constraints on the gravitational potential provided by VP shape information. The dashed line shows χ^2 for the fit to the Hernquist test model with $\Upsilon_*^{\text{true}} = 1$, when only $\mu(R)$ and $\sigma(R)$ are fitted. A range of potentials, $0.8 < \Upsilon_* < 1.2$, provides a perfect fit to the pseudo-data ($\chi^2 \approx 0$, because no noise was added). The assumption of an incorrect mass-to-light ratio can be fully compensated by a change in the orbital distribution. This is no longer true if the model must also fit $h_4(R)$ (solid curve). In this case the range of potentials that provides a perfect fit is much more narrowly centered on the true mass-to-light ratio.

and velocity dispersion profile (cf. Binney & Mamon 1982). To produce meaningful results with this method, in practice one must therefore use more orbits, or basis vectors of the phase space, than constraints. In this case, the best solution (which need not be an exact solution, i.e., $\chi^2 = 0$) need not be unique. The NNLS fit will select one of the possible solutions. The adopted solution will generally be very irregular in phase-space (trying to accommodate all the noise in the data or the orbit library), which is physically implausible. Non-negativity is the only clear-cut physical constraint on the distribution function. The ‘smoothness’ of the DF in a collisionless system ultimately depends on the efficiency of the violent relaxation during the formation of the galaxy, which is difficult to quantify. Here, we are mostly interested in constraints on the gravitational potential, independent of the detailed properties of the DF, as long as it is non negative. Therefore, we employ only a very simple smoothing procedure, which does not significantly impact the model fit to the data.

Smoothing of the DF can be achieved through regularization (Press *et al.* 1992), but there is no unique approach (e.g., Merritt 1993b). Previous authors have either maximized the entropy of phase space (Richstone & Tremaine 1988), or have enforced local smoothness of the distribution function (Merritt 1993a; Zhao 1996). Here we use a regularization scheme similar to that of Zhao. It minimizes the local curvature of the mass distribution in phase space by including an additional term to the χ^2 -function defined in equation (2.13):

$$\lambda \sum_{k=1}^{N_0} \left(\hat{\gamma}_k - \frac{1}{P} \sum_{p=1}^P \hat{\gamma}_{k_p} \right). \quad (2.15)$$

The P orbits k_1, \dots, k_P are the ‘immediate neighbors’ of orbit k in phase space. For our simple (E_i, L_{ij}) grid there are four neighbors to each point that is not on the edge of the grid. The quantities $\hat{\gamma}_k$ are defined as $\hat{\gamma}_k \equiv \gamma_k / \gamma_k^{\text{ref}}$, where the γ_k^{ref} are a set of reference weights. These could be chosen to reflect any prior knowledge or prejudice about the DF. For example, they may be set to the orbital weights that can be calculated semi-analytically (Appendix 2.A.1) for an isotropic DF, forcing the model to tend to the isotropic DF in the limit of infinite smoothing.

Here we have employed the simplest regularization by setting all the γ_k^{ref} equal to unity. The parameter λ in equation (2.15) governs the degree of regularization. Although this parameter is in principle freely adjustable, it can be chosen in a reasonably objective way, by letting the data determine the degree of permissible DF smoothing. Let χ_0^2 be the minimum chi-squared for the case without regularization ($\lambda = 0$). Any solution that matches the data with $\Delta\chi^2 = \chi^2 - \chi_0^2 \lesssim 1$ is statistically equally acceptable. We can therefore increase λ until $\Delta\chi^2 = 1$, which yields a smoother DF that provides an equally good fit to the data. This regularization procedure is followed in all the subsequent applications in this paper. As Figure 2.5 shows, this minimal regularization often leads to drastically smoother DFs with indistinguishable fits to the data.

Given the simple nature of the regularization employed here, it is important to reiterate that there is a large class of problems for which regularization is not essential. For example, this is the case when the main goal is to rule out certain potentials, e.g., those without a dark halo or without a black hole. If no good fit can be found without regularization, i.e., allowing arbitrarily un-smooth DFs, then there will certainly not exist a smooth DF that fits the data. Thus, by omitting regularization at all, one will always obtain conservative estimates of the range of potentials that are ruled out.

2.2 An illustration of the method

As an illustration and a test, we create pseudo-data drawn from the analytically known properties of an isotropic ($f(E)$) non-rotating Hernquist (1990) model. We assume a mass to light ratio $\Upsilon_\star^{\text{true}} = 1$ for the stellar population (in arbitrary units), and assume that no dark material is present. We match these pseudo-data with our technique, under the assumption of a mass-to-light ratio Υ_\star . As constraints we use the surface brightness over the radial range $0.05 R_{\text{eff}}$ to $2 R_{\text{eff}}$, with an uncertainty of 5%, the line-of-sight velocity dispersion with an error of 5%, and the (non-zero) values of h_4 with an error of ± 0.05 (based on the observational characteristics of, e.g., C95).

We calculated a library of 420 orbits ($N_E = 60$ and $N_L = 7$), with the energy grid ranging from $r_{c,1} = 0.01 R_{\text{eff}}$ to $r_{c,N_E} = 6 R_{\text{eff}}$. Each orbit was built up from 25 sub-orbits (see Section 2.1.1). Only one orbit library was calculated, for $\Upsilon_\star = 1$. The orbit library for any other Υ_\star is obtained trivially by rescaling the model velocity by a factor $\sqrt{\Upsilon_\star}$. However, the orbit contributions $h_{m,l}^k$ to the Gauss-Hermite moments at each constraint position must be calculated separately for each assumed Υ_\star , because they involve the observed velocities V_l and dispersions σ_l in a non-linear way. Similarly, the NNLS fit for the orbital weights γ_k must also be done separately for each Υ_\star .

We constructed models with $\Upsilon_\star = 0.5, 1$ and 2 . This mimics the realistic situation in which the true mass-to-light ratio of a galaxy is unknown, and has to be inferred from models with different Υ_\star . Figure 2.2 shows the match to the pseudo-data if only the surface brightness and the velocity dispersion profiles are fitted. For the correct mass-to-light ratio (middle panels), the fit is perfect and the difference between input and output h_4 is negligible, even though h_4 was not fitted. Note that $\Delta\chi^2 = 1$ regularization has been applied to all models. The bottom panel shows the orbital (mass) weights in a modified Lindblad diagram, where E and L are

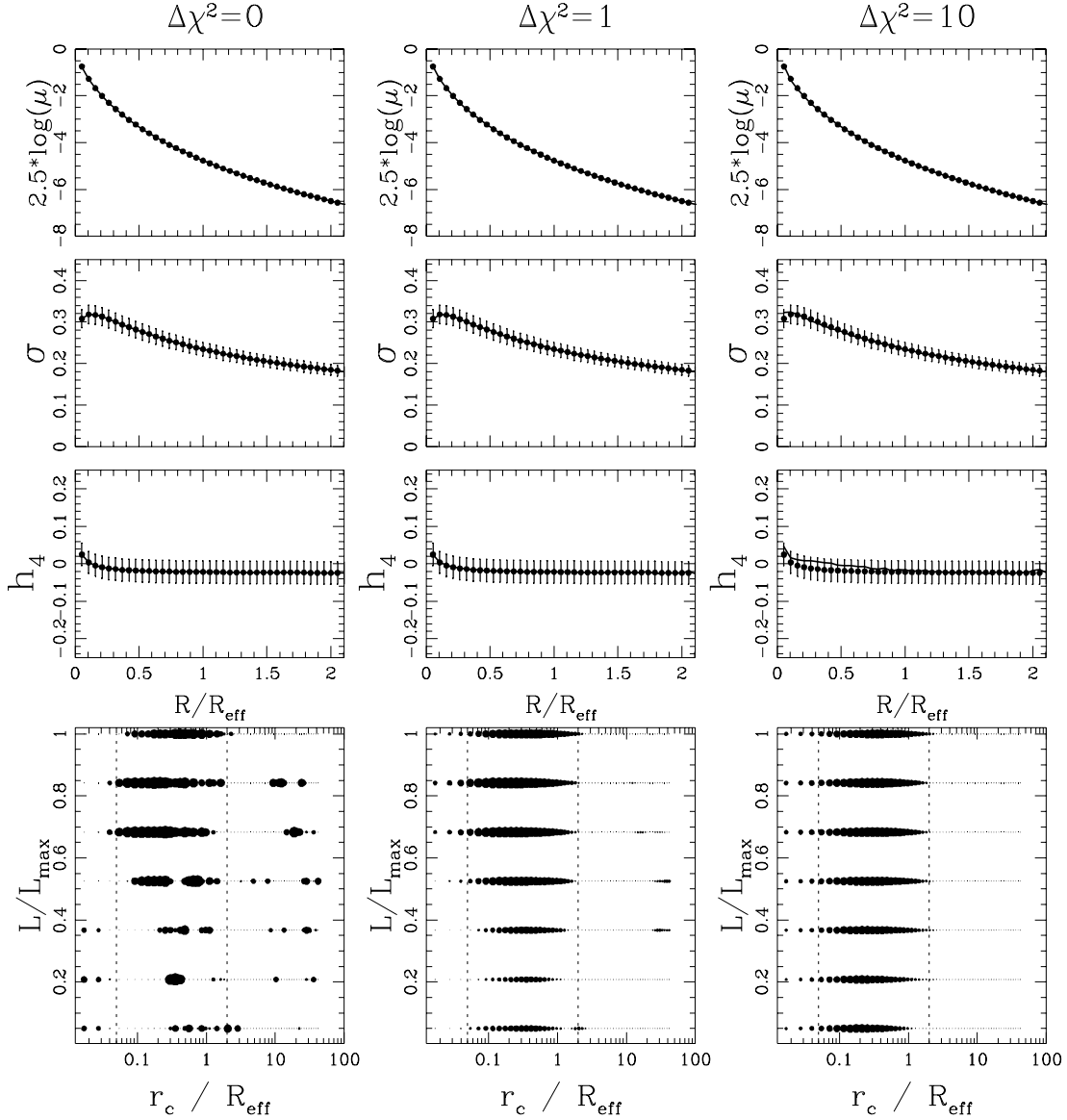


FIGURE 2.5— The effect of regularization on the modeling. All models in this figure have the correct mass-to-light ratio. The left panels show the best fit without any regularization. The phase space distribution is jagged in response to the numerical noise in the projected orbit properties. However, all models that produce a χ^2 that does not differ by more than $\Delta\chi^2 = 1$ are statistically equally acceptable. Hence, we are at liberty to select the smoothest of these models (center column; same as the center-column of Figure 2.3). The right column shows that with excessive regularization the DF is very smooth, but at the expense of a worse fit to the data.

replaced by $r_c(E)/R_{\text{eff}}$ and $L/L_{\text{max}}(E)$, respectively. The area of the dots is proportional to the logarithm of the orbital weight and the vertical dashed lines show the radial range in which observational constraints exist. For $Y_\star = Y_\star^{\text{true}}$ all the mass is attributed to orbits with r_c within the observed range (although of course for very radial orbits, e.g., $L/L_{\text{max}} = 0.05$, the apocenter lies well beyond r_c). The model assigns comparable mass to orbits with the same energy but different eccentricities. This is reassuring, because the pseudo-data were drawn from an isotropic Hernquist model. Note that we have no reason to expect that our best fitting model is precisely

isotropic, because we are constraining a function of two variables, $f(E, L)$, by two functions of one variable, the surface brightness and the velocity dispersion profile. This is insufficient to fix $f(E, L)$ uniquely, but the computations show that the remaining freedom in the DF is not large (see also Dejonghe 1987). We verified that our method does reproduce the isotropic Hernquist model exactly (to within the small discretization errors) when only the $f(E)$ -components are used (see Section 2.1.1 and Appendix 2.A).

The panels on either side show the fit for potentials with assumed mass-to-light ratios of $Y_\star = 0.5$ (left) and $Y_\star = 2$ (right). For $Y_\star = 0.5$, the resulting distribution function consists of two disjoint pieces: a tangentially biased part at $r_c \lesssim R_{\text{eff}}$ and a very radially biased part at $r_c \gtrsim R_{\text{eff}}$. These parts conspire to increase the velocity dispersion in the observed range above the isotropic value. The circular orbits have their highest projected velocity dispersion at $R \approx r_c$ and the radial orbits have their highest projected dispersion at radii $R \ll r_c$. The opposite effect is observed when $Y_\star = 2$: the orbit weights now are given to radially biased orbits at small radii and nearly circular orbits at very large radii. This combination leads to a projected dispersion smaller than the isotropic value. Despite these discrepancies, it is remarkable that over much of the radial range the velocity dispersion can be fit to much better than the factor of $\sqrt{2}$ expected from simple scaling, owing to the freedom to select a special distribution of orbits. However, the predicted h_4 profiles for both $Y_\star = 0.5$ and $Y_\star = 2$ are significantly non-Gaussian, due to the awkward phase-space structure; they differ substantially from the input profile.

Figure 2.3 shows the same orbit library, but forced to fit to surface brightness, the velocity dispersion *and* the h_4 profile. Not surprisingly, the model with $Y_\star = 1$ remains virtually unchanged. For the other two models the match to $h_4(R)$ is improved, at the slight expense of the $\sigma(R)$ fit. In this figure we also show the intrinsic velocity dispersions, σ_r and $\sigma_\theta = \sigma_\phi$. The model with the correct Y_\star is close to isotropic, as expected from the input pseudo-data (perfect isotropy cannot be expected as the constraints do not determine the DF uniquely). The other models have anisotropies that vary strongly with radius, trying to match the data. Note that this type of modeling does permit for a wide range of anisotropies if this is required to fit the data. Figure 2.4 shows the χ^2 of the model fit as a function of Y_\star . The dashed line shows the fit if only the surface brightness and velocity dispersion profiles are used as constraints (as in Figure 2.2). A range of potentials, those with $0.8 \lesssim Y_\star \lesssim 1.2$, all match the data perfectly ($\chi^2 \approx 0$, because no noise was added to the pseudo-data). This is consistent with the results of Binney & Mamon (1982); in each potential the model uses a different orbital structure to fit the data. The solid line shows the χ^2 when the h_4 profile is included in the fit (as in Figure 2.3). The range of Y_\star that produces $\chi^2 \approx 0$ is now much more narrowly centered on the true value, $Y_\star^{\text{true}} = 1$. Many of the potentials that could fit the surface brightness and velocity dispersion profiles cannot simultaneously fit the h_4 profile. The VP shape information constrains the velocity anisotropy, and therefore helps in limiting the set of allowed potentials.

Figure 2.5 shows the effect of regularizing the orbital weight distribution, for the case in which the mass-to-light ratio is correctly assumed to be $Y_\star = 1$. The left panel shows the fit without any regularization: the resulting phase-space distribution is very jagged, and in fact very different from the isotropic model that was used to generate the pseudo-data. However, since all orbital weight distributions that yield fits within $\Delta\chi^2 \sim 1$ are a statistically equally good match to the data, we are at liberty to select the smoothest distribution function amongst those (middle panel). This model is indeed very close to the isotropic model. The right panel shows excessive regularization, resulting in $\Delta\chi^2 = 10$. In the latter case one sees that the fit to the data deteriorates if too smooth an orbital weight distribution is enforced.

2.3 Improved mass modeling for elliptical galaxies

2.3.1 Choice of potentials

The modeling technique yields the relative likelihood of different gravitational potentials, given the observational constraints. However, even for the simplest, spherical case, there is an infinity of trial potentials, $\Phi(r)$. Much of the previous modeling in the literature has focused on testing the constant mass-to-light ratio hypothesis, for which the sequence of trial potentials is one-dimensional, and can be labeled by Υ_* . However, there are two cases of principal interest in which the total mass density is not proportional to the luminous stellar density, namely if (a) there is a massive black hole at the center of a galaxy, or (b) the luminous galaxy is embedded in a dark halo. While in case (a) the set of trial potentials is characterized by two parameters, Υ_* and the black hole mass M_{BH} , case (b) requires a more complex treatment. In the application of our technique below, to study the presence and properties of a dark halo in NGC 2434, we consider three classes of potentials. In each case the goal is to determine whether the given class of potentials is consistent with the data, and for what values of the parameters.

1. Constant mass-to-light ratio models, representing the case without a dark halo (or the case in which the dark and luminous matter have the same spatial distribution). The gravitational potential is derived directly from the deprojected stellar luminosity (and thus mass) distribution, with the mass-to-light ratio Υ_* as the only free parameter.
2. Logarithmic potentials, as a popular case of an ad hoc functional form for the gravitational potential of a galaxy with a dark halo, with the (constant) circular velocity V_c as the only free parameter.
3. Cosmologically motivated ‘star+halo’ models, based on the recent work by NFW. These models use dark matter mass profiles predicted from collisionless cosmological simulations, which are modified by the baryonic/stellar mass accumulating at their center under the assumption of adiabatic invariance. The motivation and construction of these potentials is described in detail in Appendix 2.B. The resulting potentials are characterized by two parameters: the stellar mass-to-light ratio Υ_* and a characteristic scale velocity V_{200} of the dark halo.

2.3.2 An application: The dark matter halo around NGC 2434

We combine the technique discussed in Section 2.1 with the sequence of trial potentials described in Section 2.3.1, to ask what range of gravitational potentials are compatible with the observed photometry and kinematics of NGC 2434. This is a nearly round (E0) elliptical galaxy at an adopted 27 Mpc, with absolute luminosity $M_B = -19.9$. Rotation is unimportant ($E_{\text{rot}}/E_{\text{kin}} \sim 0.01$). The kinematic data (σ and h_4) are from C95, and extend to $60''$; the photometry is from Carollo & Danziger (1994), and extends to $105''$. NGC 2434 is one of the few early-type galaxies where the stellar kinematics, including the shape of the VP, have been measured to $\sim 2.5R_{\text{eff}}$ ($R_{\text{eff}} \sim 24''$).

C95 showed that the kinematics of NGC 2434 could not be fit with axisymmetric constant mass-to-light ratio models with a DF of the form $f = f(E, L_z)$ (if NGC 2434 is seen edge-on and is intrinsically as round as it appears on the sky, these reduce essentially to spherical isotropic $f(E)$ models). From the sign of the discrepancy between their simple models and the data, they also *inferred* that no other model without dark matter would fit the observational constraints. Here we take the analysis two steps further: (i) we *demonstrate* that NGC 2434’s kinematics cannot be fit by any constant mass-to-light ratio model, regardless of the radial anisotropy of the orbital

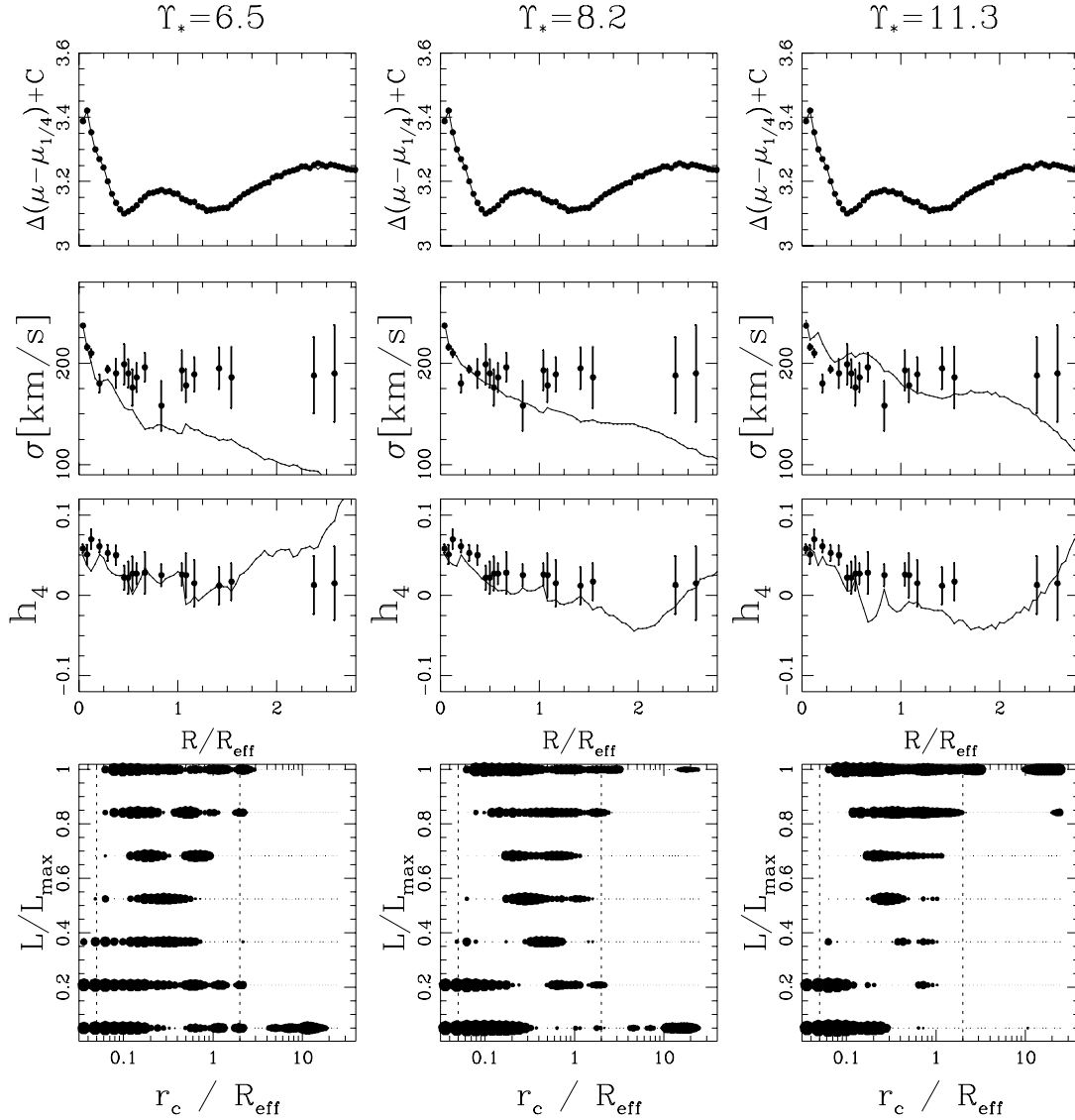


FIGURE 2.6— Predictions of constant mass-to-light ratio models compared to the data for NGC 2434. The top panels show $\Delta(\mu - \mu_{1/4})$ the difference between the surface brightness (in magnitudes) and the surface brightness of the best fitting de Vaucouleurs model. All other quantities in the panels are as in Figure 2.2. Models are shown for, from left to right, $\Upsilon_{*,B} = 6.5, 8.2$ and 11.3 . The middle column provides the best fit, but the observational data cannot be well fit for *any* value of $\Upsilon_{*,B}$. This implies that spherical models with constant mass-to-light ratio are ruled out, independent of the orbital anisotropy of the system.

distribution; and (ii) we explore the issue of *how much* dark matter is required to match the data. It is important to keep in mind for the following analysis, that we restrict ourselves to spherical modeling, and that axisymmetric face-on models have even more freedom to select different orbits.

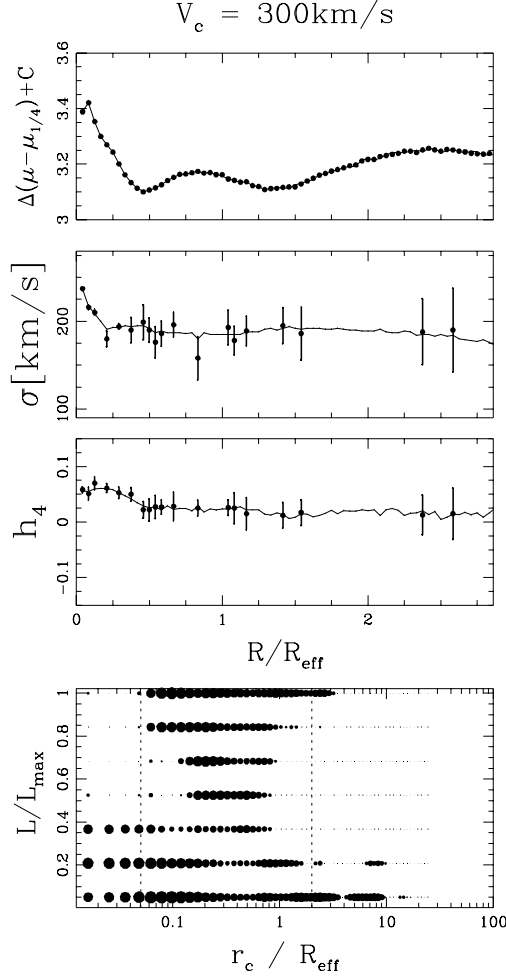


FIGURE 2.7— Predictions of a model with a logarithmic potential with $V_c = 300 \text{ km s}^{-1}$, compared to the data for NGC 2434. The model provides an excellent fit, by contrast to the constant mass-to-light ratio models in Figure 2.6. All observational constraints are matched within the error-bars with a fairly smooth and contiguous distribution of orbital weights. The value of $\chi^2 = 25$ is considerably lower than expected for the fit to 98 data points. It is apparent from the Figure that this is not because the model “fits every noise spike”, but because a portion of the errors in the photometry and kinematics is systematic; i.e. the scatter among neighboring data points is less than the size of their error bars.

Constant mass-to-light ratio models

Following C95, the luminosity density of N2434 can be parameterized as

$$\rho_* \propto (r/b)^\alpha [1 + (r/b)^2]^\beta, \quad (2.16)$$

(with $\alpha = -1.9$, $\beta = -0.52$ and $b = 13''$, see C95). This parametrized form approximates the observed R-band surface brightness (Carollo and Danziger, 1994) to $\sim 4\%$ over the range of interest. For convenience we calculated the luminous gravitational potential from it assuming a constant mass-to-light ratio Υ_* . As a consequence, even the models without dark halos will not have a perfectly constant mass-to-light ratio. However, the few percent differences are no less consistent with the stellar population data than constant $\Upsilon_*(r)$.

As in the test case of Section 2.2, an orbit library was calculated for only one value of Υ_* , which was subsequently scaled to arbitrary values of Υ_* . We used a grid of 60×7 orbits in E and L/L_{\max} , with each orbit ‘built’ from 25 sub-orbits. The energy grid ranged from $r_{c,1} = 0.1''$ to $r_{c,N_E} = 500''$; the corresponding radial spacing is $1.5''$ ($0.3''$ for the sub-orbits) at the scale radius ($b = 13''$), and $5''$ ($1''$) at the last spectroscopic data point. Our technique was used to fit the data (with regularization as discussed in Section 2.1.4), for B-band mass-to-light ratios $\Upsilon_{*,B}$ ranging from 5 to 13 (in solar units). The actual value of $\Upsilon_{*,B}$ was derived assuming a distance of 27 Mpc and a B-R color of 1.7 (Poulain and Nieto, 1994).

Figure 2.6 shows the fits for various values of $\Upsilon_{*,B}$. The best fit is found for $\Upsilon_{*,B} = 8.2 \pm 0.4$, but the observational data cannot be well fit by these models for *any* value of $\Upsilon_{*,B}$. For all values the algorithm invokes a highly anisotropic DF, in which the anisotropy is radically different between the inner and outer parts of the galaxy. However, at large radii none of the models can match a dispersion profile as flat as observed, while maintaining the observed nearly Gaussian VPs. Compared to the best-fitting models presented in Section 2.3.2, the constant mass-to-light ratio models can be ruled out at the $> 99.9\%$ confidence level, independent of the orbital anisotropy of the system.

It is worth noting that this rejection of the constant mass-to-light ratio hypothesis requires the observed constraints on the VP, i.e., the knowledge that the VP is approximately Gaussian $h_4 \ll 1$. If only the surface brightness and the velocity dispersion are used as constraints, an (almost) acceptable model can be found with constant mass-to-light ratio.

Logarithmic potentials

Scale-free logarithmic potentials have $\Phi(r) \equiv V_c^2 \log r$. Such potentials have a constant circular velocity V_c , and have therefore been popular as approximations to the potentials of galaxies with dark halos. Although the stellar kinematics of elliptical galaxies at $R \lesssim R_{\text{eff}}$ can be well fit by models with logarithmic potentials (Kochanek 1994), constant mass-to-light ratio models generally provide equally good fits (van der Marel 1991). However, studies of gravitational lensing statistics (Maoz & Rix 1993) and gravitational lensing observations for individual galaxies (e.g., Kochanek 1995) strongly rule out these constant mass-to-light ratio models, whereas logarithmic potentials do provide good fits. Logarithmic potentials therefore provide the logical next step in our modeling of NGC 2434.

We calculated orbit libraries for logarithmic potentials with V_c ranging from 250 to 370 km s^{-1} , in similar fashion as in Section 2.3.2. A model with $V_c = 300 \pm 15 \text{ km s}^{-1}$ was found to provide an excellent fit to the data, as shown in Figure 2.7. The distribution function is smooth and contiguous, and close to isotropic. The small error bar on V_c illustrates that the addition of VP shape constraints allows the normalization of the potential to be accurately determined, provided that its shape is assumed to be known a priori.

Cosmologically motivated star+halo potentials

So far, we have shown that constant mass-to-light ratio models fail, whereas a model with a logarithmic potential succeeds in fitting the data for NGC 2434. The cosmologically motivated star+halo potentials discussed in Appendix 2.B provide a continuous sequence that connect these two cases. We computed a grid of these potentials, with $\Upsilon_{*,B}$ in the range 2–13 and V_{200} in the range of 0–800 km s^{-1} ($V_{200} = 0 \text{ km s}^{-1}$ corresponds to the constant mass-to-light ratio model). Circular velocity curves for some of the resulting potentials are shown in Figure 2.8. We calculated orbit libraries in these potentials, in similar fashion as in Section 2.3.2. For each $(\Upsilon_{*,B}, V_{200})$ the orbital distribution was found that best reproduced the data, with the

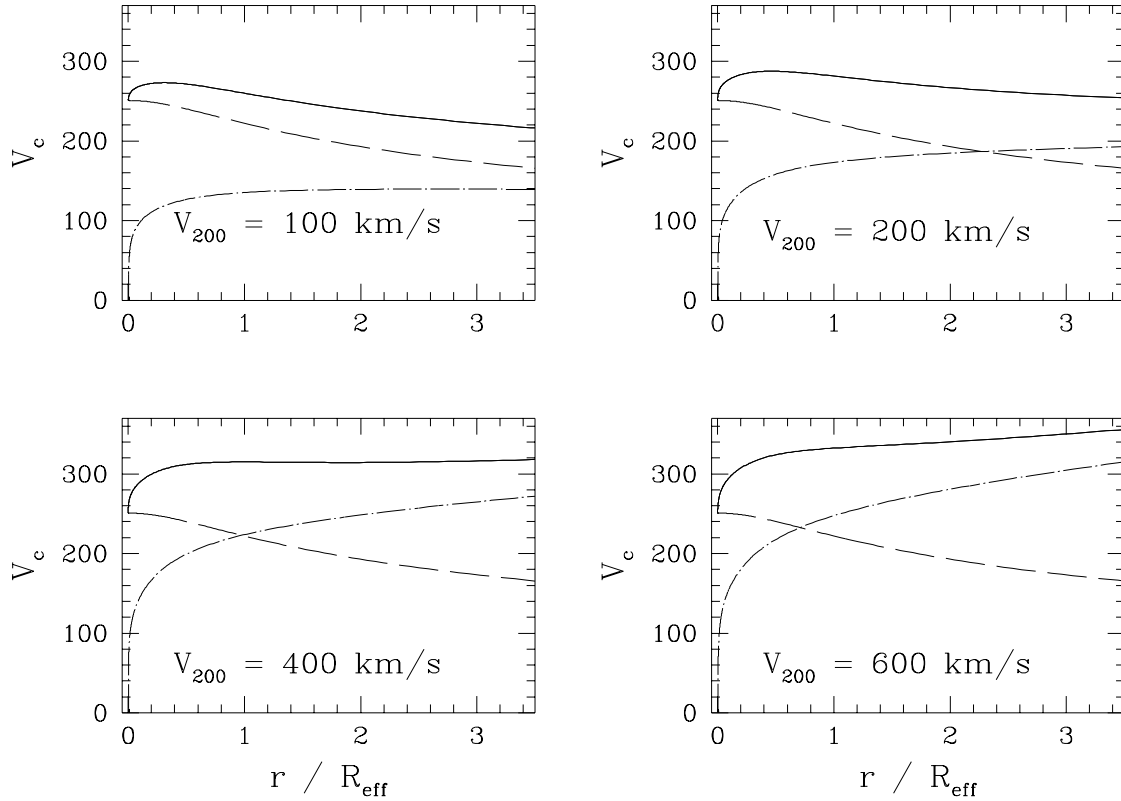


FIGURE 2.8— Rotation curves, $v_c = (R\partial\Phi/\partial R)^{1/2}$, for four of the cosmologically motivated star+halo potentials described in Appendix 2.B (solid curves). In all cases the stellar mass-to-light ratio $\Upsilon_{*,B} = 4.5$, but the characteristic halo velocity, V_{200} , changes from 100 to 600 km s^{-1} . The dashed and dash-dotted curves show the separate contributions from the stars and the dark matter, respectively. The sequence presents a way of consistently and continuously ‘bending up’ the outer part of the rotation curve. The mass distribution for $V_{200} = 400 \text{ km s}^{-1}$ leads to an approximately flat rotation curve from $0.2 < r/R_{\text{eff}} < 3$ and is very similar to the best fitting logarithmic potential (Figure 2.7).

corresponding χ^2 . With these potentials an excellent fit can be obtained to the data. As an example, Figure 2.9 shows a model with $\Upsilon_{*,B} = 4$ and $V_{200} = 400 \text{ km s}^{-1}$. This model, and the other models that fit the data well, have a radially anisotropic stellar velocity distribution ($\sigma_r/\sigma_\theta/\sigma_\phi = 1.4 : 1 : 1$). This is consistent with the results obtained from asymptotic models by de Bruijne, van der Marel, and de Zeeuw (1996).

Figure 2.10 shows the grid of models that were calculated to explore the relative likelihood of models in the $(\Upsilon_{*,B}, V_{200})$ plane. The area of each point corresponds to the logarithm of their relative likelihood. There is a clear anti-correlation between $\Upsilon_{*,B}$ and the mass of the halo (proportional to V_{200}^2). This is because the most robustly constrained quantity is the mass inside a characteristic radius ($\sim R_{\text{eff}}$), which could be either stellar or dark. This anti-correlation is quantified in Figure 2.11, which shows the 68% and 95% confidence regions for the joint distribution of $(\Upsilon_{*,B}, V_{200})$. Figure 2.11 also shows that the best fitting parameters are $\Upsilon_{*,B} = 4.35 \pm 0.35$ and $V_{200} = 450 \pm 100 \text{ km s}^{-1}$. Note that this model has a nearly constant circular velocity of $\sim 310 \text{ km/s}$ over the constrained radial range. The errors quoted refer to the 68% confidence region of each parameter individually.

Interestingly, the stellar mass-to-light ratio $\Upsilon_{*,B}$ for the best fit models is only 50% of the

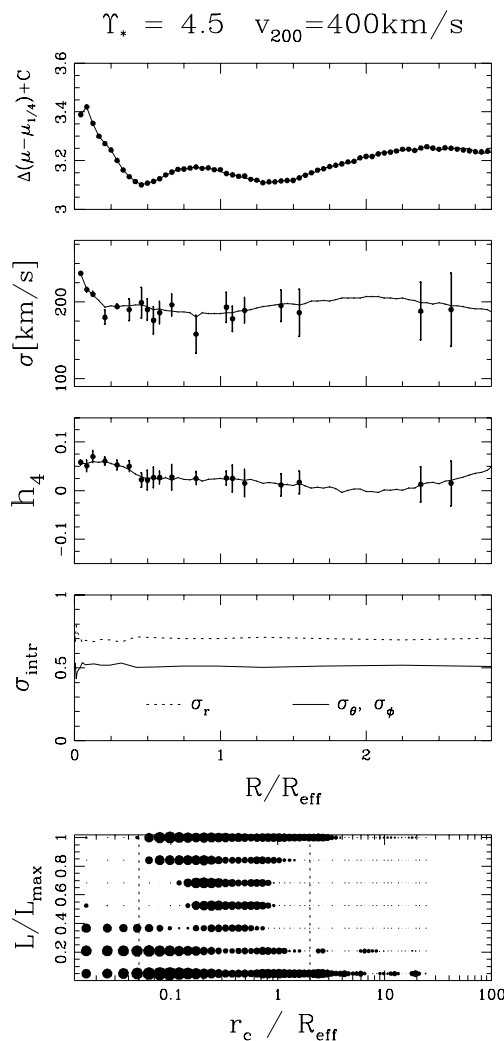


FIGURE 2.9— Predictions of a model with a cosmologically motivated star+halo potential, compared to the data for NGC 2434. The fit is excellent. The model has $\Upsilon_{*,B} = 4.5$ and $V_{200} = 400 \text{ km s}^{-1}$, and was found to provide the best fit among the parameter combinations studied, cf. Figure 2.10. The fit to the data and the orbital structure of this model are very similar to those in Figure 2.7 for the logarithmic potential, as might be expected from the similarity in the two potentials (cf. Figure 2.8). As in Figure 2.3, the fourth panel from the top shows the intrinsic velocity dispersions in units of $\sigma_{\text{tot}} = (\sigma_r^2 + \sigma_\theta^2 + \sigma_\phi^2)^{-1/2}$. The model is radially anisotropic to a nearly constant degree throughout the fit range.

value for the best-fitting constant mass-to-light ratio model shown in Figure 2.6, which is, however, a poor fit. Even with a minimum halo of $V_{200} = 250 \text{ km s}^{-1}$, roughly the smallest value allowed by the data, it is still only 62%. Thus constant mass-to-light ratio models tend to overestimate the true stellar mass-to-light ratio when a dark halo is present. The reason is that these models attempt to reproduce roughly the correct mass within a characteristic radius ($\sim R_{\text{eff}}$), but have no other option than to ascribe this mass to the stars.

The best-fitting models (e.g., that in Figure 2.10) have circular velocity curves that are nearly flat (see Figure 2.8). Their potential resembles that of the best-fitting logarithmic potential in Figure 2.9. As a result, their orbital distribution is also very similar.

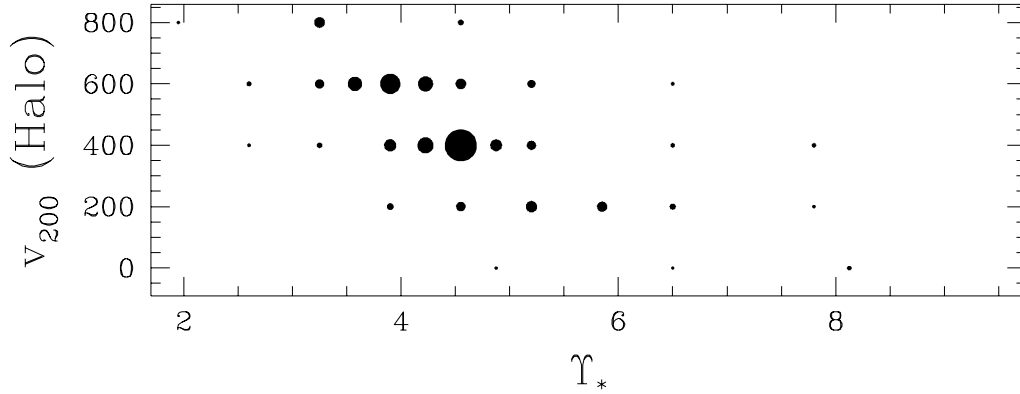


FIGURE 2.10— Relative likelihood of all star+halo models that were calculated in the parameter space of $\Upsilon_{*,B}$ and V_{200} . For each model the superposition of orbits was determined that best fits the NGC 2434 data. The area of each point is proportional to the logarithm of its relative likelihood, which is determined by $\Delta\chi^2$. Models with no dark halo ($V_{200} = 0$) are ruled out. A wide range of potentials with a dark halo can fit the data. There is a strong anti-correlation between $\Upsilon_{*,B}$ and V_{200} (see also Figure 2.11); the kinematics strongly constrain the mass within a given radius, but this mass can either be attributed to the visible or the dark matter.

2.4 Conclusions

We have described an extension of Schwarzschild’s method that is capable of modeling the full line-of-sight velocity distributions of galaxies. Similarly to what is done in the original scheme, we build galaxy models by computing a representative library of orbits in a given potential, and determine the non-negative superposition of these orbits that best fits an observed set of photometric and kinematic constraints. We have extended the technique to predict the full VP shapes for all models, and to include them in the fit to the data as a set of linear constraints. This allows us to fully exploit the high quality kinematical and VP shape data that are now becoming available, and so to constrain the anisotropy of the velocity distribution. The latter has been the main uncertainty in previous attempts to infer the gravitational potentials of elliptical galaxies from observational data. Explicit modeling of VPs also removes systematic errors in the comparison to the lowest-order velocity moments V and σ .

We characterize the VPs through their Gauss-Hermite moments. Measurements of observed VPs are now routinely specified by constraints on these moments. This is useful in practice, because data on observed VPs are often reported as constraints on the Gauss-Hermite moments. It is also computationally convenient, because it reduces VP histograms with 30–50 bins to a small number of parameters. However, our technique is not restricted to the use of Gauss-Hermite moments. In principle one may fit individual velocity bins of observed VPs, if such data is available (as it is, e.g., in the technique of Rix & White 1992). However, this does increase storage and memory requirements, while also increasing the size, and therefore the computational complexity, of the NNLS fit.

We also take into account the error on each observational constraint in finding the best matching orbit superposition. Hence we obtain an objective measure for the quality-of-fit and can determine the relative likelihoods of different models. Only projected, observable quantities are included in the fit. We have also outlined how to include the PSF and the detailed sampling of the data in the data–model comparison by convolving each orbit and sampling it over the observational apertures. This procedure is especially important in the study of galactic

nuclei. We enforce smoothness of the model DFs in phase space through a simple regularization scheme.

The scheme presented in Section 2.1 is valid for any geometry. However, in this first paper, we restricted ourselves to spherical systems which allows swift numerical calculation of the orbit libraries. The orbit superposition scheme was presented in its general form. This allows us to test new elements of our technique in the most straightforward way. We tested our method by constructing models that reproduce pseudo-data drawn from an isotropic Hernquist model, with and without inclusion of VP shape constraints in the fit. The results clearly demonstrate the importance of these constraints in narrowing down the range of potentials that can fit a given velocity dispersion profile when no assumptions are made about the form of the DF.

As an application we have considered the case of dark halos around elliptical galaxies. We used our technique to interpret the stellar kinematical and VP shape data out to $\sim 2.5R_{\text{eff}}$ for the E0 galaxy NGC 2434. Models were constructed with a constant mass-to-light ratio, with logarithmic potentials, and with cosmologically motivated star+halo potentials. The latter are based on the cosmological simulations by NFW, but are modified to incorporate the accumulation of baryonic matter under the assumption of adiabatic invariance. Models without a dark halo are ruled out. Both a logarithmic potential and a cosmologically motivated star+halo potential can provide an excellent fit to the data. The dark halo of NGC 2434 is such that roughly half of the mass within an effective radius is dark. Spherical models without a dark halo therefore tend to overestimate the mass-to-light ratio of the stellar population by a factor of ~ 2 . The best-fitting star+halo potential has a circular velocity curve that is constant ('flat') to within $\sim 10\%$ from $0.2R_{\text{eff}}$ to $3R_{\text{eff}}$. This constant circular velocity is close to that of the best-fitting logarithmic potential, which has $V_c = 300 \pm 15 \text{ km s}^{-1}$.

The restriction to spherical models for NGC 2434 is an important limitation. It is possible that NGC 2434 is a quite flattened system, seen nearly face-on. In this case, the DF has more degrees of freedom, and the potential (and the amount of dark matter) is less constrained. This issue, for NGC 2434 and other galaxies, will be fully explored with the axisymmetric extension of this technique (Cretton *et al.* 1999, van der Marel *et al.* 1998) in forthcoming papers.

If NGC 2434 is intrinsically quite round, the results presented here provide a first test of the conjecture that elliptical galaxies and spirals of the same stellar mass started out in similar dark matter halos. Obviously, the two types of galaxies differ in the degree to which the baryons were concentrated at their centers (with ellipticals being much denser). If the 'baryon contraction' led a flat rotation curve in spirals, it should have led to a centrally peaked (and hence outward falling) rotation curve in ellipticals. The nearly constant circular velocity in NGC 2434 is, however, inconsistent with this idea.

The extension to axisymmetric systems is described in detail in a forthcoming paper (Cretton *et al.* 1999). The application to the galaxy M32, to investigate the presence of a massive central black hole, is discussed in van der Marel *et al.* (1997, 1998). A further extension to tumbling triaxial systems is in progress. Another possible extension is to include radial variations in the stellar population. This can be achieved by choosing a radially varying Υ_* , or by including the color and line-strength gradients in the set of constraints, while allowing each orbit to be occupied by stars with different physical properties.

Acknowledgments

We dedicate this paper, and its companions van der Marel *et al.* (1998) and Cretton *et al.* (1999), to the memory of Martin Schwarzschild, who pioneered the modeling technique employed here. Martin's sense of purpose, his exceptional clarity of thinking, his transparent personal integrity

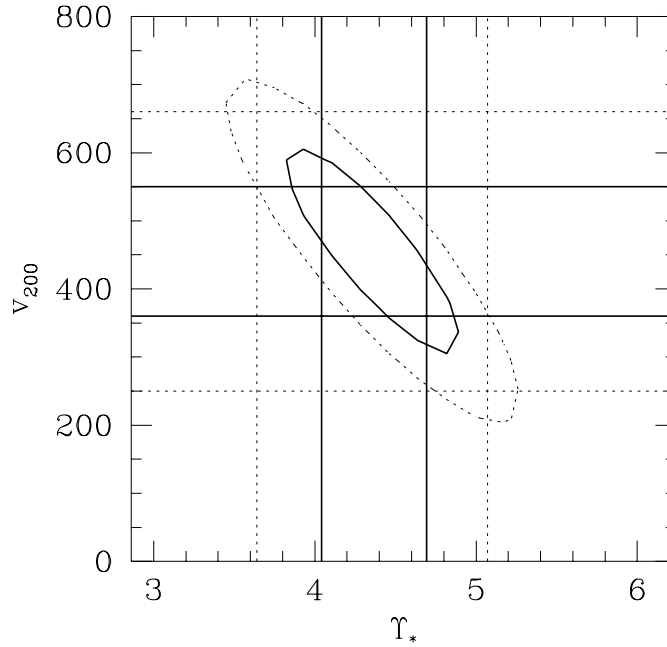


FIGURE 2.11— Confidence limits in the $\Upsilon_{*,B}$ and V_{200} plane for NGC 2434. The solid ellipse encloses the 1σ (68%) confidence region, the dotted contour encloses the 2σ (95.4%) confidence region, for the joint distribution of both parameters. The solid and dotted lines denote the 1σ and 2σ limits on the parameters individually: $\Upsilon_{*,B} = 4.35 \pm 0.35$ and $V_{200} = 450 \pm 100 \text{ km s}^{-1}$.

and, most of all, his genuine warm interest and support, remain a great source of inspiration to us. We would like to thank HongSheng Zhao for a critical reading of the manuscript. NC acknowledges financial support from NUFFIC and from the Leids Kerkhoven Bosscha Fonds, and the hospitality of Steward Observatory. CMC and RPvdM were supported by NASA through Hubble Fellowships, #HF-1079.01-96A and #HF-1065.01-94A, respectively, awarded by the Space Telescope Science Institute which is operated by the Association of Universities for Research in Astronomy, Incorporated, under NASA contract NAS5-26555.

References

- Ashman, K. M. 1992, *PASP*, 104, 1109
 Awaki, H. et al. 1994, *PASJ*, 46, L65
 Bertin, G., & Stiavelli, M. 1993, *Rep. Progr. Phys.*, 56, 493
 Bertin, G. et al. 1994, *A&A*, 292, 381
 Binney, J. J., & Mamon, G. A. 1982, *MNRAS*, 200, 361
 Binney, J. J., & Tremaine, S. D. 1987, *Galactic Dynamics* (Princeton: Princeton University Press)
 Blumenthal, G., Faber, S., Flores, R., & Primack, J. 1986, *ApJ*, 301, 27
 Carollo, C. M., & Danziger, I. J. 1994, *MNRAS*, 270, 523
 Carollo, C. M., de Zeeuw, P. T., van der Marel, R. P., Danziger, I. J., & Qian, E. E. 1995, *ApJ*, 441, L25 (C95)
 Ciardullo, R., Jacoby, G. H., & Dejonghe, H. B. 1993, *ApJ*, 414, 454
 Cole, S., & Lacey, C. 1996, *MNRAS*, 281, 716
 Cretton, N., de Zeeuw, P. T., van der Marel, R. P. & Rix, H.-W., 1999, *ApJ Supplements*, in press, (astro-ph/9902034), chapter 3 of this thesis

- de Bruijne, J. H. J., van der Marel, R. P. & de Zeeuw, P. T., 1996, *MNRAS*, 282, 909
- Dejonghe, H. B. 1987, *MNRAS*, 224, 13
- Dejonghe, H. B. 1989, *ApJ*, 343, 113
- Dejonghe, H. B., & Merritt, D. 1992, *ApJ*, 391, 531
- de Zeeuw, & P. T., Franx, M. 1991, *ARA&A* 29, 239
- de Zeeuw, P. T. 1995, *IAU Symposium* 164, *Stellar Populations*, eds G. Gilmore, P. C. van der Kruit (Dordrecht: Kluwer), p. 215
- de Zeeuw, P. T. 1996, in *Gravitational Dynamics*, eds. O. Lahav, E. Terlevich & R. Terlevich (Cambridge: Cambridge Univ. Press), p. 1
- Forman, C., Jones, C., & Tucker, W. 1985, *ApJ*, 293, 102
- Franx, M., van Gorkom, J., & de Zeeuw, P. T. 1994, *ApJ*, 436, 642
- Gerhard, O. E. 1993, *MNRAS*, 265, 213
- Hernquist, L. 1990, *ApJ* 356, 359
- Katz, N., & Richstone, D. O., 1985, *ApJ*, 296, 331
- Kochanek, C. 1994, *ApJ*, 436, 56
- Kochanek, C. 1995, *ApJ*, 445, 559
- Kormendy, J., & Richstone, D. 1995, *ARA&A*, 33, 581
- Kuijken, K. and Merrifield, M. R., 1993, *MNRAS*, 264, 712
- Lawson, C. L., & Hanson, R. J. 1974, *Solving Least Squares Problems* (Englewood Cliffs, New Jersey: Prentice-Hall)
- Levison, H. F., & Richstone, D. O. 1985, *ApJ*, 295, 349
- Maoz, D., & Rix, H.-W. 1993, *ApJ*, 416, 425
- Merritt, D. 1993a, *ApJ*, 413, 79
- Merritt, D. 1993b, in *Structure, Dynamics and Chemical Evolution of Elliptical Galaxies*, eds. I. J. Danziger, W. W. Zeilinger, & K. Kjär (Münich: ESO), p. 275
- Merritt, D., & Saha, P. 1993c, *ApJ*, 409, 75
- Merritt, D., & Oh, S. P. 1997, *AJ*, 113, 1279
- Navarro, J., Frenk, C., & White, S. D. M. 1996, *ApJ*, 462, 563 (NFW)
- Pfenniger, D. 1984, *A&A*, 141, 171
- Poulain, P. & Nieto, J.-L., 1994, *A&AS*, 103, 573
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, *Numerical Recipes* (Cambridge: Cambridge University Press)
- Richstone, D. O., 1980, *ApJ*, 238, 103
- Richstone, D. O., 1984, *ApJ*, 281, 100
- Richstone, D. O., & Tremaine, S. D. 1984, *ApJ*, 286, 27
- Richstone, D. O., & Tremaine, S. D. 1985, *ApJ*, 296, 370
- Richstone, D. O., & Tremaine, S. D. 1988, *ApJ*, 327, 82
- Rix, H.-W., 1996a, in *Proceedings of the DM 1996 Workshop at Sesto*, eds. P. Salucci and M. Seric
- Rix, H.-W., 1996b, in *The Nature of Elliptical Galaxies, Proceedings of the Second Stromlo Symposium*, eds M. Arnaboldi, G. da Costa and P. Saha., p. 65
- Rix, H.-W., & White, S. D. M. 1992, *MNRAS*, 254, 389
- Saglia, R. P., Bertin, G., & Stiavelli, M. 1992, *ApJ*, 384, 433
- Saglia, R. P. et al. 1993, *ApJ*, 403, 567
- Saglia, R. P. 1996, in *New Light on Galaxy Evolution, Proc. IAU Symp. 171*, eds. R. Bender, R. L. Davies (Dordrecht: Kluwer Academic Publishers), p. 157
- Schwarzschild, M. 1979, *ApJ*, 232, 236
- Statler, T., Smecker-Hane, T., & Cecil, G. 1996, *AJ*, 111, 1512

Tremblay, B., Merritt, D., & Williams, T. B., 1995, *ApJ*, 443, L5
 van Albada, T. S., Bahcall, J. N., Begeman, K., & Sancisi, R. 1985, *ApJ*, 295, 305
 van der Marel, R. P. 1991, *MNRAS*, 253, 710
 van der Marel, R. P., & Franx, M. 1993, *ApJ*, 407, 525 (vdMF)
 van der Marel, R. P., de Zeeuw, P. T., Rix, H.-W., & Quinlan, G. D. 1997, *Nature*, 385, 610
 van der Marel, R. P., Cretton, N., de Zeeuw, P. T., & Rix, H.-W. 1998, *ApJ*, 493, 613, chapter 4 of this thesis
 Vandervoort, P. O. 1984, *ApJ*, 287, 475
 Zhao, H. S. 1996, *MNRAS*, 283, 149

2.A The differential mass density

We describe here the connection between the orbital weights used in the Schwarzschild technique and the DF. Our treatment largely follows that of Vandervoort (1984).

2.A.1 Inferring the distribution function from the orbital weights

The general DF for a spherical system is $f(E, L)$, where the binding energy E and angular momentum L per unit mass are

$$E = \Psi(r) - \frac{1}{2}v_r^2 - \frac{L^2}{2r^2}, \quad L = rv_t, \quad (2.17)$$

and the positive gravitational potential is $\Psi \equiv -\Phi$. A solution of the Schwarzschild technique is not a direct approximation to $f(E, L)$, but to the differential mass density in the Lindblad diagram, $\gamma(E, L) \equiv dM / dE dL$. This function is fully determined by the DF and the gravitational potential.

The total mass of the system is the integral of the DF over phase-space

$$M = \int d^3\mathbf{r} d^3\mathbf{v} f(E, L) = \int 4\pi r^2 dr \int d\xi \int v_t dv_t \int dv_r f(E, L), \quad (2.18)$$

where the angle ξ is the direction of the tangential velocity vector, as defined in Section 2.1.2. The integral over ξ is trivial (since neither E nor L depends on it), so M results from a three-dimensional integral over the phase-space coordinates (r, v_t, v_r) . One may change the order of the integrations and change to the integration variables (r, L, E) to obtain

$$M = 8\pi^2 \int_{-\infty}^{\Psi(0)} dE \int_0^{L_{\max}(E)} L f(E, L) T_r(E, L) dL. \quad (2.19)$$

The radial period $T_r(E, L)$ of the orbit with integrals (E, L) is defined as

$$T_r(E, L) \equiv 2 \int_{r_{\min}(E, L)}^{r_{\max}(E, L)} \frac{dr}{v_r} = 2 \int_{r_{\min}(E, L)}^{r_{\max}(E, L)} \frac{dr}{[2(\Psi(r) - E) - (L^2/r^2)]^{1/2}}, \quad (2.20)$$

and can be easily calculated numerically for any potential. Equation (2.19) shows that the differential mass density is:

$$\gamma(E, L) = 8\pi^2 L f(E, L) T_r(E, L). \quad (2.21)$$

The Schwarzschild technique characterizes a system by the orbital weights (mass fractions) γ_k for a discrete set of orbits. Each orbit k corresponds to a combination (E_i, L_{ij}) of the integrals

of motion. Let the corresponding grid cell in (E, L) space have an area A_{ij} . The orbital weights γ_k are then related to the DF according to

$$\gamma_k \approx 8\pi^2 L_{ij} f(E_i, L_{ij}) T_r(E_i, L_{ij}) A_{ij} / M. \quad (2.22)$$

This relation allows the DF to be estimated from a set of γ_k inferred by the technique (or, it allows the γ_k to be predicted for a test model with a known DF).

2.A.2 Construction of $f(E)$ -components

It proved useful for testing our technique to use building blocks that are more complex than individual orbits. In particular, we constructed building blocks that correspond to a weighted sum of orbits with the same energy. The orbital occupancies W_{xyv} of these components are related to the orbital occupancies w_{xyv} for the individual orbits, defined in Section 2.1.2, through:

$$W_{xyv}^i = \sum_{j=1}^{N_L} \epsilon_j w_{xyv}^{ij} \quad (2.23)$$

(we denote here each orbit by the indices i and j of its position on the (E, L) grid, rather than by its index k). The weights ϵ_j are such that each component corresponds to an isotropic DF, restricted to one energy E_i . This requires that

$$\epsilon_j = L_{ij} T_r(E_i, L_{ij}) A_{ij} / \left[\sum_{j=1}^{N_L} L_{ij} T_r(E_i, L_{ij}) A_{ij} \right]. \quad (2.24)$$

This follows from equation (2.22) (which gives the relative masses contained on orbits with the same energy but different angular momenta) and from the fact that the ϵ_j must add to unity (so that the orbital occupancies W_{xyv}^i are normalized). We call the resulting building blocks ' $f(E)$ -components'. Any combination of these components yields an isotropic DF.

One may alternatively derive the construction of the $f(E)$ -components from an analysis based on delta-function DFs. Let $f_{[E_0]}^\delta$ represent a delta-function in the energy, $f_{[E_0]}^\delta(E) \equiv A_0 \delta(E - E_0)$, and let $f_{[E_0, L_0]}^\delta$ represent a delta-function in both energy and angular momentum, $f_{[E_0, L_0]}^\delta(E, L) \equiv B_0 \delta(E - E_0, L - L_0)$. These DFs are normalized to unit mass if

$$A_0 = \left[8\pi^2 \int L T_r(E_0, L) dL \right]^{-1}, \quad B_0 = \left[8\pi^2 L_0 T_r(E_0, L_0) \right]^{-1}, \quad (2.25)$$

cf. equation (2.19). Therefore,

$$f_{[E_0]}^\delta(E) = \int f_{[E_0, L_0]}^\delta(E, L) \frac{L_0 T_r(E_0, L_0)}{\int L' T_r(E_0, L') dL'} dL_0, \quad (2.26)$$

which shows how $f_{[E_0]}^\delta$ is build up as a weighted integral over the $f_{[E_0, L_0]}^\delta$. The weights provide the continuous analog of equation (2.24). The projected properties for the delta-function DF $f_{[E_0, L_0]}^\delta$ can be traced in analytical fashion (see also Merritt 1993a). For example, the VP can be evaluated through a 1D quadrature over the line of sight:

$$\begin{aligned} \text{VP}(x, y, v) &= \int f_{[E_0, L_0]}^\delta(E, L) dv_x dv_y dz \\ &= \int f_{[E_0, L_0]}^\delta(E, L) J(E, L) dE dL dz = B_0 \int J(E_0, L_0) dz, \end{aligned} \quad (2.27)$$

where $J(E, L)$ is the Jacobian for the change of integration variables, from (v_x, v_y) to (E, L) :

$$J = \left[2v_y x (xv_x + zv) \right]^{-1}, \quad (2.28)$$

with

$$v_y = \pm \sqrt{2(\Psi - E) - v^2 - v_x^2}, \quad v_x = -\frac{1}{x} \left[zv \pm \sqrt{2(\Psi - E)(z^2 + x^2) - L^2} \right]. \quad (2.29)$$

Finally isotropic building blocks $f(E_i)$ can be constructed as a weighted sum of all the orbits of the same energy E_i :

$$f(E_i) = \int f(E_i, L_i) \cdot \frac{T_r(E_i, L) L}{\int T_r(E_i, L) L dL} dL. \quad (2.30)$$

2.B Cosmologically motivated star+halo models

To model the dynamics of galaxies with dark matter halos one must choose a dark potential to add to the luminous one. The traditional approach, used widely when fitting rotation curves of spiral galaxies, has been to describe the dark matter as an isothermal sphere with asymptotic circular velocity V_c and core radius r_h (e.g., van Albada *et al.* 1985). The main drawback of this approach is that it is without physical basis: it is implausible to expect a dark matter halo with a constant density core after the baryonic mass has condensed and concentrated at its center. It also has disadvantages from a practical point of view. The dark halo is described by two new parameters, which add to the unknown mass-to-light ratio Y_* of the stellar population. This makes the sequence of potentials that must be compared to the data three-dimensional, which makes a proper comparison to data for elliptical galaxies very time consuming. Apart from this, V_c and r_h are highly correlated and cannot even be determined independently from observed rotation curves for most spiral galaxies.

Suggestive alternatives come from cosmological studies. NFW and Cole & Lacey (1996) have found that in numerical simulations of many cosmogonies, including standard CDM, the spherically averaged density profiles of the forming virialized halos can be described by a simple functional form over a wide range of radii:

$$\rho(r) = \frac{\rho_{\text{crit}} \delta_c}{(r/r_s) [1 + (r/r_s)]^2}, \quad (2.31)$$

where $\rho_{\text{crit}} = (3H_0^2)/(8\pi G)$ is the critical density of the universe. The two scale parameters δ_c and r_s are defined in terms of a dimensionless concentration parameter c , through

$$\delta_c = \frac{200}{3} c^3 / \left[\ln(1+c) - c/(1+c) \right], \quad r_s = r_{200}/c, \quad (2.32)$$

where r_{200} is defined as the radius within which the mean halo density is $200\rho_{\text{crit}}$. Each halo may also be described by a mass scale $M_{200} \equiv 200\rho_{\text{crit}}(4\pi/3)r_{200}^3$ and a velocity $V_{200} \equiv \sqrt{GM_{200}/r_{200}}$. Mathematically, any two of these parameters are sufficient to describe the halo's structure, e.g., V_{200} and c . Remarkably, NFW found in addition that for a given cosmology, the two parameters are tightly correlated (e.g., with $\sim 15\%$ scatter in c at a given V_{200}). For standard CDM this relation can be described by

$$\log_{10} c = 1.05 - 0.15 [M_{200}/(3 \times 10^{13} M_\odot)]. \quad (2.33)$$

Thus, these N-body simulations essentially suggest a cosmologically motivated, one-parameter sequence of dark matter halo models.

The NFW simulations do not contain dissipative, baryonic matter. In reality, such matter collects at the center of the potential well to form the visible part of the galaxy. Blumenthal *et al.* (1986) suggested that adiabatic invariance can be exploited to estimate how the halo structure is modified by the accumulation of luminous matter at the center. This was confirmed by N-body simulations with a crude model for the infall of the dissipative matter. In general, adiabatic invariance holds (approximately) for those orbits with periods smaller than the characteristic time scale of the changes in the potential (Binney & Tremaine 1987). An elliptical galaxy has a dynamical time of a few times 10^7 years at the effective radius, so adiabatic invariance may be a good approximation if the baryonic infall occurs over a time scale exceeding 10^{7-8} years.

If the radial mass profiles of baryons and dark matter were initially the same, then adiabatic invariance in the simplest approximation (Blumenthal *et al.* 1986) relates the initial and final radii of the mass shells r_i and r_f , respectively, by

$$r_f \left[M_*(r_f) + M_{\text{DM}}(r_f) \right] = r_i M_i(r_i). \quad (2.34)$$

Here $M_i(r_i)$ is the initial total mass within radius r_i , $M_*(r_f)$ is the final stellar mass within radius r_f , and $M_{\text{DM}}(r_f)$ is the final dark mass with radius r_f . We assume the initial profile $M_i(r_i)$ to be as given by NFW for the case of standard CDM ($\Omega = 1$), with the characteristic velocity V_{200} as free parameter. We then employ adiabatic contraction to obtain the final dark matter profile $M_{\text{DM}}(r_f)$ from the observed stellar mass profile $M_*(r_f)$, assuming a stellar mass-to-light ratio Y_* . This yields a sequence of star+halo models that are characterized by the parameters Y_* and V_{200} . This procedure does not account for mergers, which will decrease the coarse-grained phase-space density and the physical density of both the luminous and dark matter; how merging would change the ratio of luminous to dark matter is not clear. In this sense the present procedure, assuming $\Omega = 1$ and neglecting late mergers, leads to the densest possible halo *model* potentials, at a given V_{200} . Viewed differently, other scenarios will require halos of higher V_{200} to have the same amount of mass within a given radius that is accesible to observations (say, $2 R_{\text{eff}}$).

Circular velocity curves for some of the resulting potentials are shown in Figure 2.8. Several features are noteworthy: (i) the transition from the luminous to the dark matter dominated regime is always smooth, and is not marked by a feature in the circular velocity curve; (ii) because the condensing baryons drag some of the dark matter with them to the center, the dark matter fraction decreases slower towards small radii than it does for constant density core halos; and (iii) compact halos are required in order to produce an ‘effectively flat’ circular velocity curve.

Chapter 3

Axisymmetric Three-Integral Models for Galaxies

Cretton, N., de Zeeuw, P. T., van der Marel R. P., Rix, H.-W.
1999, *ApJ Supplement Series*, in press

We describe an improved, practical method for constructing galaxy models that match an arbitrary set of observational constraints, without prior assumptions about the phase-space distribution function (DF). Our method is an extension of Schwarzschild's orbit superposition technique. As in Schwarzschild's original implementation, we compute a representative library of orbits in a given potential. We then project each orbit onto the space of observables, consisting of position on the sky and line-of-sight velocity, while properly taking into account seeing convolution and pixel binning. We find the combination of orbits that produces a dynamical model that best fits the observed photometry and kinematics of the galaxy. A new element of this work is the ability to predict and match to the data the full line-of-sight velocity profile shapes. A dark component (such as a black hole and/or a dark halo) can easily be included in the models.

In an earlier paper (Rix *et al.*, chapter 2 of this thesis) we described the basic principles, and implemented them for the simplest case of spherical geometry. Here we focus on the axisymmetric case. We first show how to build galaxy models from individual orbits. This provides a method to build models with fully general DFs, without the need for analytic integrals of motion. We then discuss a set of alternative building blocks, the two-integral and the isotropic components, for which the observable properties can be computed analytically. Models built entirely from the two-integral components yield DFs of the form $f(E, L_z)$, which depend only on the energy E and angular momentum L_z . This provides a new method to construct such models. The smoothness of the two-integral and isotropic components also makes them convenient to use in conjunction with the regular orbits.

We have tested our method, by using it to reconstruct the properties of a two-integral model built with independent software. The test model is reproduced satisfactorily, either with the regular orbits, or with the two-integral components. This paper mainly deals with the technical aspects of the method, while applications to the galaxies M32 and NGC 4342 are described elsewhere (van der Marel *et al.* and Cretton & van den Bosch; respectively chapter 4 and 5 of this thesis).

IN order to understand the structure and dynamics of a galaxy, one needs to measure the total gravitational potential as well as the phase-space distribution function (DF) of the constituent stars. The DF specifies the distribution of the stars over position and velocity, and hence provides a full description of the galaxy. For a particular galaxy, one needs to explore which combinations of potential and DF are consistent with the available observations (surface brightness and kinematics). Several methods have been devised to tackle this problem.

The direct calculation of the DF generally requires analytic knowledge of the integrals of motion, and has been restricted in the past to a number of special cases: (i) spherical or other integrable potentials (e.g., Dejonghe 1984, 1986; Bishop 1987; Dejonghe & de Zeeuw 1988; Gerhard 1991; Hunter & de Zeeuw 1992); (ii) nearly integrable systems where perturbation theory can be applied (Saaf 1968; Dehnen & Gerhard 1993); or (iii) the subset of axisymmetric models in which the DF is *assumed* to depend only on E and L_z (Hunter & Qian 1993; Dehnen & Gerhard 1994; Kuijken 1995; Qian *et al.* 1995, hereafter Q95; Magorrian 1995; Merritt 1996b, hereafter M96b).

Numerical calculations of orbits in axisymmetric potentials have shown that most of the orbits admit a third integral, which in general is not known analytically (e.g., Ollongren 1962). There is no a priori physical reason to expect the DF to depend only on the two classical integrals, and in fact, there are indications for both elliptical galaxies (Binney, Davies & Illingworth 1990) and halos of spirals (Morrison, Flynn & Freeman 1990) that the DF must depend also on the third integral. For the solar neighborhood it has been known for a long time that there must be such a dependence (e.g., Binney & Merrifield 1998).

Schwarzschild (1979, 1982) devised an elegant method to circumvent our ignorance of analytic integrals of motion and to build numerically self-consistent equilibrium models of galaxies. Richstone (1980, 1984) used this technique to construct axisymmetric scale-free models. It was applied to a variety of models (spherical, axisymmetric and triaxial) by Richstone and collaborators (see e.g., Richstone & Tremaine 1984, 1985; Levison & Richstone 1985, 1987; Katz & Richstone 1985). Pfenniger (1984) used Schwarzschild's method to build two-dimensional models of barred galaxies and Merritt & Fridman (1996) and Merritt (1996a) used it to build a number of triaxial models with cusps. Zhao (1996b) modeled the Galactic bar using similar techniques. Schwarzschild's original experiment reproduced self-consistently a triaxial mass distribution, but as shown by Pfenniger (1984), one can easily include kinematic constraints in the models. Levison & Richstone (1985) modeled the observed mean line-of-sight velocities V and velocity dispersions σ to estimate the amount of counter-rotation in some well-observed galaxies.

Recent advances in detector technology have made it possible to measure full line-of-sight velocity profile (VP) shapes, instead of only the first two moments V and σ (e.g., Franx & Illingworth 1988; Rix & White 1992; van der Marel & Franx 1993, hereafter vdMF; Kuijken & Merrifield 1993). This provides further constraints on the dynamical structure of galaxies. In chapter 2, we took advantage of this development, and extended Schwarzschild's scheme to model VP shapes. We applied it to spherical models for the E0 galaxy NGC 2434, and showed that the observations imply the presence of a dark halo. Here we consider axisymmetric models and show how to use the extended Schwarzschild method to construct fully general three-integral models that can match any set of kinematic constraints. Independent implementations of the software were written by N.C. and R.v.d.M. A summary of this development is given by de Zeeuw (1997). In an earlier paper (van der Marel *et al.* 1998, chapter 4 of this thesis; see also van der Marel *et al.* 1997) we applied this modeling technique to the compact E3 elliptical M32, for which previous modeling had suggested the presence of a central massive black hole (BH) (e.g., Q95; Dehnen 1995). Cretton & van den Bosch (1999, chapter 5 of this thesis) describe an application to the edge-on S0 galaxy NGC 4342. Other groups are in the process of developing similar techniques to the one described here (e.g., Richstone *et al.* 1997; see also: Emsellem, Dejonghe & Bacon 1999; Matthias & Gerhard 1999).

This paper is organized as follows. In Section 3.1 we describe step by step how to construct the models (see Figure 3.1). We first discuss the mass models that we consider (Section 3.1.1). We describe how we choose a grid in integral space that yields a representative library of orbits (Section 3.1.2), how these orbits are calculated numerically (Section 3.1.3), how their properties are stored on a number of grids (Section 3.1.4), and how we model all aspects of the data taking and analysis, such as seeing convolution, pixel binning, and extraction of VPs (Section 3.1.5). We then present the method that we employ to determine the non-negative weight of each orbit (i.e., the number of stars traveling on each orbit), such that the global superposition of orbits produces a consistent model that best fits the observations (Section 3.1.6). Lastly, we discuss how we include optional smoothness constraints in the models (Section 3.1.7). In Section 3.2 we describe a set of alternative building blocks, the two-integral and isotropic components, for which the observable properties can be computed analytically. The smoothness of these com-

ponents makes them a convenient tool to use in conjunction with the regular orbits described in Section 3.1. Models can also be built entirely of these components, to obtain models with DFs of the form $f(E, L_z)$ or $f(E)$. In Section 3.3 we describe the tests that we have performed to establish the accuracy of our method. We present our conclusions in Section 3.4.

3.1 Construction of Dynamical Models

3.1.1 Mass Model

We study dynamical models in which all relevant quantities are axisymmetric, and symmetric with respect to the equatorial plane $z = 0$. It is sufficient to have the total gravitational potential, $\Phi \equiv \Phi_\star + \Phi_{\text{dark}}$, and the forces, $\vec{\nabla}\Phi$ available and tabulated on a grid, such that their values at any point can be recovered through interpolation. This is important, because the structure of real galaxies can be very complicated, and is not always well described in terms of analytical functions.

While the method works for arbitrary radial density profiles, it proves convenient for the purpose of presenting and testing our technique to consider models in which the mass density of the luminous material, ρ_\star , does have an analytical form:

$$\rho_\star(R, z) = \rho(s) = \rho_0 \left[\frac{s}{b} \right]^\alpha \left(1 + \left[\frac{s}{b} \right]^\gamma \right)^\beta \left(1 + \left[\frac{s}{c} \right]^\epsilon \right)^\delta, \quad (3.1)$$

where s is defined as $s^2 = R^2 + (z/q)^2$. This is an axisymmetric generalization of the spherical models studied previously by, e.g., Dehnen (1993), Tremaine *et al.* (1994) and Zhao (1996a), and includes the (α, β) models of Q95 as a limiting case. The model has a constant axial ratio q that does not vary with radius. The parameters b and c are characteristic lengths. At small radii ($r \ll b$) the density has a central cusp with logarithmic slope α (when $\alpha < 0$). At intermediate radii ($b \ll r \ll c$) the density falls off as $\rho_\star \propto r^{\alpha+\gamma\beta}$, while at large radii ($r \gg c$) $\rho_\star \propto r^{\alpha+\gamma\beta+\epsilon\delta}$. When viewed at an inclination angle i , the isophotes are ellipses of axial ratio $q' = (\cos^2 i + q^2 \sin^2 i)^{1/2}$. The luminosity density is $j = \rho_\star/\Upsilon$, where Υ is the average mass-to-light-ratio of the luminous material, which we assume to be constant.

For these models, the gravitational potential and the associated radial and vertical forces can all be obtained from one-dimensional (usually numerical) integrals (cf. eqs. [2.10]–[2.12] of Q95). We calculate the potential and forces in this way and tabulate them on a fine polar (r, θ) grid, with logarithmic sampling in radius and linear sampling in the angle. These tabulated values are used for the subsequent orbit calculations. It is straightforward to add the contributions from a dark component to the potential and the forces, as required for models with, e.g., a BH or a dark halo. In the case of a BH these contributions need not be tabulated, because they are known analytically.

For density distributions that are not stratified on similar concentric spheroids one must use more general techniques to calculate the gravitational potential and the associated forces. One possibility to determine these, while at the same time fitting a complicated surface brightness distribution, is to use a Multi-Gaussian Expansion (Emsellem *et al.* 1994). We do this in our modeling of the S0 galaxy NGC 4342 (chapter 5). Another possibility is to obtain ρ_\star through non-parametric deprojection of an observed surface brightness distribution (e.g., Dehnen 1995), and calculate the potential from a multipole or other expansion (e.g., Hernquist & Ostriker 1992; Zhao 1996a).

EXTENDED SCHWARZSCHILD METHOD

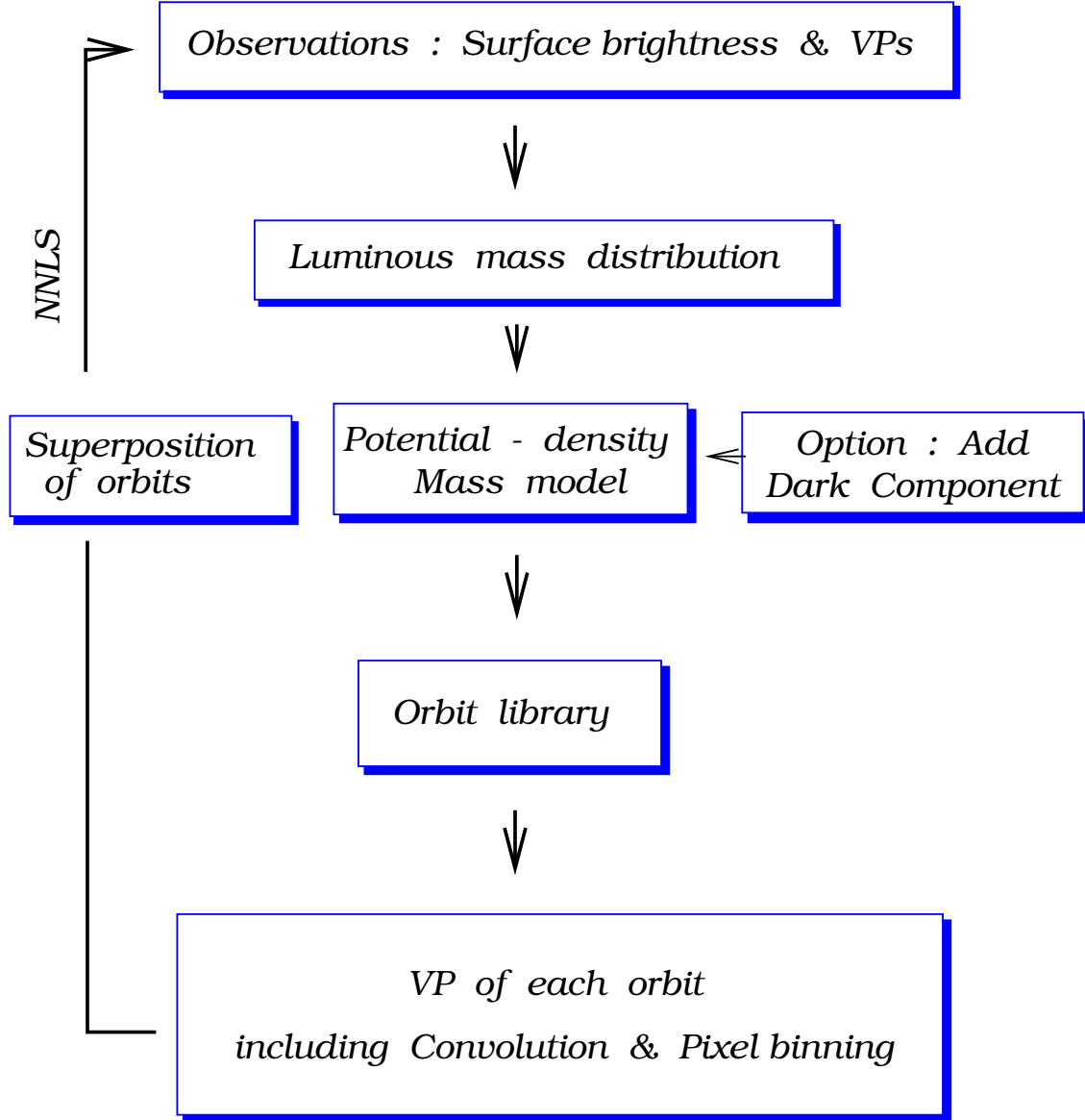


FIGURE 3.1— Flowchart of the extended Schwarzschild method. We find the non-negative superposition of the orbits with a least squares algorithm (NNLS). This combination of orbits reproduces a set of photometric (surface brightness distribution) and kinematic constraints (VPs).

3.1.2 Choice of Orbits

The results obtained with our modeling technique should not depend on the details of the orbit library. To achieve this, the library must represent the full variety of orbits in the given potential. In this section we describe how we have chosen to select orbits in order to fulfill this requirement.

In axisymmetric models, all orbits conserve at least two isolating integrals of motion: the energy E and the vertical component of the angular momentum L_z . Numerical studies have shown that many orbits conserve an additional third isolating integral I_3 , which is usually not known analytically (see e.g., Ollongren 1962; Innanen & Papp 1977; Richstone 1982; Dehnen & Gerhard 1993). These *regular* orbits are specified completely by the integrals of motion, and can be labeled by the values of E , L_z and I_3 .

For each energy E , there is one circular orbit in the equatorial plane, which has radius R_c and velocity $V_c^2 = R_c(\partial\Phi/\partial R)_{(R_c,0)}$. The angular momentum of this orbit, $R_c V_c$, is the maximum angular momentum at the given energy: $L_{\max}(E)$. We sample the energies in the model by adopting a logarithmic grid in R_c . Each R_c defines an energy E through the implicit relation $E = \Phi(R_c, 0) + \frac{1}{2}V_c^2$. The orbits in the model have $R_c \in [0, \infty)$. However, it is sufficient to adopt a grid of N_E values that covers only a finite range, $R_{c,\min}$ to $R_{c,\max}$, chosen so as to contain all but a negligible fraction of the total mass of the system. At each energy we sample the range of possible L_z values by adopting a grid in the quantity $\eta \equiv L_z/L_{\max}(E)$ ($\eta \in [-1, 1]$). Orbits with both $L_z > 0$ and $L_z < 0$ are included in the library, but the $L_z < 0$ orbits need not be calculated; they are obtained from the $L_z > 0$ orbits by reversing the velocity vector at each point along

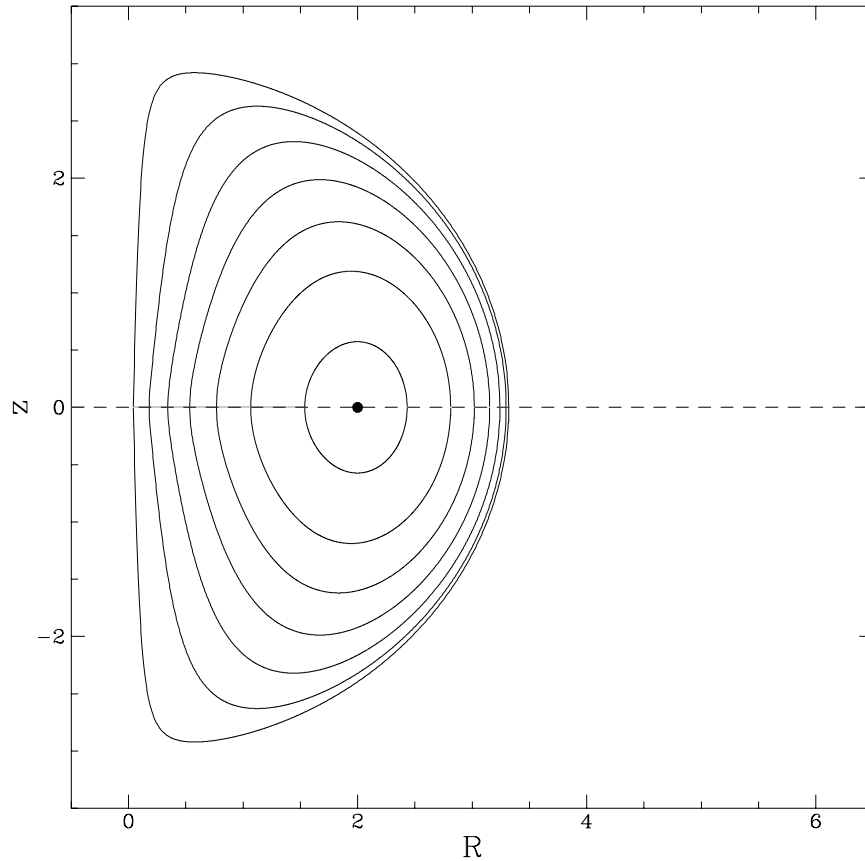


FIGURE 3.2— Zero Velocity Curves are plotted for 7 values of L_z uniformly sampled between $0.05L_z^{\max}$ and $0.95L_z^{\max}$ at an energy corresponding to a circular orbit radius $R_c = 2.0$ (indicated by the dot). The mass model is the one of our test model of Section 3.3.1, with no BH.

the orbit. We have calculated orbits for N_η values of η , spaced linearly between ϵ_1 and $1 - \epsilon_1$, where ϵ_1 is a small number. For numerical reasons, the special values $\eta = 0$ (radial orbits) and $\eta = 1$ (the circular orbit in the equatorial plane) are presumed to be represented by their closest neighbors on the grid, but are not included explicitly.

In an axisymmetric potential the orbit reduces to a two-dimensional motion in the meridional (R, z) plane in an effective gravitational potential $\Phi_{\text{eff}} = \Phi + \frac{1}{2}L_z^2/R^2$ (e.g., Binney & Tremaine 1987, hereafter BT). For fixed (E, L_z) , the position of a star is restricted to the region bounded by the ‘zero-velocity curve’ (ZVC), defined as the curve of values (R, z) that satisfy the equation $E = \Phi_{\text{eff}}$, and hence $v_R = v_z = 0$. Figure 3.2 illustrates ZVCs in the meridional plane. A regular orbit admits a third isolating integral, I_3 , that restricts its motion to a sub-region of the full region of phase-space accessible at the given (E, L_z) . This is illustrated in Figure 3.3, which shows a regular orbit viewed in the meridional plane. In our method we have chosen a numerical representation of I_3 that can be used to label the orbit. Every orbit with $L_z \neq 0$ touches the ZVC (Ollongren 1962). As suggested by Levison & Richstone (1985), we take the R coordinate of the ‘turning point’ on the ZVC (i.e., the intersection of the orbit with the ZVC), denoted by R_{ZVC} , as the third parameter to specify the orbit. At every (E, L_z) there is exactly one orbit that touches the ZVC at only one value, R_{thin} , of R : the so-called ‘thin tube’ orbit (see Figure 3.3). All other regular orbits touch the ZVC for at least two values of R_{ZVC} , one smaller than R_{thin} and one larger than R_{thin} . To sample the orbits at a given (E, L_z) , we calculate trajectories that are started with $v_R = v_z = 0$ from the ZVC, at a given radius R_{ZVC} (this radius determines v_ϕ according to $v_\phi = L_z/R$). Not every orbit launched in this way necessarily admits a third integral, since irregular orbits also touch the ZVC. Our orbit library therefore includes both regular and irregular orbits, and, as we shall see in Section 3.1.3, we have found it unnecessary to distinguish between them in all our tests and applications to date. To reduce redundancy in the library it is sufficient to consider only orbits with $R_{\text{ZVC}} \in [R_{\text{thin}}, R_{\text{max}}]$, where R_{max} is the radius at which the ZVC intersects the plane $z = 0$.

For the orbit library we have chosen to use N_{I_3} values of R_{ZVC} . Each point $(R_{\text{ZVC}}, z_{\text{ZVC}})$ on the ZVC is determined with the help of an angle w , which is sampled linearly between 0 and w_{thin} (see Figure 3.3). For numerical reasons, the special values $R_{\text{ZVC}} = R_{\text{thin}}$ (thin tube orbit) and $R_{\text{ZVC}} = R_{\text{max}}$ (equatorial orbit) are presumed to be represented by their respective closest neighbor on the grid, but are not included explicitly. Finding the starting point for the periodic orbit, R_{thin} , is straightforward (see e.g. Pfenniger & Friedli, 1993).

It is sufficient to calculate only orbits that are started from the ZVC with $z > 0$. Orbits started with $z < 0$ are obtained from those started with $z > 0$ by reversing the sign of z and v_z at each point along the orbit. Most orbits are themselves symmetric with respect to the equatorial plane (see e.g., Figure 3.3), so that this operation is redundant. However, this is not true for, e.g., the orbits parented by the 1/1 resonance between the R and z -motion (see Figure 3.11 below, or Figure 8 of Richstone 1982). Since we are only interested in constructing models that are symmetric about the equatorial plane, we do not view the orbits started with $z > 0$ and $z < 0$ from the ZVC as separate building blocks, but instead we consider only their sum.

The grid in $(R_c, \eta, R_{\text{ZVC}})$ completely specifies the orbit library. Appropriate choices for the parameters that characterize this grid are discussed in Section 3.3.

3.1.3 Orbit Calculation

For each $(R_c, \eta, R_{\text{ZVC}})$ we calculate a trajectory, started from the ZVC as described in Section 3.1.2. We have used several standard integration algorithms, including the Bulirsch-Stoer integrator (Press *et al.* 1992) and the Runge-Kutta-Fehlberg algorithm (Fehlberg 1968). We have experimented with both and found equivalent results. The former algorithm was used in chapter 4.

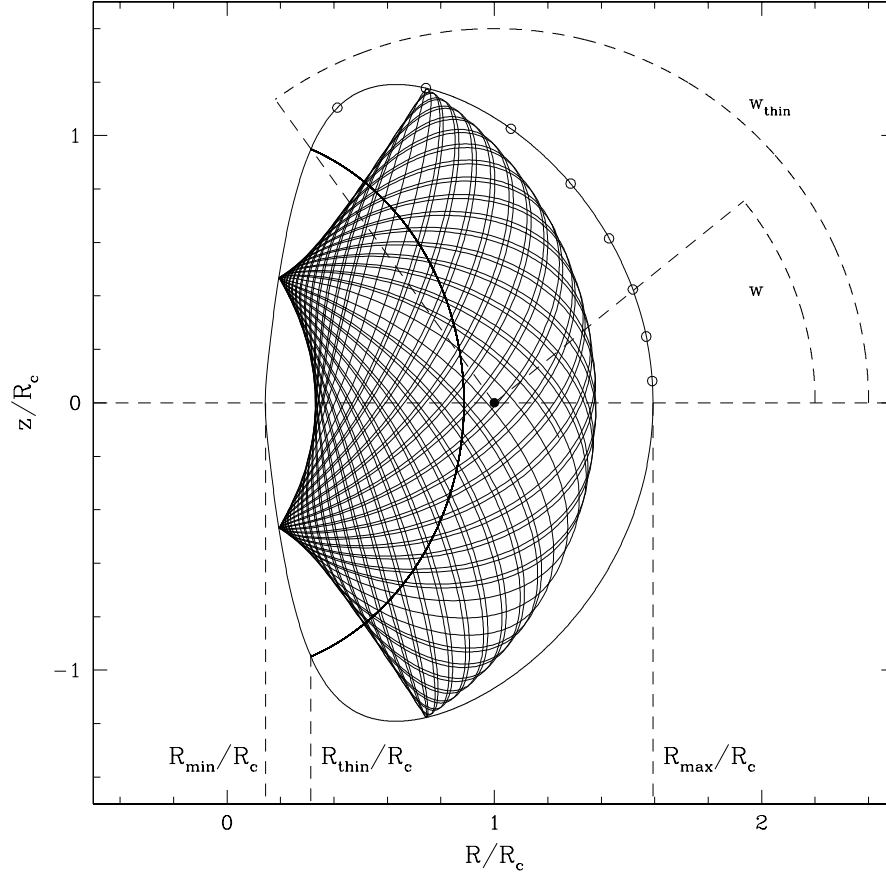


FIGURE 3.3— A regular orbit, the thin-tube periodic orbit, and the ZVC around them in the meridional (R, z) plane for the same mass model as in Figure 3.2. The radii R_{\min} , R_{\max} and R_{thin} are indicated (see text). The circular orbit is represented by a black dot. The different starting points on the ZVC are shown with open dots and the corresponding angles w and w_{thin} are indicated.

Here we use the Runge-Kutta-Fehlberg algorithm.

The results of the orbit calculations were used to approximate the ‘orbital phase-space density’ for each trajectory. Each phase-point along a calculated orbit was assigned a weight equal to the time step at that point, divided by the total integration time.

This procedure results in density distributions in phase-space, DF_{traj} and its corresponding spatial density ρ_{traj} . Orbits were calculated in the meridional plane, but all six phase-space coordinates are needed. The azimuthal velocity $v_\phi = L_z/R$ is completely specified. However, for projection onto the sky, the azimuthal angle $\phi \in [0, 2\pi]$ is also required. The distribution of stars over ϕ is homogeneous, because of the assumed symmetry. The weight at each time step was therefore divided into a number of equal ‘sub-weights’, and each was assigned a random ϕ . Furthermore, each sub-weight was divided in two, and one of the two parts was assigned phase-coordinates with (z, v_z) multiplied by -1 . This corrects for the fact that only orbits started with $z > 0$ from the ZVC were calculated (cf. Section 3.1.2). The trajectories should be integrated long enough so that the orbital phase-space densities no longer change significantly with time. Pfenniger (1984) proposed to check directly for the convergence of the orbital mass distribution. However, this may take a very long time, especially for orbits that are unusually close to a

high-order resonance, for orbits at very large radii or at very small radii close to a BH, and for irregular orbits (see also Merritt & Fridman 1996). We have used a cruder approach, in which we calculated each orbit for a fixed number (~ 200) of characteristic orbital periods. We found this to be sufficient for our purpose; longer integrations yield final models that are not significantly different. This is because the ‘noise’ in our modeling is dominated by the representation of phase space through a coarse discrete grid.

Orbits can have sharp edges in both the spatial and velocity dimensions. We found that a simple scheme to obtain smoother densities yielded slightly more accurate results for the final orbit superposition. To take into account the fact that each energy E in the orbit catalog represents all energies in some bin $[E_1, E_2]$ around it (defined by the choice of energy grid), a random energy \tilde{E} was drawn from the range $[E_1, E_2]$ for each normalized time step. The corresponding phase-coordinates (\vec{r}, \vec{v}) were then translated to the energy \tilde{E} , by replacing them by $([\tilde{R}_c/R_c]\vec{r}, [\tilde{V}_c/V_c]\vec{v})$, where R_c , \tilde{R}_c and V_c , \tilde{V}_c are the radii and circular velocities of the circular orbits at the energies E and \tilde{E} , respectively. This “dithering” approach is only approximately correct (it assumes that the potential is locally scale-free), but was found to work well in practice.

3.1.4 Storing the Orbital Properties

For each orbit we store both the intrinsic properties and the projected properties. The intrinsic properties are necessary to test for consistency of the final model. We store ρ_{traj} on an (r, θ) grid in the meridional plane, logarithmic in r and linear in $\theta \in [0, \frac{\pi}{2}]$. Angles $\theta > \frac{\pi}{2}$ need not be stored separately, because of symmetry with respect to the equatorial plane. We also store the lowest-order velocity moments of each orbit ($\rho_{\text{traj}}\langle v_i \rangle, \rho_{\text{traj}}\langle v_i v_j \rangle, i, j = r, \theta, \phi$) on the same grid, so as to be able to study the intrinsic dynamical structure of the final model.

The projected properties are necessary for comparison to observable quantities, such as the projected surface brightness and line-of-sight VP shapes. Only three coordinates of phase-space are available for comparison with observations: the projected positions x', y' (which we choose to be aligned with the photometric major and minor axis), and the line-of-sight velocity, $v_{\text{los}} (\equiv v_z')$. Given an inclination angle i of the galaxy ($i = 90^\circ$ means edge-on), these are related to the usual cylindrical coordinates (R, z, ϕ) in the following way:

$$\begin{aligned} x' &= R \sin \phi, \\ y' &= -R \cos i \cos \phi + z \sin i, \\ v_{\text{los}} &= (v_R \cos \phi - v_\phi \sin \phi) \sin i + v_z \cos i. \end{aligned} \tag{3.2}$$

To have the projected properties of the orbits accessible, we store their phase-space densities both on an (r', θ') grid on the projected plane of the sky (with similar properties as the intrinsic (r, θ) grid), and on a Cartesian (x', y', v_{los}) data cube (see Section 3.1.5). The former is used to reconstruct the projected surface density of the model. The latter is used to model observed kinematical quantities. The spatial grid size $(\Delta x, \Delta y)$ of the (x', y', v_{los}) cube is chosen to provide 2–5 times higher spatial resolution than the pixel size of the available kinematical observations. If observations with very different resolution are available for a galaxy (e.g., very high spatial resolution HST data in the central arcsec, and lower-resolution ground-based data out to an effective radius), it is best to store the data on two or more cubes with different spatial grid sizes and extents. During the orbit calculations we then store the phase-space densities simultaneously on all (x', y', v_{los}) cubes. Only the $x' \geq 0$ half of each cube needs to be stored, because each orbit has the same weight at (x', y', v_{los}) as at $(-x', -y', -v_{\text{los}})$. The size Δv of the velocity bins on the (x', y', v_{los}) cube(s) must be chosen to provide a proper sampling of the observed VPs. In

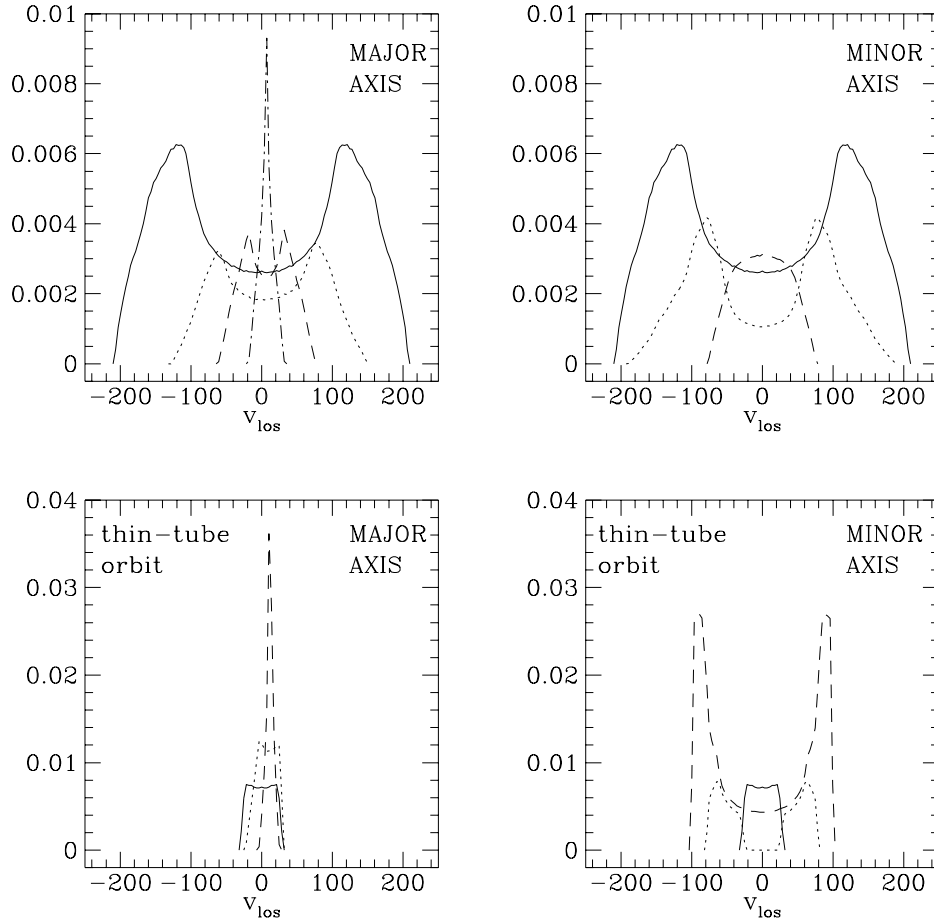


FIGURE 3.4— VPs as a function of line-of-sight velocity v_{los} (in km s^{-1}) for the two orbits of Figure 3.3. The regular and thin orbits are shown in the top and bottom panels, respectively, for viewing through ($1''$ -square) cells along the major axis (left) and minor axis (right), respectively. The orbits were not convolved with a PSF. For each panel, different lines correspond to cells at different distances r from the center: the full line corresponds to the central cell ($r = 0''$), the dotted line to $r = 1''$, the short dashed line to $r = 2''$ and the dot-short dashed line to $r = 3''$.

practice we use ~ 50 – 100 bins between $(-N_\sigma \sigma_{\text{max}}, N_\sigma \sigma_{\text{max}})$, where σ_{max} is the largest observed dispersion, and $N_\sigma = 4$ – 8 .

3.1.5 Modeling Observed Kinematical Quantities

Point-spread-function (PSF) convolution is essential when comparing model quantities with observed kinematical quantities in the central regions of galaxies. Seeing convolution correlates information in the two spatial dimensions x', y' , but not in v_{los} :

$$F_{\text{conv}}(x'_0, y'_0, v_{\text{los}}) = F \otimes \text{PSF} = \iint F(x', y', v_{\text{los}}) \text{PSF}(x' - x'_0, y' - y'_0) dx' dy', \quad (3.3)$$

where F is the function to convolve, PSF is the point-spread function, and F_{conv} is the result of the convolution of F with the PSF. The final model is a linear superposition of the orbits, so the (x', y', v_{los}) cubes for each orbit may be individually convolved with the PSF. As in chapter 2, we do the convolution for each velocity slice efficiently by multiplications in Fourier space, using Fast Fourier Transforms (e.g., Press *et al.* 1992).

Kinematical data is generally obtained either through small, discrete apertures, along a number of slit-positions, or may derive from two-dimensional integral field spectroscopy (e.g., Bacon *et al.* 1995). Any setup of this kind can be simulated by our models, including possible spatial binning along a slit. For each observational ‘aperture’, we choose the (x', y', v_{los}) cube with the most appropriate cell size, convolve it with the relevant PSF, and bin the results spatially over the aperture size. This yields a one-dimensional velocity histogram, for each orbit and for each observation. Examples of such ‘orbital VPs’ are shown in Figure 3.4.

Kinematical observations provide information on the line-of-sight VPs

$$\text{VP}(x', y', v_{\text{los}}) = \iiint \text{DF} \, dv_{x'} \, dv_{y'} \, dz', \quad (3.4)$$

at different positions (x', y') on the projected face of a galaxy. In practice, the normalization of $\text{VP}(x', y', v_{\text{los}})$ is based on the photometric data. It is often useful to parametrize observed VPs by a few numbers. A common choice for such a VP parametrization is the Gauss-Hermite expansion. We follow the notation of vdMF, in which the VP is represented as

$$\text{VP}(v_{\text{los}}) = \frac{\alpha(w)}{\sigma} \sum_{l=0}^N h_l H_l(w), \quad (3.5)$$

with

$$w = (v_{\text{los}} - V)/\sigma, \quad \alpha(w) = \frac{1}{\sqrt{2\pi}} e^{-w^2/2}. \quad (3.6)$$

The h_l are the Gauss-Hermite moments (hereafter GH-moments) defined by

$$h_l = 2\sqrt{\pi} \int_{-\infty}^{\infty} \text{VP}(v_{\text{los}}) \alpha(w) H_l(w) \, dv_{\text{los}} \quad (l = 0, \dots, L). \quad (3.7)$$

Each H_l is a Hermite polynomial (see Appendix A of vdMF). The quantities V and σ characterize the ‘weighting function’, $\alpha(w)H_l$, in the integral (3.7). When describing observations, V and σ are usually taken to be the velocity and dispersion of the Gaussian that best fits the observed VP. With this choice, $h_1 \equiv h_2 \equiv 0$. GH-moments of higher order describe deviations from a Gaussian. Only the moments of order $L \leq 6$ are generally measured from galaxy spectra, due to the finite spectral (and thus velocity) resolution of the observations.

If we envision galaxies as consisting of orbital building blocks, then the overall VP measured for a given observation is just the superposition of the individual orbital VPs. Similarly, the observed GH-moments are just a linear superposition of the GH-moments of the individual orbital VPs, provided that the observed V and σ for the given observation are used in the weighting function $\alpha(w)H_l(w)$. Thus, as described in detail in chapter 2, to fit the kinematical observations we may restrict ourselves to solving a linear superposition problem for the Gauss-Hermite moments. The constraints are then that $h_1 = h_2 = 0$, and the h_l with $l \geq 3$ should equal their observed values. It must be stressed that this approach is general, and assumes neither that the observed VPs are well-described by the lowest-order terms of a GH-series, nor that the orbital VPs are well-described by the lowest-order terms of a GH-series. Nonetheless, if a full non-parametric estimate of the observed VPs is available there is no need to restrict the analysis to the lowest-order GH-moments. Our technique can just as easily fit the individual velocity bins of the observed VPs.

3.1.6 Fitting the Constraints

Constructing a model consists of finding a weighted superposition of the stellar orbits in the library that reproduces two sets of constraints:

- **Consistency constraints** for the stellar luminosity distribution. The model should reproduce the initially assumed luminous stellar density $\rho_*(R, z)$ (Section 3.1.1), for each cell of the meridional (r, θ) grid, for each cell of the projected plane (r', θ') grid, and for each aperture for which there is kinematical data. In theory, it is sufficient to fit only the meridional plane masses, because projected densities are then fit automatically. In practice this might not exactly be the case, because of discretization. To circumvent this, the projected masses may be included as separate constraints. Note that for axisymmetric models the projected density does not uniquely specify the intrinsic density (e.g., Rybicki 1986; Gerhard & Binney 1996).
- **Kinematical constraints.** The model should reproduce the observed kinematics of the galaxy, including VP shapes. As discussed, we express this as a set of linear constraints on the GH-moments of the VPs (chapter 2).

Finding the orbit superposition that best fits these constraints amounts to solving a linear problem, which can be written in matrix notation as $\mathbf{B}\vec{\gamma} = \vec{c}$ (chapter 2). The matrix \mathbf{B} contains the mass that each orbit contributes to each relevant intrinsic or projected grid cell, and the GH-moments that each orbit contributes to each kinematical observation. The vector \vec{c} contains the mass predicted by $\rho_*(R, z)$ for each relevant intrinsic or projected grid cell, and the observed GH-moments for all kinematical observations; the vector $\vec{\gamma}$, which is to be solved for, contains the weight of each orbit, i.e., the total mass of stars on each orbit. These weights should be non-negative. Using the terminology introduced in Section 3.1.5, the basic Schwarzschild equation becomes

$$\sum_{j=1}^{N_{\text{orbits}}} \gamma_j \text{VP}_{ij} = \text{VP}_i, \quad (3.8)$$

with VP_{ij} the individual VP of orbit j at constraint point i , and VP_i the observed VP for the same constraint point.

The superposition problem can be expressed as a non-negative least squares (NNLS) fit for the above matrix equation. We have used the NNLS routine of Lawson & Hanson (1974) to solve it. The NNLS routine finds a combination of non-negative orbital occupancies (which need not be unique) that minimizes the usual \mathcal{L}^2 norm $\|\mathbf{B}\vec{\gamma} - \vec{c}\|$. This norm can be viewed as a χ^2 quantity that measures the quality of the fit to the constraints. The NNLS routine always finds a best solution. It need not be acceptable in light of the observations; this must be assessed through the χ^2 of the best fit, and by comparison of the model predictions to the constraints.

As is customary in least-squares fitting, the model predictions for each constraint and the actual constraint values (the elements of the vector \vec{c}) are weighted by the errors in the constraints. Observational errors are available for the kinematical constraints. In principle one would like the consistency constraints to be fit with machine precision. It turns out that this is generally unfeasible, because of discretization. It was found that models with no kinematical constraints could at best simultaneously fit both the intrinsic and the projected masses with a fractional error of $\sim 5 \times 10^{-3}$ (when using ~ 1000 orbits). We therefore assigned fractional errors of this size to the masses in the consistency constraints. In principle one would like to include also the observational surface brightness errors in the analysis. Unfortunately, this requires the exploration of a large set of three-dimensional mass densities (that all fit the surface photometry to within the errors), which is prohibitively time-consuming.

3.1.7 Regularization

Our orbit superposition models are not generally smooth in integral space, as a result of the ‘ill-conditioned’ numerical nature of the NNLS matrix equation that is being solved. There are no physical theories that describe exactly how smooth the DF of a stellar system should be, but some degree of smoothness should be expected. Our technique can be extended in a straightforward manner to yield smooth solutions, by adding linear regularization constraints to the NNLS matrix equation (e.g., Press *et al.* 1992; Merritt 1993). This has the same effect as the addition of ‘maximum entropy’ constraints (Richstone & Tremaine 1988). For linear regularization, each regularization constraint must be of the form

$$\sum_k s_{k,l} \gamma_k = 0 \pm \Delta, \quad (3.9)$$

thus providing an extra row to the matrix equation. The γ_k are the orbital weights that make up the vector $\vec{\gamma}$, and l is the number of the regularization constraint. The parameter Δ sets the amount of regularization. Models with $\Delta \rightarrow \infty$ have no regularization, while models with $\Delta \rightarrow 0$ give infinite weight to the regularization constraints. Alternatively, one may view this as adding a term $\lambda \|\mathbf{S}\vec{\gamma}\|$ to the norm $\|\mathbf{B}\vec{\gamma} - \vec{c}\|$ that is minimized by the NNLS routine, where \mathbf{S} is the matrix with elements $\{s_{k,l}\}$, and $\lambda \equiv 1/\Delta$ is a regularization parameter (Zhao 1996b).

Many choices are possible for the matrix \mathbf{S} , with the only requirement that the norm $\|\mathbf{S}\vec{\gamma}\|$ should provide a measure of the smoothness of the solution. Our choice is based on the fact that we consider the NNLS solution $\gamma(R_c, \eta, R_{\text{ZVC}})$ to be “smooth” if the second derivatives of the (unitless) function $\gamma(R_c, \eta, R_{\text{ZVC}})/\gamma_0(R_c)$ are small. Here the “reference weights” $\gamma_0(R_c)$ are a rough approximation to the energy dependence of the model. These are determined beforehand, e.g., by studying the spherical isotropic limit of the given mass density. We view the three-dimensional numerical grid in integral space as a Cartesian lattice, and we approximate the second derivatives by second order divided differences (eq. 18.5.10 of Press *et al.* 1992). We assume that the distance between adjacent grid points on the lattice is unity, independent of the cartesian direction in which they are adjacent. This (arbitrarily) solves the problem that the axes of the grid in integral space have different units and yields three regularization constraints for each grid point (i, j, l) that is not on a boundary: $-\gamma_{i-1,j,l} + 2\gamma_{i,j,l} - \gamma_{i+1,j,l} = 0$, $-\gamma_{i,j-1,l} + 2\gamma_{i,j,l} - \gamma_{i,j+1,l} = 0$, and $-\gamma_{i,j,l-1} + 2\gamma_{i,j,l} - \gamma_{i,j,l+1} = 0$.

3.2 Two-integral Components and Isotropic Components

3.2.1 Definition

Individual regular orbits correspond to building blocks with a DF proportional to $\delta(E - E_0) \delta(L_z - L_{z,0}) \delta(I_3 - I_{3,0})$, for given $(E_0, L_{z,0}, I_{3,0})$. These are not the only building blocks that can be used to construct models. One may also use ‘two-integral components’, which correspond to the DF

$$f_{[E_0, L_{z,0}]}^\delta \equiv C_{[E_0, L_{z,0}]} \delta(E - E_0) \delta(L_z - L_{z,0}), \quad (3.10)$$

or ‘isotropic components’, which correspond to the DF (cf. Richstone 1982)

$$f_{[E_0]}^\delta \equiv C_{[E_0]} \delta(E - E_0). \quad (3.11)$$

We choose the normalization coefficients $C_{[E_0, L_{z,0}]}$ and $C_{[E_0]}$ such that the total mass of each component is equal to unity; explicit expressions are derived in Appendix 3.A.1.

The two-integral components are smoother building blocks than the regular orbits, since they fill completely the ZVC and do not have the sharp edges of the regular orbits. It is useful

to view them as a particular combination of all orbits that could be numerically integrated at the given $(E_0, L_{z,0})$, both regular orbits that fill only a subset of the area enclosed by the ZVC (therefore admitting 3 independent integrals of motion) *and* irregular orbits that occupy a larger area (admitting only 2 integrals; note that an irregular orbit does not necessarily fill the entire phase-space region defined by $(E_0, L_{z,0})$; see Merritt & Valluri 1996 for a discussion of the triaxial case). Similarly, an isotropic component is a weighted combination of all two-integral components (i.e., all orbits) at the given energy E_0 . The region in space occupied by such a component is bounded by the equipotential surface $\Phi(R, z) = E_0$.

The two-integral and isotropic components are useful, because their properties can be calculated semi-analytically. By using *only* two-integral components in the NNLS orbit superposition, one can construct $f(E, L_z)$ models for arbitrary spheroidal potentials. This provides a new and convenient way of constructing such models, which adds to the several techniques already in existence for this purpose (Hunter & Qian 1993; Dehnen & Gerhard 1994; Kuijken 1995; Magorrian 1995; M96b). Using only isotropic components in the NNLS orbit superposition is generally less useful, because these components follow equipotential surfaces, which are rounder than isodensity surfaces. Thus, they cannot be used to build self-consistent isotropic axisymmetric models. In chapter 2, we describe how to use them to build spherical isotropic models. Alternatively, the two-integral and isotropic components may be used in the superposition *in conjunction with* the regular orbits. This has two advantages. First, these components are smoother, and their inclusion therefore reduces numerical noise that arises from the discrete representation of phase space (see also Zhao 1996b). Second, addition of these components provides a way to include all irregular orbits in the models.

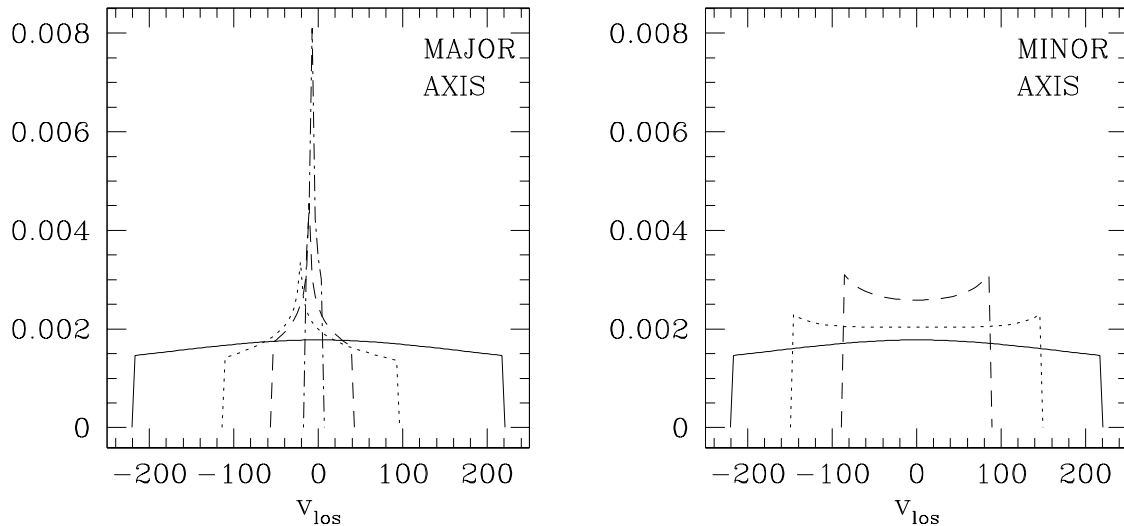


FIGURE 3.5— VPs of an individual two-integral component with the same (E, L_z) as the orbits in Figure 3.3, viewed along the major axis (left) and minor axis (right), respectively. Line types are the same as in Figure 3.4. However, unlike in Figure 3.4, these VPs were calculated for a point along either axis, and were not ‘integrated’ over cells.

3.2.2 Velocity Profiles

The VP of a two-integral component is obtained by substitution of the DF of equation (3.10) into equation (3.4). The resulting integral may be written as

$$\text{VP}_{[E_0, L_{z,0}]}(x', y', v_{\text{los}}) = C_{[E_0, L_{z,0}]} \int dz' J_{[E_0, L_{z,0}]}, \quad (3.12)$$

where

$$J_{[E, L_z]} \equiv \left| \frac{\partial(v_{x'}, v_{y'})}{\partial(E, L_z)} \right|_{[E, L_z]}, \quad (3.13)$$

is the Jacobian for the change of variables from $(v_{x'}, v_{y'})$ to (E, L_z) . In Appendix 3.A.2 we give an explicit expression for this Jacobian. The integration in equation (3.12) extends over those z' for which there exist velocities $(v_{x'}, v_{y'})$ such that $E(x', y', z', v_{x'}, v_{y'}, v_{\text{los}}) = E_0$ and $L_z(x', y', z', v_{x'}, v_{y'}, v_{\text{los}}) = L_{z,0}$. Similarly, the VP for an isotropic component may be written as

$$\text{VP}_{[E_0]}(x', y', v_{\text{los}}) = C_{[E_0]} \int dz' \int dL_z J_{[E_0, L_z]}. \quad (3.14)$$

The projected density for a two-integral or isotropic component, at projected position (x', y') , is obtained as the integral of $\text{VP}(x', y', v_{\text{los}})$ over v_{los} .

Equations (3.12) and (3.14) can be used to calculate the VPs of the two-integral and isotropic components through numerical quadratures, without the need for calculating orbital trajectories. The only difficulty lies in finding the domain of integration in z' . We illustrate this for the case of an edge-on system ($i = 90^\circ$). In this case $J_{[E_0, L_{z,0}]} = (z' v_{y'})^{-1}$, with

$$v_{y'} = \sqrt{2(E_0 - \Phi) - \left(\frac{L_{z,0} + x' v_{\text{los}}}{z'} \right)^2 - v_{\text{los}}^2} \quad (3.15)$$

(cf. eqs. [3.32, 3.33]). We will refer to the expression under the square-root as W . The integration in equation (3.12) extends over those z' for which $W \geq 0$. We find the roots of W numerically. We start by finding the roots of $2(E_0 - \Phi) - v_{\text{los}}^2 = 0$. This gives an interval that encompasses *all* real roots of W , because $2(E_0 - \Phi) - v_{\text{los}}^2 \geq W$. Then we subdivide this interval in many (~ 100 – 1000) small segments, and check whether the sign of W differs at the ends of each segment. If it does, we find the root between these two points through bisection. The continuity of the resulting VP was used to check whether all required integration domains in z' were found. For the potentials studied here, we typically find two or four roots.

The VP calculation for edge-on *isotropic* components is less complicated. The Jacobian is quadratic in L_z , and can be written as $J_{[E_0, L_z]} = [(L_z^+ - L_z)(L_z - L_z^-)]^{-1/2}$. One can show that L_z^- and L_z^+ are real if $2(E_0 - \Phi) - v_{\text{los}}^2 \geq 0$. In this case the integral over dL_z in equation (3.14) extends from L_z^- to L_z^+ , and is always equal to π . Thus, if there is a z'_{max} for which $2(E_0 - \Phi) - v_{\text{los}}^2 = 0$, then

$$\text{VP}_{[E_0]}(x', y', v_{\text{los}}) = 2\pi C_{[E_0]} z'_{\text{max}}. \quad (3.16)$$

If there is no such z'_{max} (i.e., if v_{los} exceeds the escape velocity at the tangent point), then $\text{VP}_{[E_0]}(x', y', v_{\text{los}}) = 0$.

Figure 3.5 shows examples of the VP of a two-integral component along the major and minor axes of an edge-on system.

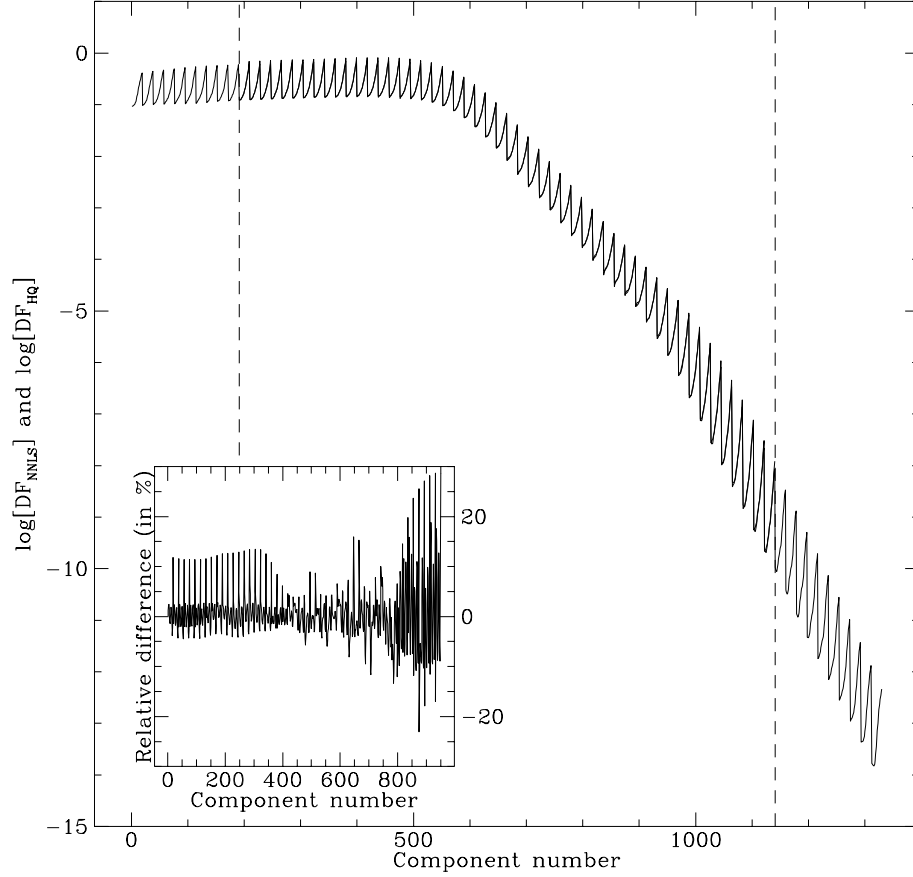


FIGURE 3.6— The DF $f(E, L_z)$ for the test model with a $3 \times 10^6 M_\odot$ BH described in Section 3.3.1. The DF from the semi-analytical HQ algorithm and from the extended Schwarzschild technique, using only two-integral components, are plotted as function of the component number. The components run in order or energy, and in order of L_z for each energy; this causes the jagged appearance of the curves. The two curves mostly overlap in the comparison interval (dashed vertical lines). The insert shows the relative difference (in per cent) between the two DFs. The agreement between the two methods of calculating the DF is acceptable.

3.3 Tests

3.3.1 The Test Model

The most useful tests for our axisymmetric implementation are those for which the results can be compared to analytical results, or to semi-analytical or numerical results that were obtained with an independent algorithm. Models with $f(E, L_z)$ DFs have been widely studied in the past five years (e.g., Evans 1993, 1994; Hunter & Qian 1993; Dehnen & Gerhard 1994; Evans & de Zeeuw 1994; Kuijken 1995; Q95; Magorrian 1995; M96b). Their properties can be derived semi-analytically, and a variety of algorithms and numerical implementations have been presented to derive the DF $f(E, L_z)$ that generates a given luminous mass density $\rho_\star(R, z)$ in a given potential $\Phi(R, z)$. These models therefore provide an ideal test case. Here we present two tests where we use our method to reproduce the properties of an edge-on $f(E, L_z)$ model.

We consider a model with a luminous mass density of the form (3.1), with parameters: $\alpha = -1.435$, $\beta = -0.423$, $\gamma = \epsilon = 2.0$, $\delta = -1.298$, $b = 0.55''$, $c = 102.0''$, $q = 0.73$, $\rho_0 = j_0 Y M_\odot / L_{\odot, V}$,

$j_0 = 0.463 \times 10^5 L_{\odot, V} \text{pc}^{-3}$ (for an assumed distance of 0.7 Mpc). We calculate the potential of the test model, $\Phi \equiv \Phi_{\star} + \Phi_{\text{dark}}$, under the assumption that $\Upsilon = 2.5$, and with the option of a nuclear BH ($\Phi_{\text{dark}} = -GM_{\text{BH}}/r$) of mass $M_{\text{BH}} = 3 \times 10^6 M_{\odot}$. All these parameters are based on the application of our technique to the case of the galaxy M32, which was presented in chapter 4. This analogy with M32 was chosen mainly to demonstrate the accuracy of our method for a realistic galaxy model.

The luminous mass density $\rho_{\star}(R, z)$ and potential $\Phi(R, z)$ determine uniquely only the even part $f_{\text{even}}(E, L_z)$ of the DF, $f(E, L_z) \equiv f_{\text{even}}(E, L_z) + f_{\text{odd}}(E, L_z)$. For our test model we specify (arbitrarily) the extreme case that $f_{\text{odd}}(E, L_z) = f_{\text{even}}(E, L_z)$ for $L_z > 0$, and that $f_{\text{odd}}(E, L_z) = -f_{\text{even}}(E, L_z)$ for $L_z < 0$ (and by definition, $f_{\text{odd}}(E, 0) = 0$). Thus, the $f(E, L_z)$ test model is ‘maximally rotating’: all the stars are rotating in the same sense and have $L_z > 0$.

First, the unique $f(E, L_z)$ DF of the test model was calculated using the approach described in chapter 4. We will refer to the resulting DF as DF_{HQ} (for Hunter & Qian, 1993). The kinematical predictions (VPs, GH-moments, etc.) for the test model DF were subsequently calculated using the expressions and software of Q95. The Jeans equations were used as in chapter 5 to compute the intrinsic second-order velocity moments $\langle v_{\phi}^2 \rangle$ and $\langle v_r^2 \rangle = \langle v_{\theta}^2 \rangle$ in the meridional plane. Our tests in Sections 3.3.2 and 3.3.3 are aimed at assessing how well our algorithm can reproduce the test model properties thus calculated with independent methods. This allows us to test all key aspects of the orbit model construction, including the sampling of integral space, orbit calculation, discreteness effects of the spatial grids, projection into the data cubes, seeing convolution, and the NNLS algorithm. Hence, it is no great drawback that our tests are restricted to two-integral models.

3.3.2 Reproducing the Test Model with Two-integral Components

We first describe tests of the extended Schwarzschild technique with only two-integral components. We used an (E, L_z) grid as described in Section 3.1.2, with $N_E = 70$, $R_{c, \text{min}} = 10^{-4.2}$ arcsec, $R_{c, \text{max}} = 10^{4.2}$ arcsec, $N_{\eta} = 19$, and $\epsilon_1 = 0.01$. Only components with $L_z > 0$ were included in the superposition; the resulting models are therefore by definition maximally rotating with a DF of the form $f(E, L_z)$. The DF is determined uniquely by the mass density. Kinematical constraints are therefore not required in the NNLS fit, but only constraints on the consistency of the stellar luminosity distribution (see Section 3.1.6). For these, the polar (r, θ) and (r', θ') grids in the meridional plane and on the projected plane of the sky (see Section 3.1.4) were chosen to have 16 bins in the radial coordinate between $R_{c, \text{min}}$ and $R_{c, \text{max}}$, and $N_{\theta} = N_{\theta'} = 7$ bins in the angular coordinate (a rather modest resolution, but similar tests with finer grids yielded similar accuracies). We semi-analytically (eq. [4-140b] of BT) calculated the isotropic DF $f(E)$ for the spherical version of the test model, and used the corresponding masses on our energy grid as reference weights for the regularization (see Section 3.1.7).

The NNLS algorithm yields the mass on each (E, L_z) grid cell, i.e., the integral of $dM / dE dL_z$ over the grid cell. It does *not* directly yield the DF $f(E, L_z)$, which by definition is the density in the six dimensional phase-space. However, for a two-integral model there is a simple relation between $dM / dE dL_z$ and the DF $f(E, L_z)$, as derived in Appendix 3.B.1. With equation (3.39) the NNLS fit provides an estimate of the DF, which we will denote DF_{NNLS}. We compare it with DF_{HQ} on the same grid, but to avoid possible border effects, we restrict the comparison to the $N_E = 50$ energy grid points with R_c between 10^{-3} and 10^3 arcsec (see Figures 3.6 and 3.7 for the test models with and without a central BH, respectively).

The DFs agree well over 10 and 20 orders of magnitude, respectively. The inserts show the percentage errors in the DF calculation. Note that the largest errors occur at grid points that carry little mass, e.g., at large radii. The orbit library in these figures is numbered as follows.

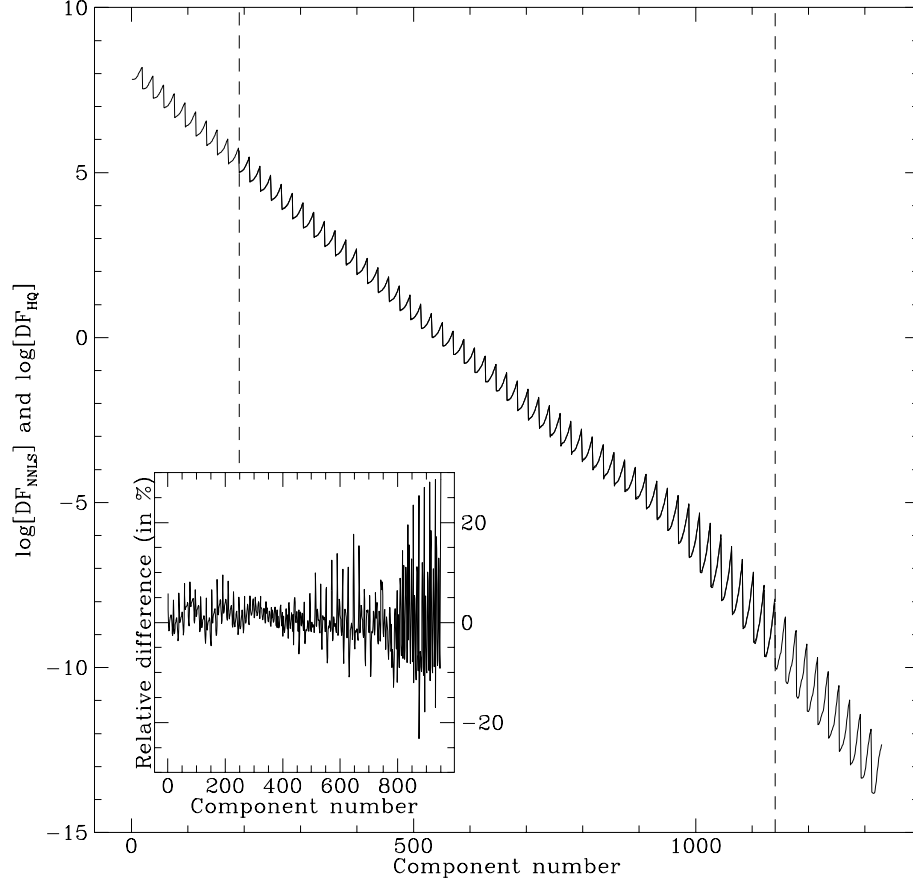


FIGURE 3.7— Similar to Figure 3.6, but for the same model without a central BH.

For each value of $R_c(E)_j$ of the energy, L_z runs monotonically from $L_{z,\min}$ to $L_{z,\max}$, covering N_η ($= 19$ here) components. The next orbit corresponds to $L_{z,\min}$ of $R_c(E)_{j+1}$, etc. This choice of numbering causes the jagged appearance of the DF. Q95 plotted $f(E, L_z = 0)$ and $f(E, L_z = L_{z,\max})$ as a function of E (see their Figure 8), and such curves would appear as envelopes in Figures 3.6 and 3.7.

To assess the influence of smoothing on the accuracy with which our technique recovers the DF, we have studied the dependence of the RMS logarithmic residual

$$\text{RMS}_{\log \text{DF}} \equiv \left[\frac{1}{N_E N_\eta} \sum_{i=1}^{N_E} \sum_{j=1}^{N_\eta} (\log \text{DF}_{\text{NNLS}} - \log \text{DF}_{\text{HQ}})^2 \right]^{1/2}, \quad (3.17)$$

on the regularization parameter Δ (see Section 3.1.7). When Δ tends to zero, the regularization constraints receive infinite weight. This yields a very smooth DF, but one that doesn't fit the consistency (mass) constraints very well, and therefore doesn't approximate DF_{HQ} very well. At the other extreme, when Δ is very large there is hardly any smoothing, and the mathematically ill-conditioned nature of the problem yields a very jagged solution that also doesn't match DF_{HQ} . Figure 3.8 shows $\text{RMS}_{\log \text{DF}}(\Delta)$ for the following three cases: (a) the case in which only the masses on the meridional plane (r, θ) grid are included as constraints in the NNLS fit; (b) the case in which only the masses on the projected plane (r', θ') grid are included as constraints; and

(c) the case in which both are included as constraints. In principle, deprojection of the projected mass density is unique for an edge-on system, so these approaches should recover equivalent results. However, this is not true in practice because of discretization effects. In all three cases, $\text{RMS}_{\log \text{DF}}$ has a minimum near $\log \Delta \approx 2.0$, which is thus the optimum smoothing. The value of $\text{RMS}_{\log \text{DF}}$ at the minimum is only mildly different for the different cases, but (a) yields the slightly better results. We have therefore adopted case (a) for all our further test calculations. The minimum $\text{RMS}_{\log \text{DF}}$ is ~ 0.02 ; this corresponds to a 5% RMS difference between DF_{NNLS} and DF_{HQ} . The mass-weighted RMS difference,

$$\text{RMS}_{\text{DF}} = \left[\iint \text{DF}_{\text{HQ}} \left[\frac{\text{DF}_{\text{NNLS}} - \text{DF}_{\text{HQ}}}{\text{DF}_{\text{HQ}}} \right]^2 d^3 \vec{x} d^3 \vec{v} \right] / \left[\iint \text{DF}_{\text{HQ}} d^3 \vec{x} d^3 \vec{v} \right]^{1/2}, \quad (3.18)$$

for the model with the optimum smoothing is also $\sim 5\%$. This level of accuracy in the determination of the DF is similar to that obtained with other techniques (see Gerhard *et al.* 1998; Matthias & Gerhard 1999).

Figure 3.9 compares the predictions for the meridional plane velocity moments to the results of the Jeans equations, for the model without a BH and with the optimum smoothing. The top and bottom row of the figure show $\langle v_\phi^2 \rangle^{1/2}$ and $\langle v_r^2 \rangle^{1/2} = \langle v_\theta^2 \rangle^{1/2}$, respectively. We plot separately each angular sector of the polar grid in the meridional plane. In each row, the first panel is closest to the symmetry axis and the last one is closest to the equatorial plane. Full lines show predictions of the extended Schwarzschild technique, and dashed lines the results obtained from the Jeans equations. The model predictions were interpolated between the (E, L_z) grid points to get smoother results. Overall the agreement is very good, and better than 1%. This is better than the $\sim 5\%$ agreement in the DF, because the velocity moments are integrals over the DF (such that errors tend to cancel). The errors in the velocity moments are largest near the symmetry axis, since in the extended Schwarzschild technique only a few components with very low L_z can reach this region of the meridional plane. However, the errors are always $\lesssim 2 \text{ km s}^{-1}$.

3.3.3 Reproducing the Test Model with Regular orbits

The next step in our testing procedure is to try to reproduce the properties of the test model with regular orbits, rather than two-integral components. The first obvious question is whether we can give the orbit superposition algorithm constraints that force it to generate a model with a DF of the form $f(E, L_z)$, which can then be compared to the distribution function DF_{HQ} calculated as in Section 3.3.1. Unfortunately, there is no set of simple linear kinematic constraints that force the NNLS algorithm to produce an $f(E, L_z)$ model. One can certainly impose the necessary conditions that $\langle v_r^2 \rangle = \langle v_\theta^2 \rangle$ and $\langle v_r v_\theta \rangle = 0$, but these conditions are not sufficient; an $f(E, L_z)$ model is fully determined only by constraints on *all* its higher order velocity moments (e.g., Magorrian & Binney 1994).

We therefore restrict ourselves here to a simpler test. We calculate an orbit library in the gravitational potential of the test model, but do not do a subsequent NNLS fit. Instead, we fix the orbital weights γ_j to those appropriate for an $f(E, L_z)$ model, and merely calculate the projected kinematical quantities for some observational test setup, given these orbital weights. The results are compared to the same quantities but now calculated from DF_{HQ} as described in Section 3.3.1. This tests all of the important parts of our method that were not already tested by the calculations in Section 3.3.2, namely the orbit calculation, the projection into data cubes and VPs, and the seeing convolution.

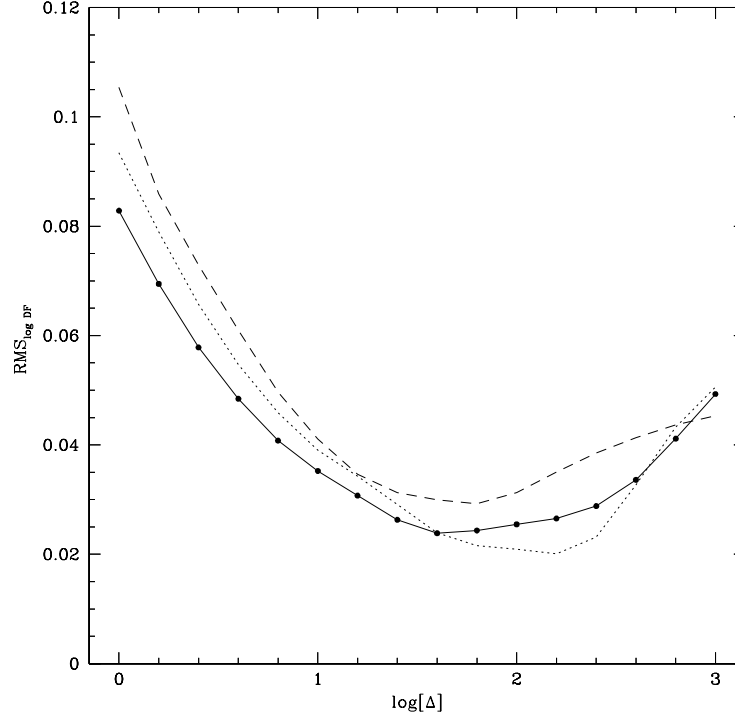


FIGURE 3.8— The RMS logarithmic residual $\text{RMS}_{\log \text{DF}}$ for the test model of Figure 3.7, as function of the logarithm of the regularization parameter Δ . The residual measures the difference between the DF as calculated with the semi-analytical HQ algorithm, and as calculated with the extended Schwarzschild technique, using only two-integral components. The different line-types indicate the cases in which only masses in the meridional plane are included as consistency constraints (dotted line), in which only projected plane masses are included (dashed line), or in which both are included (full line with dots).

The main difficulty with this test is that it requires knowledge of the orbital weights for an $f(E, L_z)$ model, i.e., of the differential mass density $dM / dE dL_z dR_{\text{zvc}}$, on the grid of quantities (E, L_z, R_{zvc}) that we use to sample orbit space (cf. Section 3.1.2). This is not as straightforward as the calculation of $dM / dE dL_z$ described in Appendix 3.B.1. In fact, the orbital weights can only be easily calculated if an explicit expression exists for the third integral, which is not the case for our test model. However, if the model has a central BH then at small radii, or high energies, the potential is Keplerian and spherical (as it is at large radii, or low energies, because of the finite mass of the model). In this potential, all the integrals of motion are known, and these limits are therefore analytically tractable.

At high energies (small radii) the test model reduces to a scale-free axisymmetric mass density cusp with an $f(E, L_z)$ DF in a spherical Kepler potential. This limit was studied analytically by de Bruijne, van der Marel & de Zeeuw (1996). In this limit, the normalized distribution of mass over $(\eta, R_{\text{zvc}}/R_c)$ at fixed energy, which we will denote as $h(\eta, R_{\text{zvc}}/R_c)$, is a known function that is independent of energy. An explicit expression for $h(\eta, R_{\text{zvc}}/R_c)$ is derived in Appendix 3.B.2. We use this result to approximate the differential mass density of our test model, restricting ourselves to the case with a $3 \times 10^6 M_\odot$ BH. First, we calculate the differential mass density of the model over energy alone: $G(E) \equiv dM / dE = \int [dM / dE dL_z] dL_z$; where $dM / dE dL_z$ is obtained from equation (3.39). Then, we assume that the distribution of mass

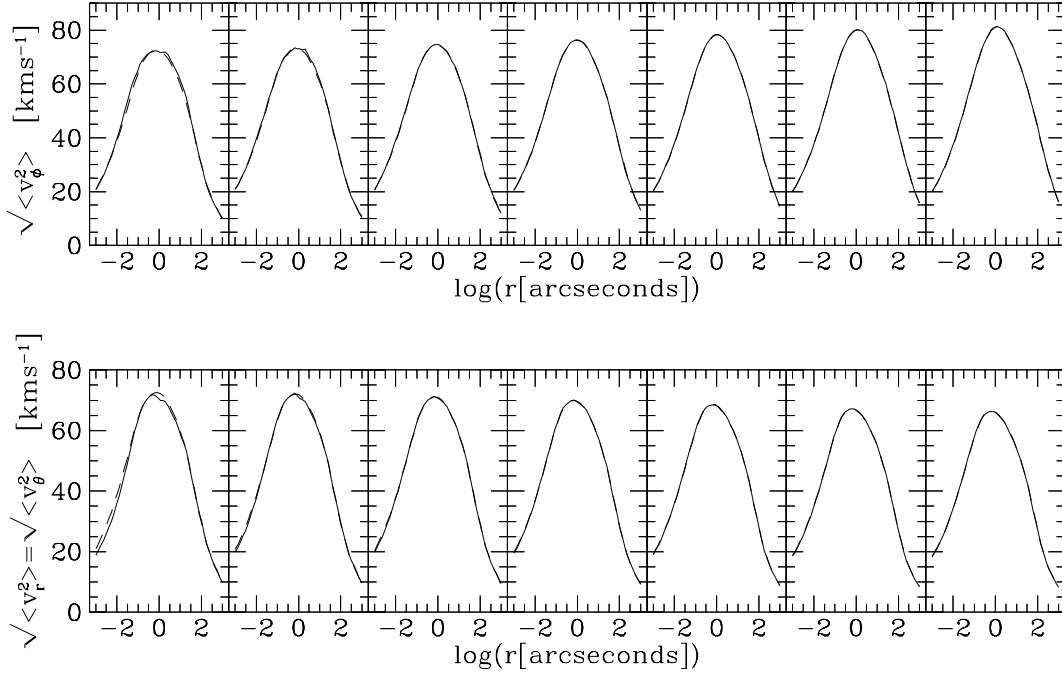


FIGURE 3.9— Comparison of meridional plane velocity moments, calculated either with the extended Schwarzschild technique using only two-integral components (full curves), or with the Jeans equations (dashed curves), for the model of Figure 3.7. The top and bottom row of panels show $\langle v_\phi^2 \rangle^{1/2}$ and $\langle v_r^2 \rangle^{1/2} = \langle v_\theta^2 \rangle^{1/2}$, respectively. The meridional (r, θ) plane is divided in 7 sectors. In each row, the first panel is the sector closest to the symmetry axis and the last panel is the sector closest to the equatorial plane. The discrepancies are largest near the symmetry axis, but are acceptable everywhere.

over (L_z, R_{zvc}) at fixed energy is always the same as in the high-energy limit, so that

$$\frac{dM}{dE d\eta d(R_{zvc}/R_c)} = G(E) h(\eta, R_{zvc}/R_c). \quad (3.19)$$

This relation is only correct at asymptotically high energies. We found it was sufficiently good at energies with $R_c(E) \lesssim 0.5''$, which is where the mass density of the model is a pure power law and where the potential is Keplerian.

For our test we picked an observational test-setup with the same set of 8 square apertures, roughly aligned on the major axis, as used by van der Marel *et al.* (1997b) in their HST observations of M32 (see their Figure 3). These apertures all lie at projected radii $R \lesssim 0.5''$, so most of the light seen in these apertures originates from stars with energies for which our approximation of the differential mass distribution is adequate. We chose the same PSF as in van der Marel *et al.* (1997b), which is a sum of three Gaussians that approximates the HST PSF. Subsequently, we picked a grid in (E, L_z, R_{zvc}) space with $N_E = 20$, $N_\eta = 7$, and $N_{L_3} = 7$, and we calculated an orbit library for this grid. Then finally we calculated orbital weights from equation (3.19), by integrating at each point of our (E, L_z, R_{zvc}) grid the approximation $dM/[dE d\eta d(R_{zvc}/R_c)]$ over the corresponding grid cell. Predictions for the projected kinematics then follow by superposing the VPs for individual orbits in the library as in equation (3.8).

Figure 3.10 shows the results thus obtained for the kinematical quantities $V, \sigma, h_3, \dots, h_6$. As

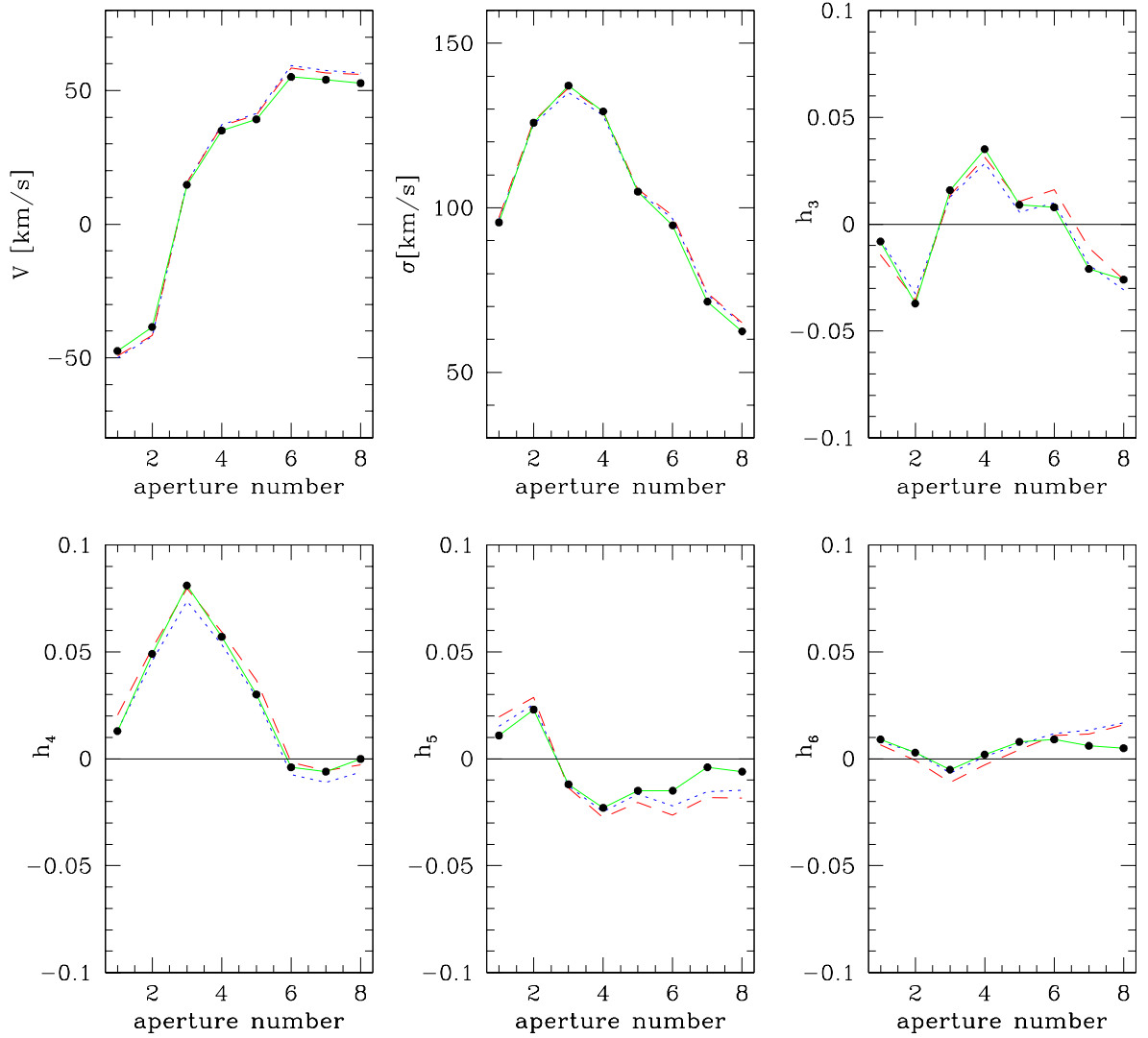


FIGURE 3.10— Kinematical predictions for the edge-on $f(E, L_z)$ test model with a $3 \times 10^6 M_\odot$ BH discussed in Section 3.3.3. The kinematical apertures are the same as for the HST observations of M32 by van der Marel *et al.* (1997b). They are aligned along the major axis. Data points are plotted equidistantly along the abscissa. Dotted and dashed curves are predictions obtained with the extended Schwarzschild technique, using the software of Cretton and van der Marel, respectively. The solid curves show predictions obtained from direct integration over the DF, using the software of Q95. The three curves agree well, demonstrating the numerical accuracy of the orbit superposition technique.

mentioned in the Introduction, independent software implementations of the extended Schwarzschild method were programmed by both N.C. and R.v.d.M. Dotted curves in the figure show the results from N.C.’s software, while dashed curves show the results from R.v.d.M.’s software. For comparison, solid curves show the results obtained by direct integration over the known DF_{HQ} , using the (completely independent) software of Q95 as described in Section 3.3.1. The RMS difference between the different predictions is $\sim 2 \text{ km s}^{-1}$ in V and σ , and ~ 0.01 in the Gauss-Hermite moments. Kinematical data typically have larger observational errors than this, so the numerical accuracy of our method is entirely adequate for modeling real galaxies.

Finally, let us say a few words about the orbits in the library for this test model. Figure 3.11

shows the orbits in the library at the energy corresponding to $R_c = 0.25''$. Figure 3.12 shows the orbits at the same R_c , in the same model, but now without a BH. The orbits are all regular and have a stable periodic parent. The parents were determined using surfaces-of-section (see also: Richstone 1982; Lees & Schwarzschild 1992; Evans 1994; Evans, Häfner & de Zeeuw 1997) and are indicated in the figures. Most orbits in Figures 3.11 and 3.12 are tubes, and are parented by the thin tube. Orbits that are not tubes are indicated. A minority of the low $|\eta|$ orbits in Figure 3.12 is parented by higher-order resonances, such as the 3:2 and 4:3 (being the ratio of the R - and z -frequencies of the parent). By contrast, most of the low $|\eta|$ orbits in Figure 3.11 is parented by the 1:1 resonance. Thus the orbital structure of the models with and without BH is very clearly different. An analysis of the orbital structure of these models as a function of the BH mass is beyond the scope of the present paper, but does seem worth further study.

3.4 Concluding Remarks

In this paper we have described an extension of Schwarzschild’s method for building anisotropic axisymmetric dynamical models of galaxies. We compute a set of orbits in a given mass model and find the non-negative superposition of these orbits that best reproduces a set of (photometric and kinematic) constraints. Our method includes the full VP shape as kinematic constraint. We parametrize the VP using a GH expansion so that it is specified by a few numbers. The modeling method is valid for any kind of parametric (or non-parametric) VP representation and properly takes into account the observational setup (seeing convolution, pixel binning, error on each constraint). We obtain smooth models by imposing a regularization scheme in integral-space.

In chapter 2, we have described in detail several aspects of this method and applied it to the spherical case. However, it is not restricted to this simple geometry, and we have described here the axisymmetric extension. The mass model used to compute the orbit library may be complex: it can have a central density cusp, a stellar disk, a central black hole or an extended dark halo. Applications of our code to the flattened systems M32 and NGC 4342 were presented in chapter 4 and 5 of this thesis.

We have also devised a new semi-analytic method for constructing simpler dynamical models, for which the DF has the special form $DF = f(E, L_z)$ or $DF = f(E)$. These DFs are obtained by using NNLS with analytic building blocks for which the VPs are obtained by one-dimensional quadratures. This technique is general and does not require the density to be expressed analytically as a function of the potential, but can be used with any complex mass model. Previous techniques assuming $DF = f(E, L_z)$ that are also free of this condition include those of Hunter & Qian (1993), Dehnen & Gerhard (1994), Kuijken (1995), Magorrian (1995) and M96b.

We have tested our new method by having it reproduce the properties of $f(E, L_z)$ models for which the DF and projected properties can be calculated with independent algorithms. This allowed us to test all aspects of the superposition method, including the sampling of integral space, orbit calculation, discreteness effects of the spatial grids, projection into the data cubes, seeing convolution, and the NNLS algorithm. Tests with only two-integral components reproduced the DF with a mass-weighted RMS accuracy of $\lesssim 5\%$, and the meridional plane velocity moments to better than 2 km s^{-1} . Tests with a regular orbit library indicated accuracies in the projected quantities of $\sim 2 \text{ km s}^{-1}$ in V and σ , and 0.01 in the GH-moments. All the tests that we have done indicate that the accuracy of our method is adequate for the interpretation of kinematical data obtained with realistic setups.

Our technique can be extended to triaxiality. Several parts of the method will be unaltered for this geometry: the use of projected quantities, e.g., $VP(x', y', v_{\text{los}})$, the fitting procedure, the

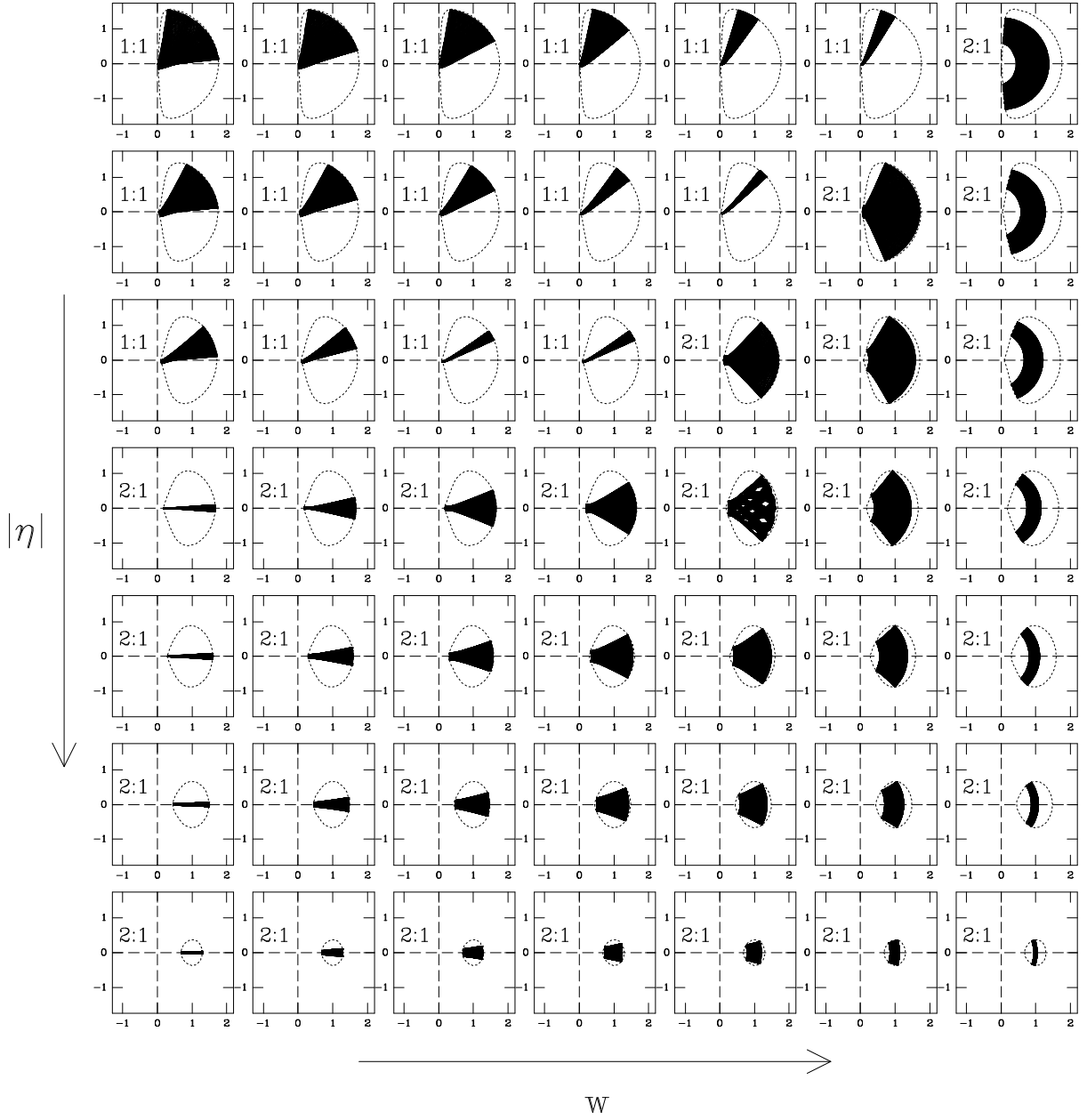


FIGURE 3.11— Examples of orbits at the energy with $R_c = 0.25''$, in the test model of Section 3.3.1 for the case with a $3 \times 10^6 M_\odot$ BH. The axes for each orbit are in units of R_c . Each line corresponds to a different value of $|\eta|$ and each column to a different value of the third integral. The ratio of the R - and z -frequencies for the parent orbit is indicated for those orbits that are not parented by the thin tube.

seeing convolution, etc. However, the orbital structure is much richer than in the axisymmetric case. New orbit families appear (e.g., box orbits) as well as numerous chaotic regions associated with resonances. During the numerical integration of a trajectory in such a mass model, all six phase-space coordinates need to be computed, since there is no azimuthal symmetry. Consequently, the computing overhead is significantly higher for triaxial geometries. Work along these lines is in progress.

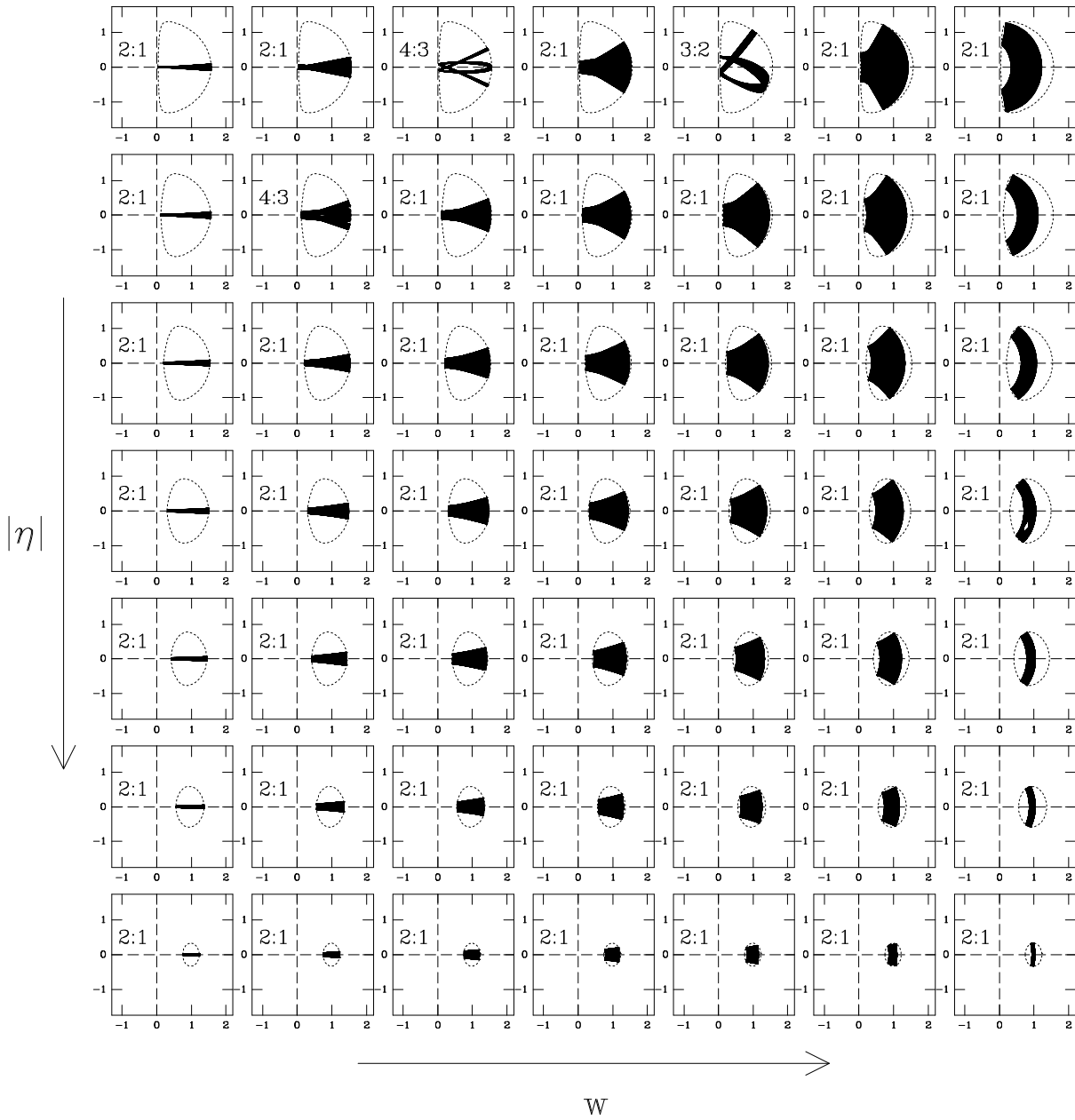


FIGURE 3.12— Similar as Figure 3.11, but now for the same model without a central BH.

Acknowledgments

NC expresses his thanks to Richard Arnold for his help. It is a pleasure to thank Frank van den Bosch, Eric Emsellem, Hong-Sheng Zhao, Yannick Copin, Walter Jaffe and Luis Aguilar for fruitful discussions on the modeling technique. NC acknowledges the hospitality of MPA (Garching) and Steward Observatory, and thanks D. Pfenniger for discussions at initial stages of this project and visits to Geneva Observatory. He was financially supported by a grant from the Swiss government (Etat du Valais) and by an exchange grant from NUFFIC. He and PTdZ acknowledge travel support from the Leids Kerkhoven Bosscha Fonds. RPvdM was supported

by NASA through grant number # GO-05847.01-94A, Hubble Fellowship # HF-1065.01-94A and an STScI Institute Fellowship, all awarded by the Space Telescope Science Institute which is operated by the Association of Universities for Research in Astronomy, Incorporated, under NASA contract NAS5-26555. PTdZ acknowledges the hospitality of the Institute for Advanced Study where this work was initiated.

References

- Bacon, R. *et al.* 1995, *A&AS*, 113, 247
- Binney, J. J., & Merrifield, M. R. 1998, *Galactic Astronomy* (Princeton: Princeton University Press)
- Binney, J. J., & Tremaine, S. D. 1987, *Galactic Dynamics* (Princeton: Princeton University Press) (BT)
- Binney, J. J., Davies, R. L., & Illingworth, G. D. 1990, *ApJ*, 361, 78
- Bishop, J. L. 1987, *ApJ*, 322, 618
- Cretton, N., & van den Bosch, F. C. 1999, *ApJ*, 514, 704, chapter 5 of this thesis
- de Bruijne, J. H. J., van der Marel, R. P., & de Zeeuw P. T. 1996, *MNRAS*, 282, 909
- Dehnen, W. D. 1993, *MNRAS*, 265, 250
- Dehnen, W. D. 1995, *MNRAS*, 274, 919
- Dehnen, W. D., & Gerhard, O. E. 1993, *MNRAS*, 261, 311
- Dehnen, W. D., & Gerhard, O. E. 1994, *MNRAS*, 268, 1019
- Dejonghe, H. 1984, *A&A*, 133, 225
- Dejonghe, H. 1986, *Phys. Rep.*, 133, 217
- Dejonghe, H., & de Zeeuw, P. T. 1988, *ApJ*, 333, 90
- de Zeeuw, P.T. 1997, in *The Nature of Elliptical Galaxies*, Proceedings of the Second Stromlo Symposium, eds. Arnaboldi, M., da Costa, G., & Saha, P., p. 44
- Emsellem, E., Monnet, G., & Bacon, R. 1994, *A&A*, 285, 723
- Emsellem, E., Dejonghe, H., & Bacon, R. 1999, *MNRAS*, 303, 495
- Evans, N. W. 1993, *MNRAS*, 260, 191
- Evans, N. W. 1994, *MNRAS*, 267, 333
- Evans, N. W., & de Zeeuw, P. T. 1994, *MNRAS*, 271, 202
- Evans, N. W., & Häfner R. M., & de Zeeuw, P. T. 1997, *MNRAS*, 286, 315
- Fehlberg, E. 1968, NASA Technical Report R-287
- Franx, M., & Illingworth, G. D. 1988, *ApJ*, 327, L55
- Gerhard, O. E. 1991, *MNRAS*, 250, 812
- Gerhard, O. E., & Binney, J. J. 1996, *MNRAS*, 279, 993
- Gerhard, O. E., Jeske, G., Saglia, R. P., & Bender, R. 1998, *MNRAS*, 295, 197
- Hernquist, L., & Ostriker, J. P. 1992, *ApJ*, 386, 375
- Hunter, C., & de Zeeuw, P. T. 1992, *ApJ*, 389, 79
- Hunter, C., & Qian, E. E. 1993, *MNRAS*, 262, 401 (HQ)
- Innanen, K. P., & Papp, K. A. 1977, *AJ*, 82, 322
- Katz, N., & Richstone, D. O. 1985, *ApJ*, 296, 331
- Kuijken, K. 1995, *ApJ*, 446, 194
- Kuijken, K., & Merrifield, M. R. 1993, *MNRAS*, 264, 712
- Lawson, C. L., & Hanson, R. J. 1974, *Solving Least Squares Problems* (Englewood Cliffs, New Jersey: Prentice-Hall)
- Lees, J. F., & Schwarzschild, M. 1992, *ApJ*, 384, 491
- Levison, H. F., & Richstone, D. O. 1985, *ApJ*, 295, 349

- Levison, H. F., & Richstone, D. O. 1987, *ApJ*, 314, 476
- Magorrian, J., & Binney J. J. 1994, *MNRAS*, 271, 949
- Magorrian, J. 1995, *MNRAS*, 277, 1185
- Matthias, M., & Gerhard, O. E. 1999, *MNRAS*, submitted (astro-ph/9901036)
- Merritt, D. 1993, *ApJ*, 413, 79
- Merritt, D. 1996a, *ApJ*, 486, 102
- Merritt, D. 1996b, *AJ*, 112, 1085 (M96b)
- Merritt, D., & Fridman, T. 1996, *ApJ*, 460, 136
- Merritt, D., & Valluri, M. 1996, *ApJ*, 471, 82
- Morrison, H., Flynn, B., & Freeman, K. C. 1990, *AJ*, 100, 1191
- Ollongren, A. 1962, *Bull. Astron. Inst. Neth.*, 16, 241
- Pfenniger, D. 1984, *A&A*, 141, 171
- Pfenniger, D., & Friedli D. 1993, *A&A*, 270, 561
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, *Numerical Recipes* (Cambridge: Cambridge University Press)
- Qian, E. E., de Zeeuw, P. T., van der Marel, R. P., & Hunter, C. 1995, *MNRAS*, 274, 602 (Q95)
- Richstone, D. O. 1980, *ApJ*, 238, 103
- Richstone, D. O. 1982, *ApJ*, 252, 496
- Richstone, D. O. 1984, *ApJ*, 281, 100
- Richstone, D. O., *et al.* 1997, in *The Nature of Elliptical Galaxies*, Proceedings of the Second Stromlo Symposium, eds. Arnaboldi, M., da Costa, G., & Saha, P., p. 123
- Richstone, D. O., & Tremaine S. 1984, *ApJ*, 286, 27
- Richstone, D. O., & Tremaine S. 1985, *ApJ*, 296, 370
- Richstone, D. O., & Tremaine S. 1988, *ApJ*, 327, 82
- Rix, H-W., & White, S. D. M. 1992, *MNRAS*, 254, 389
- Rix, H-W., de Zeeuw, P. T., Cretton N., van der Marel, R. P., & Carollo, C. M. 1997, *ApJ*, 488, 702, chapter 2 of this thesis
- Rybicki, G. B., 1986, in *Structure and Dynamics of Elliptical Galaxies*, ed. P.T. de Zeeuw (Dordrecht: Reidel), p. 397
- Saaf, A. 1968, *ApJ*, 154, 483
- Schwarzschild, M. 1979, *ApJ*, 232, 236
- Schwarzschild, M. 1982, *ApJ*, 263, 599
- Tremaine S., *et al.* 1994, *AJ*, 107, 634
- van der Marel, R. P., & Franx, M. 1993, *ApJ*, 407, 525 (vdMF)
- van der Marel, R. P., de Zeeuw, P. T., Rix, H. W., & Quinlan, G. D. 1997, *Nature*, 385, 610
- van der Marel, R. P., de Zeeuw, P. T., & Rix, H. W. 1997b, *ApJ*, 488, 119
- van der Marel, R. P., Cretton, N., de Zeeuw, P. T., & Rix, H. W. 1998, *ApJ*, 493, 613, chapter 4 of this thesis
- Vandervoort, P. O. 1984, *ApJ* 287, 475
- Zhao, H. S. 1996a, *MNRAS*, 278, 488
- Zhao, H. S. 1996b, *MNRAS*, 283, 149

3.A Construction of $f(E, L_z)$ - and $f(E)$ -components

3.A.1 Normalization

To determine the normalization coefficients $C_{[E_0, L_z, 0]}$ and $C_{[E_0]}$ in the definitions of the two-integral and isotropic components (eqs. [3.10, 3.11]), we seek expressions for the total mass of a single component. The phase-space volume in cylindrical coordinates is $d^3\vec{x} d^3\vec{v} = R^2 dR d\phi dz$

$d\dot{R} d\dot{\phi} d\dot{z}$. We use

$$\dot{R}^2 = 2[E - \Phi(R, z)] - \frac{L_z^2}{R^2} - \dot{z}^2, \quad \dot{\phi} = L_z/R^2, \quad (3.20)$$

to switch, at fixed (R, z) , from the variables (\dot{R}, \dot{z}) to (E, L_z) . We then find for the total mass of a two-integral component:

$$m_{[E_0, L_{z,0}]} \equiv \int d^3\vec{x} d^3\vec{v} f_{[E_0, L_{z,0}]}^\delta = \int R^2 dR \int d\phi \int dz \int \frac{dE}{\dot{R}} \int \frac{dL_z}{R^2} \int d\dot{z} f_{[E_0, L_{z,0}]}^\delta. \quad (3.21)$$

Upon substitution of $f_{[E_0, L_{z,0}]}^\delta$ from equation (3.10), the integration over ϕ , E and L_z becomes trivial, and we obtain

$$m_{[E_0, L_{z,0}]} = 4\pi C_{[E_0, L_{z,0}]} \iint dR dz \int \left\{ 2[E_0 - \Phi(R, z)] - \frac{L_{z,0}^2}{R^2} - \dot{z}^2 \right\}^{-1/2} d\dot{z}. \quad (3.22)$$

The integral over $d\dot{z}$ extends over the region $|\dot{z}| \leq \dot{z}_{\max}$, where \dot{z}_{\max} is defined as the root of the expression in the square root. Therefore,

$$m_{[E_0, L_{z,0}]} = 8\pi C_{[E_0, L_{z,0}]} \iint dR dz \int_0^{\dot{z}_{\max}} \frac{d\dot{z}}{\sqrt{\dot{z}_{\max}^2 - \dot{z}^2}} = 4\pi^2 C_{[E_0, L_{z,0}]} \iint dR dz, \quad (3.23)$$

where the remaining double integral is over the region for which $E_0 - \Phi(R, z) - (L_{z,0}^2/2R^2) \geq 0$. This is exactly the region $\Phi_{\text{eff},0}(R, z) \leq E_0$ bounded by the ZVC at the given $(E_0, L_{z,0})$, where $\Phi_{\text{eff},0}$ is the effective gravitational potential at the given $L_{z,0}$. To obtain $m_{[E_0, L_{z,0}]} = 1$, we choose

$$C_{[E_0, L_{z,0}]} = \left[4\pi^2 \iint_{\Phi_{\text{eff},0}(R,z) \leq E_0} dR dz \right]^{-1} = \left[2\pi^2 \oint_{\text{ZVC}[E_0, L_{z,0}]} (R dz - z dR) \right]^{-1}, \quad (3.24)$$

where the second equality was obtained with a variant of Stokes' theorem.

Following similar arguments, we obtain for the mass of an isotropic component:

$$\begin{aligned} m_{[E_0]} &\equiv \int d^3\vec{x} d^3\vec{v} f_{[E_0]}^\delta \\ &= 8\pi C_{[E_0]} \iiint dR dz \int dL_z \int_0^{\dot{z}_{\max}} \frac{d\dot{z}}{\sqrt{\dot{z}_{\max}^2 - \dot{z}^2}} = 4\pi^2 C_{[E_0]} \iiint_{\Phi_{\text{eff}}(R,z) \leq E_0} dR dz dL_z, \end{aligned} \quad (3.25)$$

where the effective gravitational potential $\Phi_{\text{eff}}(R, z)$ is a function of L_z , at fixed (R, z) . Evaluation of the integral over dL_z yields

$$m_{[E_0]} = 8\sqrt{2} \pi^2 C_{[E_0]} \iint_{\Phi(R,z) \leq E_0} R \sqrt{E_0 - \Phi(R, z)} dR dz. \quad (3.26)$$

To obtain $m_{[E_0]} = 1$, we choose

$$C_{[E_0]} = \left[8\sqrt{2} \pi^2 \iint_{\Phi(R,z) \leq E_0} R \sqrt{E_0 - \Phi(R, z)} dR dz \right]^{-1}. \quad (3.27)$$

For the special case of a spherical potential, $\Phi = \Phi(r)$, we have

$$m_{[E_0]} = 16\sqrt{2} \pi^2 C_{[E_0]} \int_{\Phi(r) \leq E_0} r^2 \sqrt{E_0 - \Phi(r)} dr, \quad (3.28)$$

which can be recognized as the 'density-of-states' function for an isotropic spherical system (BT). Calculations for spherical components with DFs proportional to $\delta(E - E_0) \delta(L - L_0)$ were presented in Appendix A of chapter 2.

3.A.2 Velocity profiles

We derive here the Jacobian J for the transformation from $(v_{x'}, v_{y'})$ to (E, L_z) , which enters into the expressions for the VPs of two-integral and isotropic components (eqs. [3.12,3.14]). The energy is

$$E = \frac{1}{2}(v_{x'}^2 + v_{y'}^2 + v_{\text{los}}^2) + \Phi(x', y', z'). \quad (3.29)$$

For inclination angle i ,

$$x = -y' \cos i + z' \sin i, \quad z = y' \sin i + z' \cos i, \quad (3.30)$$

and the angular momentum is therefore

$$L_z = xv_y - yv_x = v_{x'}(-y' \cos i + z' \sin i) + v_{y'}(x' \cos i) - v_{\text{los}}(x' \sin i). \quad (3.31)$$

This yields for the Jacobian

$$J = \left| x'v_{x'} \cos i + y'v_{y'} \cos i - z'v_{y'} \sin i \right|^{-1}, \quad (3.32)$$

in which $v_{x'}$ and $v_{y'}$ are functions of E and L_z determined by:

$$v_{x'} = \frac{L_z - v_{y'} x' \cos i + v_{\text{los}} x' \sin i}{-y' \cos i + z' \sin i}, \quad v_{y'}^2 = 2(E - \Phi) - v_{x'}^2 - v_{\text{los}}^2. \quad (3.33)$$

Substitution of $v_{x'}$ in the expression for $v_{y'}^2$ yields a quadratic equation for $v_{y'}$:

$$av_{y'}^2 + bv_{y'} + c = 0, \quad (3.34)$$

where

$$\begin{aligned} a &= (-y' \cos i + z' \sin i)^2 + (x' \cos i)^2, \\ b &= -2(L_z + v_{\text{los}} x' \sin i) x' \cos i, \\ c &= v_{\text{los}}^2 (-y' \cos i + z' \sin i)^2 + (L_z + v_{\text{los}} x' \sin i)^2 - 2(E - \Phi)(-y' \cos i + z' \sin i)^2. \end{aligned} \quad (3.35)$$

Therefore,

$$v_{y'} = \frac{2(L_z + v_{\text{los}} x' \sin i) x' \cos i \pm \sqrt{\Delta}}{2[(-y' \cos i + z' \sin i)^2 + (x' \cos i)^2]}, \quad (3.36)$$

with $\Delta \equiv b^2 - 4ac$. Equations (3.32), (3.33) and (3.36) define the Jacobian J .

3.B Relation Between Orbital Weights and the DF

3.B.1 $dM/dE dL_z$ for an $f(E, L_z)$ model

For a two-integral model there is a simple relation between the differential mass density $dM/dE dL_z$ and the DF $f(E, L_z)$. To derive this relation (see also Vandervoort 1984) we write the trivial identity

$$f(E, L_z) = \iint f(E_0, L_{z,0}) (f_{[E_0, L_{z,0}]}^\delta / C_{[E_0, L_{z,0}]}) dE_0 dL_{z,0}, \quad (3.37)$$

where $f_{[E_0, L_{z,0}]}^\delta$ and $C_{[E_0, L_{z,0}]}$ are as defined in equation (3.10). The total mass of the system is

$$M \equiv \int d^3\vec{x} d^3\vec{v} f(E, L_z) = \iint f(E_0, L_{z,0}) (m_{[E_0, L_{z,0}]} / C_{[E_0, L_{z,0}]}) dE_0 dL_{z,0}, \quad (3.38)$$

where the second identity follows upon substitution of equation (3.37), exchange of the integration order, and use of the definition of $m_{[E_0, L_{z,0}]}$ from equation (3.21). Substitution of equations (3.23) and (3.24), relabeling of the integration variables from $(E_0, L_{z,0})$ to (E, L_z) , and differentiation then yields

$$dM / dE dL_z = f(E, L_z) \times \left[2\pi^2 \oint_{\text{ZVC}[E, L_z]} (R dz - z dR) \right]. \quad (3.39)$$

The mass weight γ_j for a cell in integral space is

$$\gamma_j = \iint_{\text{cell}_j} \frac{dM}{dE dL_z} dE dL_z. \quad (3.40)$$

3.B.2 Scale-free Density in a Kepler Potential

We summarize here the asymptotic case of a scale-free axisymmetric density in a spherical Kepler potential, which was discussed in detail by de Bruijne, van der Marel & de Zeeuw (1996). We adopt the same units as in that paper. In those units, $\rho_* = s^{-\xi}$ and $\Psi = 1/r$, where s is defined as in eq. (3.1) by $s^2 = R^2 + (z/q)^2$, with q the axial ratio. The associated eccentricity is $e = \sqrt{1 - q^2}$. It is convenient to work with the integrals of motion

$$\mathcal{E} = -E, \quad \eta^2 \equiv L_z^2 / L_{\text{max}}^2(\mathcal{E}), \quad \zeta^2 \equiv L^2 / L_{\text{max}}^2(\mathcal{E}), \quad (3.41)$$

where \mathcal{E} is the binding energy, $\eta^2 \in [0, 1]$ and $\zeta^2 \in [\eta^2, 1]$. The unique (even) two-integral DF is

$$\text{DF}(\mathcal{E}, \eta^2) = C_0 g(\mathcal{E}) j(e^2 \eta^2), \quad (3.42)$$

where

$$C_0 = \frac{q^\xi}{2\pi \mathcal{B}(\xi - \frac{1}{2}, \frac{3}{2})}, \quad g(\mathcal{E}) = \mathcal{E}^{\xi - \frac{3}{2}}, \quad j(e^2 \eta^2) = {}_3F_2\left(\frac{\xi}{2}, \frac{\xi+1}{2}, \frac{\xi+2}{2}; \frac{1}{2}, \frac{2\xi-1}{2}; e^2 \eta^2\right). \quad (3.43)$$

The special functions \mathcal{B} and ${}_3F_2$ are the beta-function and a generalized hypergeometric function, respectively. Upon substitution of $\xi = -\alpha$ this yields the $R_c \rightarrow 0$ limit of our test model (Section 3.3.1); upon substitution of $\xi = -\alpha - (\beta\gamma) - (\delta\epsilon)$, it yields the $R_c \rightarrow \infty$ limit.

We wish to calculate the mass weight γ_j contained in a cell number j of integral space (see eq. [3.8]). According to equations (35) and (37) of de Bruijne *et al.*, this is given by:

$$\gamma_j = \int \int \int_{\text{cell}_j} d\mathcal{E} d\eta^2 d\zeta^2 w(\mathcal{E}, \eta^2, \zeta^2) \text{DF}(\mathcal{E}, \eta^2), \quad w(\mathcal{E}, \eta^2, \zeta^2) = \frac{\pi^3}{4} \mathcal{E}^{-5/2} (\eta^2)^{-1/2} (\zeta^2)^{-1/2}. \quad (3.44)$$

In the Kepler potential the binding energy is related to the circular radius according to $\mathcal{E} = 1/(2R_c)$. The ZVC at a given (R_c, η^2) is therefore defined by

$$1 = \frac{2R_c}{r} - \left(\frac{R_c \eta}{R}\right)^2. \quad (3.45)$$

A particle on the ZVC has $v_r = v_\theta = 0$, $L^2 = r^2 v_\phi^2$ and $L_z^2 = R^2 v_\phi^2$. Therefore, we have $\zeta^2 = (r\eta/R)^2$. Combined with the expression for the ZVC this yields

$$\zeta^2 = \left[\frac{2\eta (R_{\text{zvc}}/R_c)}{\eta^2 + (R_{\text{zvc}}/R_c)^2} \right]^2, \quad (3.46)$$

which is a one-to-one relation if (R_{zvc}/R_c) is chosen between η and $1 + \sqrt{1 - \eta^2}$. Substitution of the DF from equation (3.42) into equation (3.44), and transformation to the variables $\log R_c$, $\eta \in [0, 1]$ and $(R_{\text{zvc}}/R_c) \in [\eta, 1 + \sqrt{1 - \eta^2}]$ yields

$$\frac{dM}{d(\log R_c) d\eta d(R_{\text{zvc}}/R_c)} = 2 C_0 \pi^3 \left(\frac{R_c}{2}\right)^{3-\xi} \eta j(e^2 \eta^2) \left\{ \frac{\eta^2 - (R_{\text{zvc}}/R_c)^2}{[\eta^2 + (R_{\text{zvc}}/R_c)^2]^2} \right\}. \quad (3.47)$$

Hence, the normalized distribution of mass over $(\eta, R_{\text{zvc}}/R_c)$ at fixed energy, which we will denote as $h(\eta, R_{\text{zvc}}/R_c)$, is independent of energy. In particular:

$$h(\eta, R_{\text{zvc}}/R_c) = 2\eta j(e^2 \eta^2) \left\{ \frac{\eta^2 - (R_{\text{zvc}}/R_c)^2}{[\eta^2 + (R_{\text{zvc}}/R_c)^2]^2} \right\} / \int_0^1 d\eta^2 j(e^2 \eta^2) \int_{\eta}^{1+\sqrt{1-\eta^2}} d(R_{\text{zvc}}/R_c) \left\{ \frac{\eta^2 - (R_{\text{zvc}}/R_c)^2}{[\eta^2 + (R_{\text{zvc}}/R_c)^2]^2} \right\}. \quad (3.48)$$

Chapter 4

Improved evidence for a black hole in M32

van der Marel, R. P., Cretton, N., de Zeeuw, P. T., Rix, H.-W.
1998, *ApJ*, 493, 613

Axisymmetric dynamical models are constructed for the E3 galaxy M32 to interpret high spatial resolution stellar kinematical data obtained with the Hubble Space Telescope (HST). Models are studied with two-integral, $f(E, L_z)$, phase-space distribution functions, and with fully general three-integral distribution functions. The latter are built using an extension of Schwarzschild's approach: individual orbits in the axisymmetric potential are calculated numerically, and populated using non-negative least-squares fitting so as to reproduce all available kinematical data, including line-of-sight velocity profile shapes. The details of this method are described in companion papers by Rix *et al.* and Cretton *et al.* (chapter 2 and 3 of this thesis).

Models are constructed for inclinations $i = 90^\circ$ (edge-on) and $i = 55^\circ$. No model without a nuclear dark object can fit the combined ground-based and HST data, independent of the dynamical structure of M32. Models with a nuclear dark object of mass $M_{\text{BH}} = 3.4 \times 10^6 M_\odot$ (with 1σ and 3σ error bars of $0.7 \times 10^6 M_\odot$ and $1.6 \times 10^6 M_\odot$, respectively) do provide an excellent fit. The inclined models provide the best fit, but the inferred M_{BH} does not depend sensitively on the assumed inclination. The models that best fit the data are not two-integral models, but like two-integral models they are azimuthally anisotropic. Two-integral models therefore provide useful low-order approximations to the dynamical structure of M32. We use them to show that an extended dark object can fit the data only if its half-mass radius is $r_h \lesssim 0.08''$ ($= 0.26$ pc), implying a central dark matter density exceeding $1 \times 10^8 M_\odot \text{ pc}^{-3}$.

The inferred M_{BH} is consistent with that suggested previously by ground-based kinematical data. However, radially anisotropic axisymmetric constant mass-to-light ratio models are now ruled out for the first time, and the limit on the dark matter density implied by the HST data is now stringent enough to rule out most plausible alternatives to a massive black hole.

The dynamically inferred M_{BH} is identical to that suggested by existing models for HST photometry of M32 that assume adiabatic growth (over a time scale exceeding 10^6 yr) of a black hole into a pre-existing core. The low activity of the nucleus of M32 implies either that only a very small fraction of the gas that is shed by evolving stars is accreted onto the black hole, or alternatively, that accretion proceeds at very low efficiency, e.g., in an advection-dominated mode.

It is generally believed that active galaxies and quasars are powered by the presence of massive black holes (BHs) in their nuclei, and that such BHs are present in many, possibly all, quiescent galaxies as well (see Kormendy & Richstone 1995, Lynden-Bell 1996 and Rees 1996 for reviews of this paradigm and its history). Evidence for this can be derived from studies of the dynamics of stars and gas in the nuclei of individual galaxies. The high spatial resolution data that can now be obtained with the Hubble Space Telescope (HST) allows the existing evidence to be strengthened considerably. The present paper is part of a new HST study of the quiescent galaxy M32, in which the presence of a BH has long been suspected based on the steep central rotation velocity gradient and nuclear peak in the velocity dispersion seen in ground-based data (e.g., Tonry 1987; van der Marel *et al.* 1994a; Bender, Kormendy & Dehnen 1996). The main results of our project were summarized, and discussed in the context of other recent work,

in van der Marel *et al.* (1997a). The acquisition and reduction of the stellar kinematical HST data were described in van der Marel, de Zeeuw & Rix (1997b; hereafter Paper I). Here we present new dynamical models that we have used to interpret the combined HST and ground-based data.

Two types of self-consistent¹ dynamical models have been constructed previously to interpret the ground-based data for M32. Dressler & Richstone (1988) and Richstone, Bower & Dressler (1990) used a method based on Schwarzschild's (1979) technique, in which individual orbits are calculated and superposed, to provide a self-consistent model that fits a given set of data. These 'maximum entropy' (Richstone & Tremaine 1988) models could fit the (then available) data only by invoking the presence of a central dark mass of $(0.7-8) \times 10^6 M_\odot$. The models are general in the sense that they make no assumptions about the dynamical structure of the galaxy. However, a drawback was that only spherical geometry was considered. Even though the models can be made to rotate, it remains unclear what systematic errors are introduced when they are applied to a flattened (E3) galaxy like M32. An alternative approach has been to construct axisymmetric models with phase-space distribution functions (DFs) that depend only on the two classical integrals of motion, $f = f(E, L_z)$, where E is the binding energy and L_z is the angular momentum component along the symmetry axis, both per unit mass. These models properly take flattening and rotation into account. To fit the M32 data, they require the presence of a central dark mass between $1.8 \times 10^6 M_\odot$ (van der Marel *et al.* 1994b; Qian *et al.* 1995, hereafter Q95; Dehnen 1995) and $3 \times 10^6 M_\odot$ (Bender, Kormendy & Dehnen 1996). The disadvantage of these models is that they have a special dynamical structure. The velocity dispersions in the meridional plane are isotropic, $\sigma_r = \sigma_\theta$, which might not be the case in M32. However, the models do fit the observed line-of-sight velocity profile (VP) shapes without invoking freely adjustable parameters, which provides some reason to believe that the M32 DF may not be too different from the form $f(E, L_z)$.

The previous work on M32 has shown that models with a BH can fit the ground-based data, but the modeling has not been general enough to demonstrate that a BH is required.² In fact, the spatial resolution of the ground-based data might have been insufficient for this to be the case (see Appendix 4.A for a discussion of this issue). It has certainly not been sufficient to rule out a cluster of dark objects (as opposed to a central BH) on the basis of theoretical arguments; Goodman & Lee (1989) showed that this requires a resolution of $\lesssim 0.1''$. Our new HST data of the nuclear region of M32 were obtained with the HST Faint Object Spectrograph (FOS) through square apertures of $0.068''$ and $0.191''$, respectively, yielding the highest spatial resolution stellar kinematical data for M32 obtained to date. The results (see, e.g., Figure 4.2 below) show a steeper rotation curve and higher central velocity dispersion than the best ground-based M32 data. The primary goals of our project are to determine whether these new HST data rigorously rule out models without any dark mass, and to what extent they constrain the mass and size of the dark object in M32.

To obtain constraints on the presence of a dark object that are least dependent on a priori assumptions about the DF, we need to compare the HST data not only to the predictions of axisymmetric $f(E, L_z)$ models, but also to the predictions of models with a fully general dynamical structure. Orbit superposition techniques provide the most straightforward approach

¹We use the term 'self-consistent' for models in which the luminous mass density is in equilibrium in the combined gravitational potential due to the luminous mass density and some (known) dark matter density. This definition is broader than the traditional one, which excludes dark matter.

²Dressler & Richstone (1988) and Richstone, Bower & Dressler (1990) argued that their data could not be fit by any spherical model without a BH, but we show in Figure 4.15 below that their data can be fit by an axisymmetric model without a BH.

to construct such models. However, orbit superposition is more difficult to implement for the axisymmetric case than for the spherical case: the orbits are not planar and typically possess an additional integral of motion, so that the orbit library must sample three rather than two integrals of motion. Furthermore, a larger parameter space must be explored because of the unknown inclination angle. This implies that larger amounts of CPU time and computer memory are required. However, other than that, there are no reasons why such models would be infeasible. Motivated by the increased speed and memory capacity of computers, we therefore developed a technique to construct fully general axisymmetric orbit superposition models, that fit any given number of observed photometric and kinematic constraints. Independent software implementations were written by H.-W.R, N.C. and R.v.d.M. Our technique may be viewed as the axisymmetric generalization of the spherical modeling used by Richstone and collaborators, with the important additional feature that we calculate VP shapes and include an arbitrary number of Gauss-Hermite moments in the fit. We take into account the error on each observational constraint to obtain an objective χ^2 measure for the quality-of-fit. Our basic algorithm is described in Rix *et al.* (1997; chapter 2) and summarized in de Zeeuw (1997). Chapter 2 provides an application to the spherical geometry; Cretton *et al.* (1999; chapter 3) present the extension to the axisymmetric case. Here we summarize the main steps of the axisymmetric algorithm briefly, and focus on the application to M32. The resulting models are the most general yet constructed for M32.

The paper is organized as follows. In Section 4.1 we discuss our parametrizations for the stellar mass density and for the potential of the dark object. In Section 4.2 we describe the construction of models with $f(E, L_z)$ DFs, and in Section 4.3 we compare the predictions of these models, both with and without BHs, to the kinematical data. In Section 4.4 we outline the orbit-superposition technique for constructing models with a fully general dynamical structure, and in Section 4.5 we compare the predictions of these models to the data. We construct models with an extended dark object in Section 4.6. We summarize and discuss our main conclusions in Section 4.7. Readers interested primarily in the results of our models may wish to skip Sections 4.1, 4.2 and 4.4.

4.1 Mass density and potential

We adopt a parametrized form for the axisymmetric mass density of M32:

$$\rho(R, z) = \rho_0 (m/b)^\alpha [1 + (m/b)^2]^\beta [1 + (m/c)^2]^\gamma, \quad m^2 \equiv R^2 + (z/q)^2. \quad (4.1)$$

The mass density ρ is related to the luminosity density j according to $\rho = Yj$, where Y is the average mass-to-light ratio of the stellar population (hereafter given in solar V-band units). Both Y and the intrinsic axial ratio q are assumed to be constant (as a function of radius). The projected axial ratio q_p is determined by the inclination i according to $q_p^2 = \cos^2 i + q^2 \sin^2 i$. The parameters Y and i can be freely specified; all other parameters are determined by fitting to the available M32 surface photometry.

The highest spatial resolution surface photometry available for M32 is that presented by Lauer *et al.* (1992), based on pre-COSTAR HST/WFPC images. Their measurements extend to $\sim 4''$ from the nucleus. At larger radii ground-based data are available from Kent (1987) and Peletier (1993). Figure 4.1 shows the major axis surface brightness measurements from these sources. The solid curve shows the surface brightness profile predicted by our model, for $\alpha = -1.435$, $\beta = -0.423$, $\gamma = -1.298$, $b = 0.55''$, $c = 102.0''$, $q_p = 0.73$, $\rho_0 = j_0 Y M_\odot / L_{\odot, V}$, $j_0 = 0.463 \times 10^5 (q_p/q) L_{\odot, V} \text{ pc}^{-3}$, and an assumed distance of 0.7 Mpc. The factor $[1 + (m/c)^2]^\gamma$ in equation (4.1) ensures that the model has finite mass, and that it provides an adequate fit

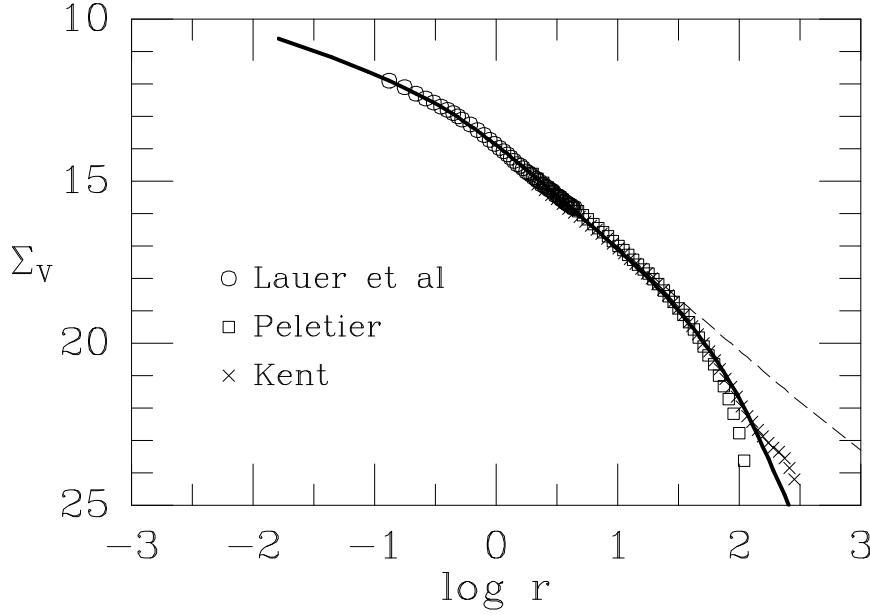


FIGURE 4.1— Data points are observations of the M32 major axis V-band surface brightness from Lauer *et al.* (1992), Peletier (1993) and Kent (1987). The R-band data from Kent and Peletier were transformed to the V-band by assuming a constant V–R color. The differences between the data sets at large radii are due to uncertainties in the sky subtraction. Measurements are not plotted at radii where the PSF introduces large uncertainties ($\sim 0.1''$ for the Lauer *et al.* HST data, $\sim 2''$ for the ground-based data). The solid curve is the brightness profile for the axisymmetric luminous density model used in the present paper. The dashed curve is the profile for the model used by van der Marel *et al.* (1994b) and Q95.

to the observed surface brightness profile out to $\gtrsim 100''$. Apart from this factor, the model is identical to that used by van der Marel *et al.* (1994b) and Q95 (dashed curve in Figure 4.1).

Our model for the mass density is somewhat less general than that used by Dehnen (1995), who deprojected the surface photometry in an unparametrized manner. His approach avoids possible biases resulting from the choice of an *ad hoc* parametrization (Merritt & Tremblay 1994; Gebhardt *et al.* 1996). It also allows the axial ratio of M32 to vary with radius. The observed axial ratio is very close to constant at $q_p = 0.73$ in the central $\sim 10''$, but increases slowly to 0.86 at $\sim 100''$. Even though our model does not reproduce this modest variation, overall it provides an excellent fit, and is fully adequate for a study of the nuclear dynamics. The uncertainties in the interpretation of the kinematic data for the center of M32 are due almost entirely to our ignorance of the dynamical structure of M32. The uncertainties introduced by errors in the brightness profile or by the non-uniqueness of the deprojection are relatively minor (van den Bosch 1997). The effect of possible triaxiality is more difficult to assess, but we will argue in Section 4.7.4 that triaxiality is unlikely to modify any of the major conclusions of our paper.

The gravitational potential is assumed to be $\Psi = \Psi_{\text{lum}} + \Psi_{\text{dark}}$, where Ψ_{lum} is the potential generated by the luminous matter with mass density (4.1), and Ψ_{dark} allows for the possibility of a massive dark object in the nucleus. We assume the latter to be

$$\Psi_{\text{dark}} = GM_{\text{BH}}(r^2 + \epsilon^2)^{-1/2}, \quad (4.2)$$

which is the potential generated by a cluster with a Plummer model mass density (e.g., Binney & Tremaine 1987). For $\epsilon = 0$ one obtains the case of a dark nuclear point mass, i.e., a nuclear

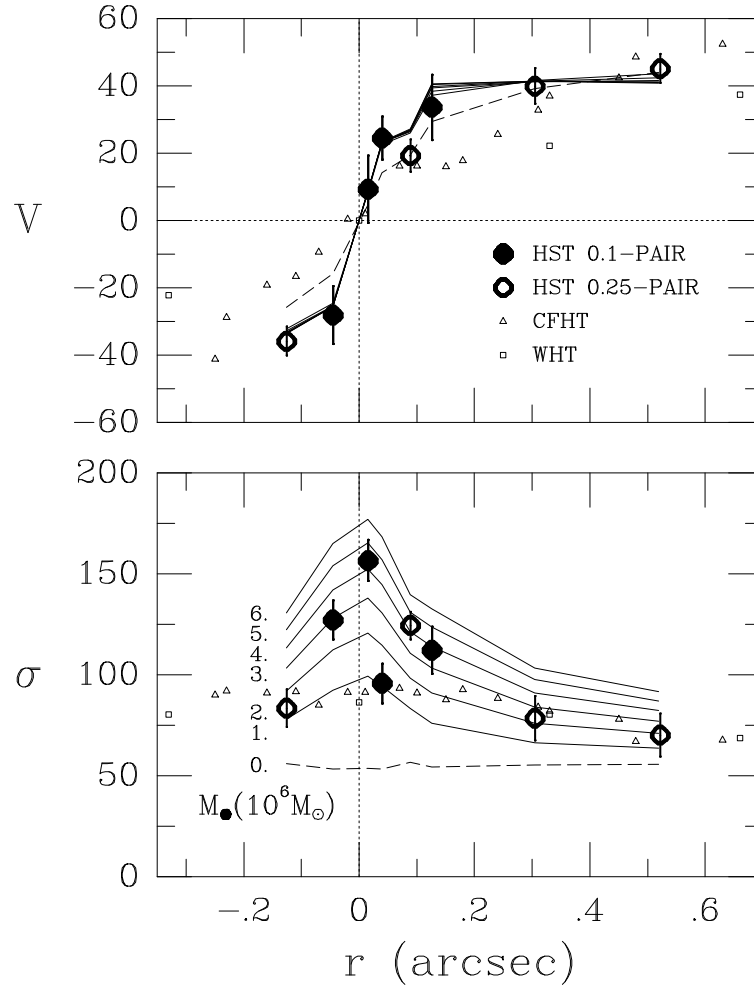


FIGURE 4.2— Data points show the rotation velocities V and velocity dispersions σ derived from the HST/FOS data of Paper I, and from the ground based WHT and CFHT data of van der Marel *et al.* (1994a) and Bender, Kormendy & Dehnen (1996). Errors for the ground-based data are $\sim 1 \text{ km s}^{-1}$ for the WHT data and $\sim 6 \text{ km s}^{-1}$ for the CFHT data, but are not plotted for clarity. The abscissa r is the major axis distance. Curves show the predictions for the HST setup of edge-on $f(E, L_z)$ models with no nuclear dark mass (dashed curves) and with nuclear point masses (BHs) of $1, 2, 3, 4, 5$ and $6 \times 10^6 M_\odot$ (solid curves). Models with $M_{\text{BH}} \approx (3 \pm 1) \times 10^6 M_\odot$ best reproduce the trend in the observed dispersions.

BH. We do not include the potential of a possible dark halo around M32. There are no (strong) observational constraints on the possible presence and characteristics of such a dark halo, and even if present, it will not affect the stellar kinematics near the nucleus of M32.

4.2 Construction of two-integral models

The regular orbits in general axisymmetric potentials are characterized by three integrals of motion, the binding energy $E = \Psi - \frac{1}{2}v^2$, the component of the angular momentum around the symmetry axis $L_z = Rv_\phi$, and a non-classical, or effective, third integral I_3 (Ollongren 1962; Richstone 1982; Binney & Tremaine 1987). In any given axisymmetric potential there is an infinity of DFs $f(E, L_z, I_3)$ that generate a given axisymmetric mass density $\rho(R, z)$. Such models

are difficult to construct, primarily because the third integral cannot generally be expressed explicitly in terms of the phase-space coordinates. However, for any mass density $\rho(R, z)$ there is exactly one DF that is even in L_z , and does not depend on I_3 . This unique even ‘two-integral’ DF, $f_e(E, L_z)$, provides a useful low-order approximation to any axisymmetric model, and has the convenient property that many physical quantities, including the DF itself, can be calculated semi-analytically. We study models of this type for M32 because they have successfully reproduced ground-based M32 data, and because they provide a useful guide for the interpretation of more general three-integral models, which are discussed in Sections 4.4 and 4.5.

To calculate the $f_e(E, L_z)$ DFs for our models we have used a combination of the techniques described in Q95 and Dehnen (1995), even though either technique by itself could have been used to get the same result (in fact, yet another technique to address this problem is described in chapter 3 of this thesis). Initially, four radial regimes are considered: $m \ll b$; $m \approx b$; $m \approx c$; and $m \gg c$, with m and $b \ll c$ as defined in equation (4.1). In these regimes the mass density is approximately: $\rho \propto m^\alpha$; $\rho \propto (m/b)^\alpha [1 + (m/b)^2]^\beta$; $\rho \propto (m/b)^{\alpha+2\beta} [1 + (m/c)^2]^\gamma$; and $\rho \propto m^{\alpha+2\beta+2\gamma}$, respectively. For each of these mass densities $f_e(E, L_z)$ can be calculated with the technique and software of Q95. The DFs for the four regimes are then smoothly patched together in energy, to yield an approximation to the full DF. This approximation is then used as the starting point for Lucy-Richardson iteration as described in Dehnen (1995). This yields the DF $f_e(E, L_z)$, reproducing the model mass density to $\lesssim 0.3$ per cent RMS.

The total DF is the sum of the part f_e that is even in L_z , and the part f_o that is odd in L_z . In principle, f_o is determined completely by the mean streaming velocities $\langle v_\phi \rangle(R, z)$, but these are not determined well enough by the data to make an inversion practicable. Instead, therefore, f_o can be freely specified so as to best fit the data, with the only constraint that the total DF should be positive definite. None of the main conclusions of our paper depend sensitively on the particular parametrization used for f_o , so we restrict ourselves here to a simple choice (Section 4.3.1). Once the complete DF is known, the projected line-of-sight VPs can be calculated for any particular observational setup as in Q95. From the VPs, predictions can be calculated for the observable kinematical quantities.

4.3 Predictions of two-integral models

4.3.1 Data-model comparison

Figure 4.2 shows the HST/FOS data presented in Paper I, obtained with the apertures ‘0.1-PAIR’ (0.068'' square) and ‘0.25-PAIR’ (0.191'' square). The figure also shows the highest available spatial resolution ground-based data, obtained by van der Marel *et al.* (1994a) with the William Herschel Telescope (WHT), and by Bender, Kormendy & Dehnen (1996) with the Canada-France Hawaii Telescope (CFHT). The spatial resolution for these ground-based observations is roughly 0.9'' and 0.5'' FWHM, respectively. The predictions of $f(E, L_z)$ models for the WHT and CFHT data have already been discussed in detail by previous authors, and we therefore focus here on a comparison of $f(E, L_z)$ models to the new HST data.

The curves in Figure 4.2 are the predictions of edge-on $f(E, L_z)$ models with dark nuclear *point masses* (i.e., BHs). These models have an intrinsic axial ratio $q = 0.73$. The dark mass only influences the kinematical predictions in the central few arcsec, and the mass-to-light ratio Υ was therefore chosen to fit the normalization of the kinematical data at larger radii. A good fit to the WHT data between $\sim 5''$ and $\sim 12''$ is obtained with $\Upsilon = 2.51$. Predictions were calculated for each individual observation, taking into account the aperture position, aperture size and PSF for the HST data as given in Paper I. Connecting curves in the figure were drawn to guide the eye. It proved sufficient to study only models with a very simple odd part f_o , namely those that

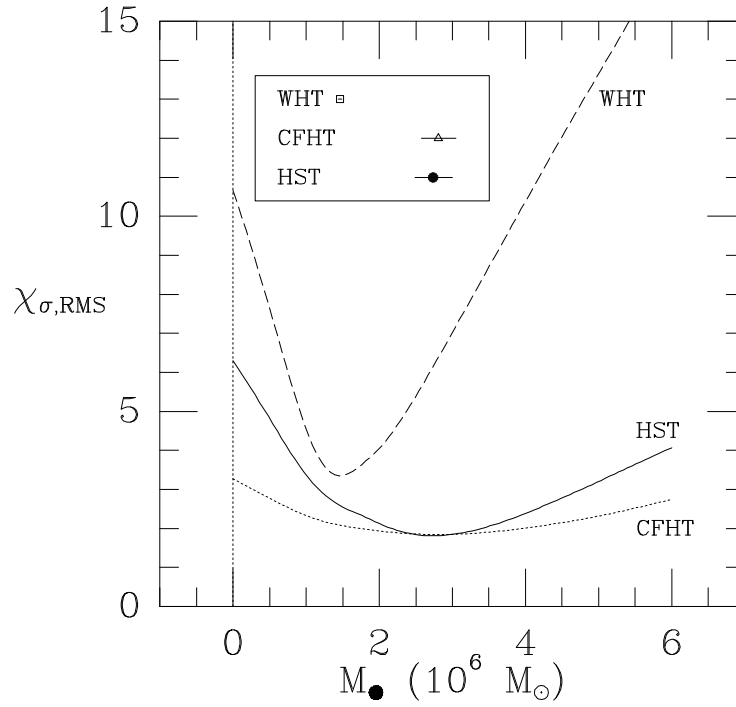


FIGURE 4.3— The relative RMS velocity dispersion residual $\chi_{\sigma, \text{RMS}}$ for edge-on $f(E, L_z)$ models with nuclear point masses, as function of M_{BH} . Results are shown for the HST dispersion measurements of Paper I, and for the ground-based WHT and CFHT measurements. The $\chi_{\sigma, \text{RMS}}$ for the best fit to the WHT data is larger than for the CFHT data, because the WHT data have much smaller errors (such that differences between the predictions and the data are statistically more significant). The symbols in the box mark the position of the best-fitting M_{BH} for each data set and its formal 1σ error bar (the latter is determined not only by the curvature at the $\chi_{\sigma, \text{RMS}}$ minimum, but also by the number of data points, which is different for each data set). Models with $M_{\text{BH}} \approx 3 \times 10^6 M_{\odot}$ best fit the HST and the CFHT data. The WHT data are not well fitted by this M_{BH} . The fact that data of different spatial resolution cannot be fit with the same M_{BH} implies that M32 does not have a DF of the form $f = f(E, L_z)$.

produce a total DF in which at every $(E, |L_z|)$ a fraction F of the stars has $L_z > 0$, and a fraction $(1 - F)$ has $L_z < 0$. For each M_{BH} , the fraction $F \leq 1$ of stars with $L_z > 0$ in the model was chosen to optimize the χ^2 of the fit to the rotation curve. The displayed models *with* a BH fit the rotation curve well, more or less independent of M_{BH} ; models with higher M_{BH} require smaller F . The models differ primarily in their predictions for the velocity dispersions. The observed trend of increasing velocity dispersion towards the nucleus is successfully reproduced by models with a nuclear point mass of $M_{\text{BH}} \approx (3 \pm 1) \times 10^6 M_{\odot}$. The model without a BH predicts a roughly constant velocity dispersion with radius, and is strongly ruled out. In fact, this model also fails to fit the observed rotation velocity gradient in the central arcsec. The displayed model for the no-BH case is maximally rotating ($F = 1$), and it is thus not possible to improve this by choosing a more general form for the odd part of the DF.

For a quantitative analysis of the best-fitting M_{BH} , we define a χ^2 statistic that measures the quality of the model fit to the observed HST velocity dispersions:

$$\chi_{\sigma}^2 \equiv \sum_{i=1}^N \left(\frac{\sigma_{\text{model}} - \sigma_{\text{obs}}}{\Delta\sigma_{\text{obs}}} \right)^2, \quad \chi_{\sigma, \text{RMS}} \equiv [\chi_{\sigma}^2 / N]^{1/2}. \quad (4.3)$$

Figure 4.3 shows the relative RMS residual $\chi_{\sigma, \text{RMS}}$ as function of M_{BH} . The best fit to the HST dispersions is obtained for $M_{\text{BH}} = (2.7 \pm 0.3) \times 10^6 M_{\odot}$. The quoted error is a formal 1σ error based on the assumption of Gaussian statistics (probably an underestimate, as there is some hint

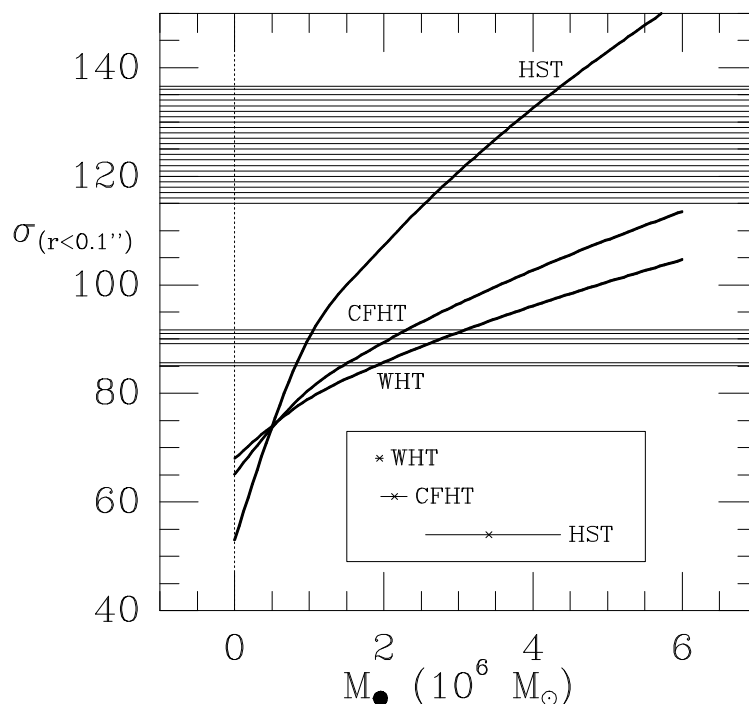


FIGURE 4.4— Horizontally hatched regions show the average velocity dispersion and its 1σ error for those HST, CFHT and WHT observations with the aperture center within $0.1''$ from the M32 nucleus. Curves show the predictions for edge-on $f(E, L_z)$ models with nuclear point masses, as function of M_{BH} . Models with no nuclear dark mass are ruled out under the $f(E, L_z)$ hypothesis. Models with nuclear point masses reproduce the increase in the observed nuclear velocity dispersion with increasing spatial resolution. The horizontal bars in the boxed region indicate for each data set the range of M_{BH} values that predict a nuclear dispersion within the observed hatched region. As in Figure 4.3, the data of different spatial resolution cannot be fit with one single value of M_{BH} .

for systematic errors in the data in addition to random errors). An alternative way of estimating M_{BH} is to model the average of the four data points within $0.1''$ from the center: $\sigma_{r \leq 0.1''} = (126 \pm 10) \text{ km s}^{-1}$, where the 1σ error is based on the scatter between the data points. Figure 4.4 shows the observed $\sigma_{r \leq 0.1''}$ as a hatched region; a solid curve shows the predicted value as a function of M_{BH} . The predictions fall in the observed range for $M_{\text{BH}} = (3.4 \pm 0.9) \times 10^6 M_{\odot}$.

The inclination of M32 cannot be derived from the observed photometry and is therefore a free parameter in the modeling. However, the predictions of $f(E, L_z)$ models are rather insensitive to the assumed inclination (van der Marel *et al.* 1994b). This was verified by also calculating the predictions of inclined models with $i = 55^\circ$, which have an intrinsic axial ratio $q = 0.55$. A mass-to-light ratio $\Upsilon = 2.55$ was adopted, so that at large radii one obtains the same RMS projected velocity on the intermediate axis (between the major and minor axes) as for an edge-on model. On the major axis the $i = 55^\circ$ model then predicts a slightly higher RMS velocity than the edge-on model, resulting in a ~ 10 per cent smaller best-fitting M_{BH} . Apart from this, the conclusions from the inclined models were found to be identical to those for the edge-on models (cf. Figure 4.13 below).

4.3.2 Is the M32 distribution function of the form $f(E, L_z)$?

The $f(E, L_z)$ models that fit the HST data can only be correct if they also fit the ground-based WHT and CFHT data. One may define similar χ^2 quantities as for the HST data, to determine the best-fitting M_{BH} for either of these data sets. Figures 4.3 and 4.4 show the relative

RMS residual $\chi_{\sigma, \text{RMS}}$ of the fit to all the dispersion measurements, and the average dispersion $\sigma_{r \leq 0.1''}$ of the dispersion measurements centered with $0.1''$ from the nucleus. For the WHT data, $\chi_{\sigma, \text{RMS}}$ is minimized for $M_{\text{BH}} = (1.46 \pm 0.03) \times 10^6 M_{\odot}$, whereas for the CFHT data it is minimized for $M_{\text{BH}} = (2.8 \pm 0.2) \times 10^6 M_{\odot}$. The central velocity dispersion measured with the WHT is best fit with $M_{\text{BH}} = (1.94 \pm 0.05) \times 10^6 M_{\odot}$, whereas for the CFHT data it is best fit with $M_{\text{BH}} = (2.1 \pm 0.2) \times 10^6 M_{\odot}$. These results are roughly consistent with those of previous authors. Van der Marel *et al.* (1994b), Q95 and Dehnen (1995) found $M_{\text{BH}} \approx (1.8 \pm 0.3) \times 10^6 M_{\odot}$ for the best fitting edge-on $f(E, L_z)$ model to the ground-based WHT data, while Bender, Kormendy & Dehnen obtained a best fit to their higher spatial resolution CFHT data with $M_{\text{BH}} \approx (3.0 \pm 0.5) \times 10^6 M_{\odot}$. The latter value is somewhat higher than the one we find here, because it was chosen only to provide a good fit to the CFHT rotation curve; it does not fit the CFHT velocity dispersions very well.

These results indicate that, under the assumption of an $f(E, L_z)$ DF, the different observations cannot all be fit simultaneously with the same M_{BH} , even after accounting for the different observational setups. The lowest spatial resolution WHT data require a significantly lower M_{BH} than the highest spatial resolution HST data. This implies that M32 has a DF that is not of the form $f = f(E, L_z)$.

4.4 Construction of three-integral models

To construct more general three-integral models for M32, we extended Schwarzschild's orbit superposition algorithm. Its basic structure is to calculate an orbit library that samples integral space in some complete and uniform way, to store the time-averaged intrinsic and projected properties of the orbits, and to search for the weighted superposition of orbits that best fits the observed kinematics, while reproducing the mass density $\rho(R, z)$ for self-consistency. Here we summarize the main steps, with emphasis on those aspects that are unique to the M32 application. Complete descriptions of the technique are given in chapter 2 and 3.

We sample integral space with an (R_c, η, w) grid. The quantity $R_c(E)$ is the radius of the circular orbit in the equatorial plane with energy E . Its angular momentum, $L_{\text{max}}(E)$, is the maximum angular momentum at the given energy. We define $\eta(E, L_z) \equiv L_z / L_{\text{max}}(E)$. For fixed (R_c, η) , the position of a star in the meridional (R, z) plane is restricted to the region bounded by the 'zero-velocity-curve' (ZVC), defined by the equation $E = \Psi_{\text{eff}}$, where $\Psi_{\text{eff}} = \Psi - \frac{1}{2} L_z^2 / R^2$ is the 'effective gravitational potential' (Binney & Tremaine 1987). We parametrize the third integral at each (R_c, η) using an angle w , which fixes the position at which an orbit touches the ZVC (cf. Figure 4.5).

The quantity R_c was sampled using 20 logarithmically spaced values between $R_{c, \text{min}} = 6.12 \times 10^{-4}$ arcsec, and $R_{c, \text{max}} = 7.55 \times 10^3$ arcsec. This range of radii contains all but a fraction 10^{-4} of the stellar mass of M32. The quantity η was sampled using an 'open' grid (in the same sense that numerical quadrature formulae can be open or closed, e.g., Press *et al.* 1992) of $N_{\eta} = 14$ values, spaced linearly between -1 and 1 , i.e., $\eta_i = -1 + (2i - 1) / N_{\eta}$, for $i = 1, \dots, N_{\eta}$. The quantity w was sampled using an 'open' grid of 7 values, spaced linearly between 0 and w_{th} . Here w_{th} is the angle w for the 'thin tube' orbit at the given (E, L_z) (see Figure 4.5). The special values $\eta = 0$ and $\eta = \pm 1$ (meridional plane and circular orbits) and the special values $w = 0$ and $w = w_{\text{th}}$ (equatorial plane and thin tube orbits) are presumed to be represented by their closest neighbors on the grid, but are not included explicitly.

An orbit was integrated for each (R_c, η, w) combination, starting with $v_R = v_z = 0$ from the ZVC. The integration time was 200 times the period of the circular orbit at the same energy. This is sufficient to properly sample the phase-space trajectory for the large majority of orbits,

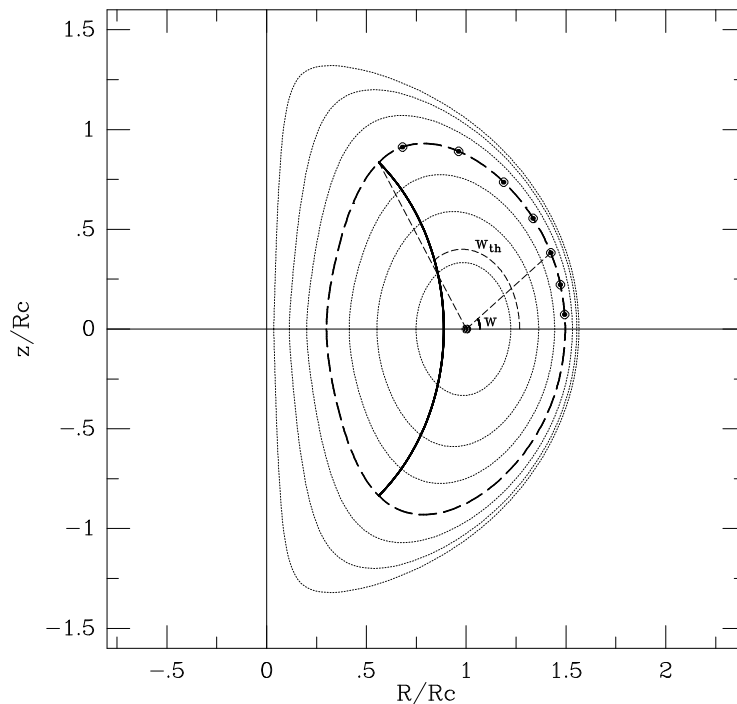


FIGURE 4.5— Example of the meridional plane at a fixed energy E . The axes are in units of $R_c(E)$, the radius of the circular orbit at the given energy. The library in our orbit-superposition models uses an open grid of 7 values of $|\eta| \in [0, 1]$, where $\eta \equiv L_z/L_{\max}(E)$; orbits for $L_z > 0$ and $L_z < 0$ are identical, except for a reversal of the three-dimensional velocity vector at each phase-point. The oval curves are the zero-velocity curves (ZVCs) for each $|\eta|$. Positions on the ZVC can be parametrized using the angle w . Because every orbit with $L_z \neq 0$ touches the ZVC (Ollongren 1962), all orbits at a given (E, L_z) can be sampled by starting stars with $v_R = v_z = 0$ from the ZVC (but $v_\phi \neq 0$ for $L_z \neq 0$). The ‘thin tube’ orbit (heavy solid curve, for the case of the highlighted ZVC) is the only orbit that touches the ZVC at only a single value of $|w|$, referred to as w_{th} . All other orbits touch the ZVC at at least two values of $|w|$, one smaller than w_{th} and one larger than w_{th} (see chapter 3 for examples of actual orbits). Orbits with $w \in [-\pi, 0]$ follow trivially from those with $w \in [0, \pi]$ upon multiplication of (z, v_z) by -1 , at each phase-point along the orbit. It is therefore sufficient to consider only orbits with $w \in [0, w_{\text{th}}]$. We sample this range using an open grid of 7 values (indicated by the dots, for the case of the highlighted ZVC). The figure shows the meridional plane at the energy for which $R_c(E) = 0.25''$, in the edge-on model with $M_{\text{BH}} = 3 \times 10^6 M_\odot$ and $\Upsilon = 2.51$. The ZVCs at other energies and in other models differ in the details, but are topologically similar.

although it exceeds the Hubble time only at radii $\gtrsim 100''$. The integrations yield the ‘orbital phase-space density’ for each orbit, as described in chapter 3. These were binned onto: (i) an (r, θ) grid in the meridional plane; (ii) an (r', θ') grid on the projected (x, y) plane of the sky; and (iii) several Cartesian (x, y, v) cubes, with v the line-of-sight velocity. The first two grids were chosen logarithmic in r, r' , with identical bins to those used for R_c , and linear in $\theta, \theta' \in [0, \frac{\pi}{2}]$. The (x, y, v) cubes were centered on $(0, 0, 0)$, with 211×211 square spatial cells, and 91 velocity bins of 15 km s^{-1} . Spatial cell sizes were adopted of $0.025'', 0.08''$ and $0.5''$, respectively. The (x, y, v) cubes were used to calculate, for each orbit, the predicted line-of-sight velocity histograms for all positions and setups for which kinematical data are available, taking into account the observational point-spread-functions (PSFs) and aperture positions, orientations and sizes, as described in chapter 3.

Construction of a model consists of finding a weighted superposition of the orbits in the library that reproduces two sets of constraints:

- Self-consistency constraints. The model should reproduce the masses predicted by the luminous density $\rho(R, z)$ (Section 4.1), for each cell of the meridional (r, θ) grid, for each

cell of the projected (r', θ') grid, and for each aperture on the sky for which there is data³.

- Kinematical constraints. The model should reproduce the observed kinematics of the galaxy, including VP shapes. We chose to express this as a set of linear constraints on the Gauss-Hermite moments h_i of the VPs (chapter 2).

Smoothness of the solutions in integral space may be enforced by adding extra regularization constraints (see Section 4.5.3 below). The quality of the fit to the combined constraints can be measured through a χ^2 quantity (chapter 2). The assessment of the fit to the kinematical constraints includes the observational errors. In principle, one would like to fit the self-consistency constraints with machine precision. In practice this is unfeasible, because the projected mass constraints are not independent from each other (aperture positions for different data sets partly overlap) and from the meridional plane mass constraints. It was found that models with no kinematical constraints could at best fit the masses with a fractional error of $\sim 5 \times 10^{-3}$. Motivated by this, fractional ‘errors’ of this size were assigned to all the masses in the self-consistency constraints⁴. As described in chapter 2 and 4, we use the NNLS routine of Lawson & Hanson (1974) to determine the combination of non-negative orbital occupancies (which need not be unique) that minimizes the combined χ^2 .

The model predictions have a finite numerical accuracy, due to, e.g., gridding and discretization. Tests show that the numerical errors in the predicted kinematics of our method are $\lesssim 2 \text{ km s}^{-1}$ for the rotation velocities and velocity dispersions, and $\lesssim 0.01$ in the Gauss-Hermite moments (chapter 3). Numerical errors of this magnitude have only an insignificant effect on the data-model comparison (see Appendix 4.A).

4.5 Predictions of three-integral models

4.5.1 Implementation

We have studied three-integral models with dark central *point*-masses, i.e., $\epsilon = 0$ in equation (4.2). There are then three free model parameters: the inclination i , the mass-to-light ratio Y , and the BH mass M_{BH} . The parameter space must be explored through separate sets of orbit libraries⁵. As for the two-integral models, we have studied only two, widely spaced, inclinations: $i = 90^\circ$ and $i = 55^\circ$. For each inclination we have sampled the physically interesting range of (Y, M_{BH}) combinations. For each (Y, M_{BH}) combination we determined the orbital weights (and hence the dynamical structure) that best fit the data, and the corresponding goodness-of-fit quantity $\chi^2(Y, M_{\text{BH}})$. All available HST/FOS, CFHT and WHT data were included as kinematical constraints on the models. Older kinematical data for M32 were not included because of their lower spatial resolution and/or poorer sky coverage. In total, each NNLS fit had 1960 orbits to fit 782 constraints: 366 self-consistency constraints, and 416 kinematical constraints (for 86 positions on the projected plane of the galaxy). We focus primarily (Figures 4.6–4.8) on models without additional constraints that enforce smoothness in integral-space.

³In theory, it is sufficient to fit only the meridional plane masses. Projected masses are then fit automatically. In practice this is not exactly the case, because of discretization. Projected masses were therefore included as separate constraints.

⁴In principle one would like to include the observational surface brightness errors in the analysis. Unfortunately, this requires the exploration of a large set of three-dimensional mass densities (that all fit the surface photometry to within the errors), which is prohibitively time-consuming. However, the observational errors in the surface brightness are small enough that they are not believed to influence the conclusions of our paper.

⁵Only one orbit library needs to be calculated for models with the same M_{BH}/Y . The potentials of such models are identical except for a normalization factor, and the orbits are therefore identical except for a velocity scaling. Each (Y, M_{BH}) combination does require a separate NNLS fit to the constraints.

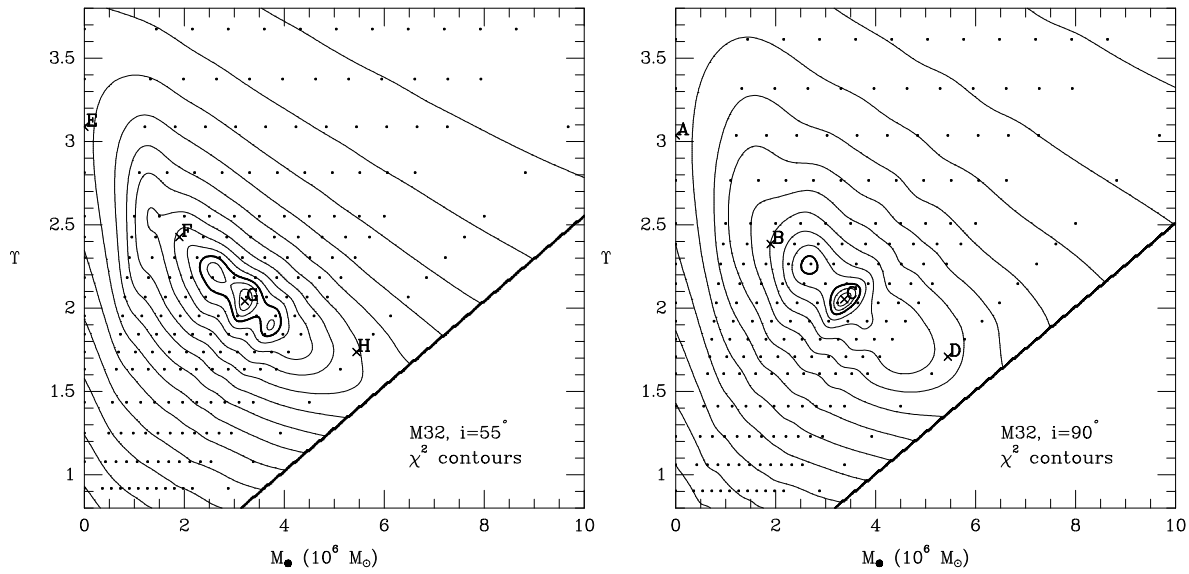


FIGURE 4.6— Contour plots of the χ^2 that measures the quality of the fit to the combined HST, CHFT and WHT data, for orbit-superposition models with $i = 90^\circ$ and $i = 55^\circ$. The model parameters along the abscissa and ordinate are the BH mass M_{BH} and the (V-band) mass-to-light ratio Υ , respectively. The dots indicate models that were calculated, the contours were obtained through spline interpolation (the first three contours define the formal 68.3, 95.4 and (heavy contours) 99.73 per cent confidence regions; subsequent contours are characterized by a factor two increase in $\Delta\chi^2$). The bottom right corner in each panel is a region for which no models were calculated. The labeled positions define models that are discussed in detail in the text and in subsequent figures and in Table 4.1. Models C and G provide the overall best fits.

This is a sufficient and conservative approach for addressing the primary question of our paper: *which models are ruled out by the M32 data?* If the data cannot even be fit with an arbitrarily unsmooth DF, they certainly cannot be fit with a smooth DF.

4.5.2 Data-model comparison

Figure 4.6 shows the main result: contour plots of $\chi^2(\Upsilon, M_{\text{BH}})$ for both inclinations that were studied. The displayed χ^2 measures the quality of the fit to the kinematical constraints only; the actual NNLS fits were done to both the kinematical and the self-consistency constraints, but contour plots of the total χ^2 look similar. The overall minimum χ^2 values are obtained for: $\Upsilon = 2.1$ and $M_{\text{BH}} = 3.4 \times 10^6 M_\odot$ for $i = 90^\circ$; and $\Upsilon = 2.0$ and $M_{\text{BH}} = 3.2 \times 10^6 M_\odot$ for $i = 55^\circ$.

Figure 4.7 compares the kinematical predictions of the best-fitting edge-on and $i = 55^\circ$ models to the data. Two problems with the data must be taken into account when assessing the quality of the fit. First, the HST velocity dispersions show a scatter between some neighboring points that is much larger than the formal errors, most likely due to some unknown systematic effect. The models cannot be expected to reproduce this. Second, the CFHT rotation velocities at radii $\gtrsim 0.5''$ exceed the WHT measurements by an amount which cannot be attributed to differences in spatial resolution, but must be due to some unknown systematic error in either of the two data sets. The WHT data have the smallest error bars, and therefore receive most weight in the NNLS fit. As a result, the models tend to underpredict the CFHT rotation curve.

These systematic problems with the data preclude the use of χ^2 as a meaningful statistic to assess which models provide an acceptable fit: if the observations themselves are not mutually consistent, then clearly no model can be statistically consistent with all of them. Although the use of any statistical test is suspect in the presence of systematic errors, one may still assign

confidence regions on the model parameters by using the *relative likelihood* statistic $\Delta\chi^2 \equiv \chi^2 - \chi_{\min}^2$. This statistic merely measures which parameter combinations provide an equally good (or bad) fit as the one(s) that yield the minimum χ^2 . If we assume that the observational errors are normally distributed (which, as mentioned, is likely to be an oversimplification), then $\Delta\chi^2$ follows a χ^2 probability distribution with the number of degrees of freedom equal to the number of model parameters (Press *et al.* 1992)⁶.

The best-fitting edge-on model in Figure 4.7 has $\chi^2 = 690.0$ while the best-fitting $i = 55^\circ$ model has $\chi^2 = 602.5$, both for $N = 416$ degrees of freedom. The fact that $\chi^2 > N$, even for these optimum fits, is due primarily to the systematic errors in the data. To the eye, the models appear to fit the data as well as could be hoped for. The χ^2 values do indicate that the $i = 55^\circ$ model provides a significantly better fit than the edge-on model, implying that M32 is not seen edge-on. However, the results presented here do not allow us to derive the actual inclination of M32. That would require a detailed study of the entire range of possible inclinations, which would be more computer-intensive. The important conclusion in the present context is that the topology of the χ^2 contours in Figure 4.6 is virtually identical for both inclinations: the allowed range for M_{BH} is therefore uninfluenced by our ignorance of the true inclination of M32.

The $\Delta\chi^2$ statistic was used to assign confidence values to the contours in Figure 4.6. At the 68.3 per cent confidence level (1σ for a Gaussian probability distribution), the allowed M_{BH} fall in the range $(3.2\text{--}3.5) \times 10^6 M_\odot$ for $i = 90^\circ$, and in the range $(3.1\text{--}3.4) \times 10^6 M_\odot$ for $i = 55^\circ$. At the 99.73 per cent confidence level (3σ for a Gaussian probability distribution), they fall in the ranges $(2.5\text{--}3.7) \times 10^6 M_\odot$ and $(2.3\text{--}3.9) \times 10^6 M_\odot$, respectively. In reality, small numerical errors in the models might have distorted the χ^2 contours. We address this issue in Appendix 4.A. Any numerical errors are small enough that they have no influence on our conclusion that models without a dark mass are firmly ruled out. However, the possibility of small numerical errors does increase the confidence bands on M_{BH} . Based on the analysis in Appendix 4.A, we conclude that $M_{\text{BH}} = (3.4 \pm 0.7) \times 10^6 M_\odot$ at 68.3 per cent confidence, and $M_{\text{BH}} = (3.4 \pm 1.6) \times 10^6 M_\odot$ at 99.73 per cent confidence. These estimates take into account both the observational errors in the data and possible numerical errors in the models, and are valid for both inclinations that were studied.

Eight selected models in Figure 4.6 are labeled as A–H, and are listed in Table 4.1. Figure 4.8 compares the model predictions to the observed rotation velocities and velocity dispersions for all the models labeled in Figure 4.6. Models C&G are the overall best fits for the two inclinations. Models B&F and D&H are (approximately) the best-fitting models for $M_{\text{BH}} = 1.9$ and $5.4 \times 10^6 M_\odot$, respectively. The latter models are marginally ruled out at the $> 99\%$ confidence level (cf. the above discussion), although to the eye they do appear to reproduce the main features of the data. They differ from the overall best-fitting models primarily in their predictions for the HST velocity dispersions. The differences in the predictions for the ground-based data are smaller (and invisible to the eye in Figure 4.8), but nonetheless more statistically significant because of the smaller error bars for these data. Models A&E, the best fits without a central dark mass, are indisputably ruled out. The main problem for these models is to fit the central peak in the velocity dispersion. They come rather close to fitting the WHT observations, and predict a central dispersion of $\sim 84 \text{ km s}^{-1}$. However, the models without a dark mass fail to reproduce the higher central dispersion of $\sim 91 \pm 2 \text{ km s}^{-1}$ measured with the CFHT (although

⁶A more robust way to incorporate the effects of *random* errors in the assignment of confidence bands would be to use ‘bootstrapping’, in which one directly calculates the statistical distribution of model parameters by finding the best-fit parameter combinations for different ‘realizations’ of the data set. Unfortunately, this is computationally infeasible in the present context: even the analysis of the single (available) data set for M32 already takes weeks of CPU time on a high-end workstation.

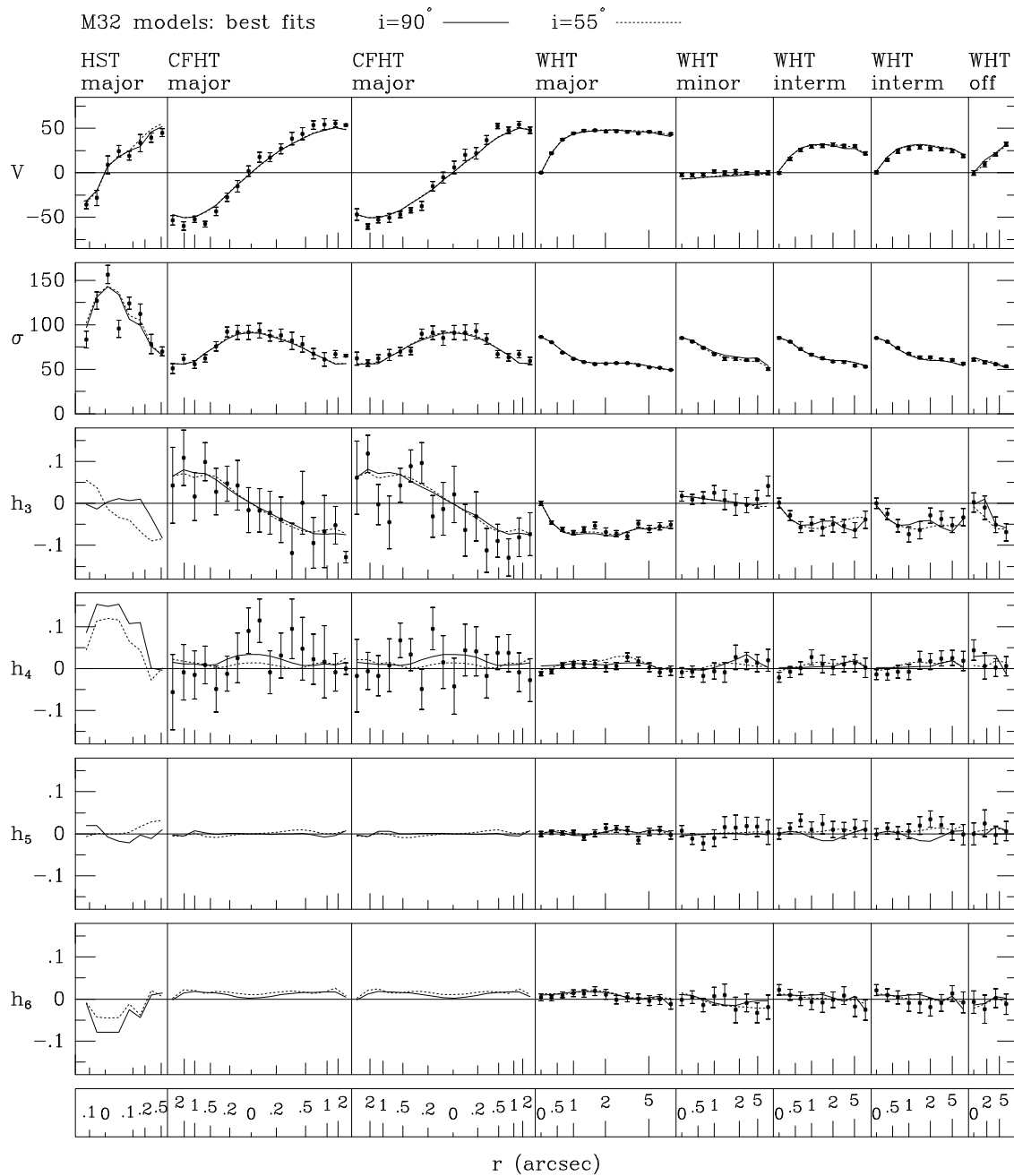


FIGURE 4.7— Predictions of the best-fitting orbit-superposition models for $i = 90^\circ$ and $i = 55^\circ$ (labeled C and G in Figure 4.6 and Table 4.1), compared to the kinematical HST, CFHT and WHT data. The models have nuclear BHs of $3.4 \times 10^6 M_\odot$ and $3.2 \times 10^6 M_\odot$, respectively. The HST data have the highest spatial resolution, and were taken with a set of apertures aligned along the major axis. The ground-based data are long-slit measurements. For the CFHT observations two independent sets of major axis data are available with a similar setup. For the WHT observations data are available with the slit along the major axis, minor axis, two intermediate axes (major $\pm 45^\circ$), and an axis parallel to the major axis but $4''$ offset from it. Shown from top to bottom are: rotation velocities, velocity dispersions, and the Gauss-Hermite moments (when available) that measure deviations of the line-of-sight VP shapes from a Gaussian. The data points are arranged equidistantly along the abscissa. The corresponding distance from the nucleus in arcsec is illustrated schematically in the bottom panel. The WHT data were analyzed by averaging spectra at positive and negative radii, so for these data only positive radii are shown. The model fits to the data are excellent.

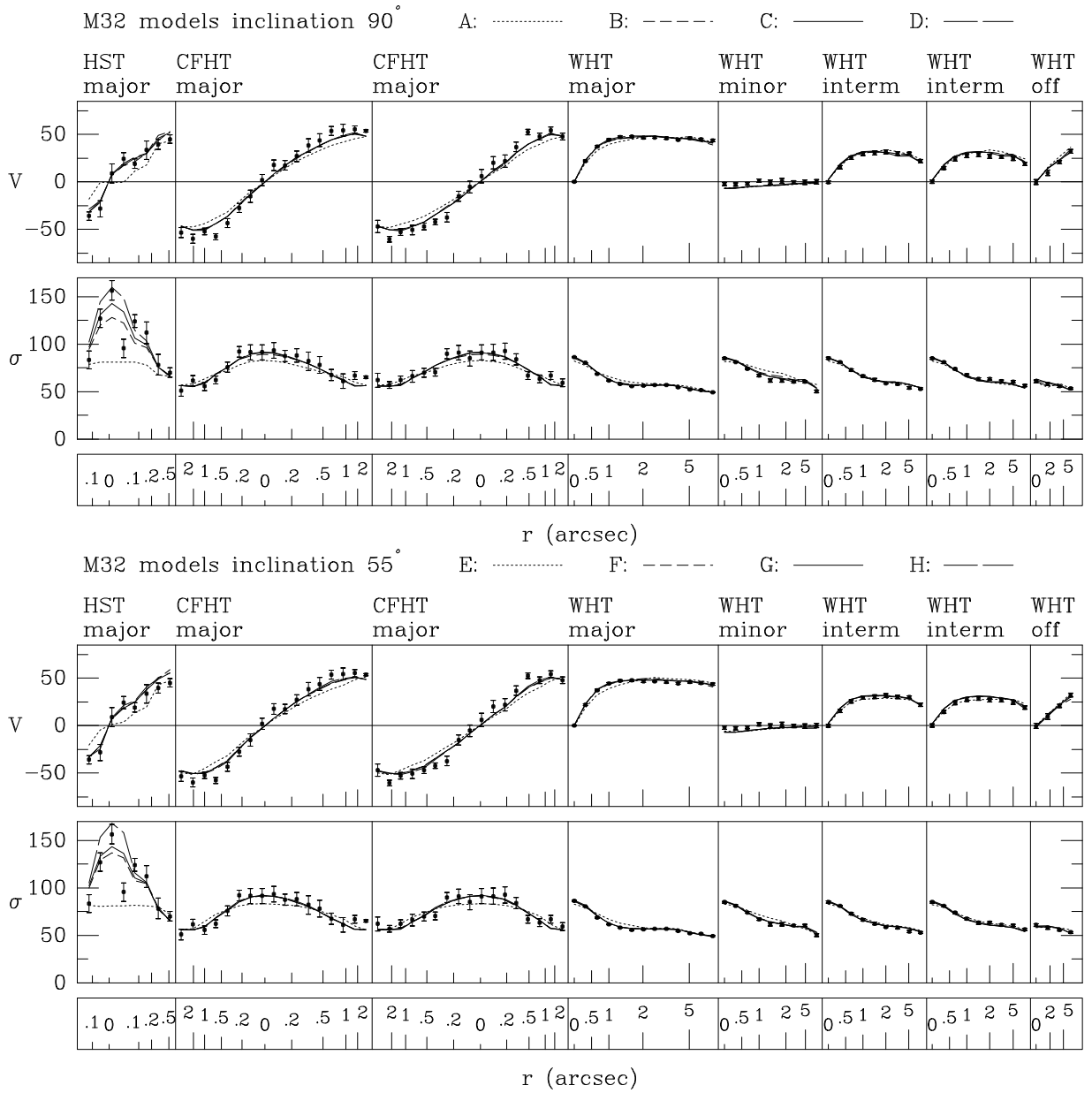


FIGURE 4.8— Predicted rotation velocities and velocity dispersions for the orbit-superposition models A–D (top panel) and E–H (bottom panel) defined in Figure 4.6 and Table 4.1. Models C&G are the overall best fits, models A&E are the best fits without a central dark mass, and models B&F and D&H are (approximately) the best-fitting models for $M_{\text{BH}} = 1.9$ and $5.4 \times 10^6 M_\odot$, respectively. The data are as in Figure 4.7. The models without a BH manage to fit the WHT data reasonably well, but are firmly ruled out by the HST data.

still only marginally), and don't even come close to reproducing the HST dispersions, which exceed 100 km s^{-1} in the central $0.1''$.

4.5.3 Smooth solutions

The top row in Figure 4.9 illustrates the orbital occupancies in integral space for the best-fitting edge-on model (model C). Only a small fraction of the orbits in the library is used to fit the constraints, while the remainder receives zero weight. This yields an equilibrium solution of

the collisionless Boltzmann equation, but is not physically plausible.

Smoothness of the solutions in integral space can be enforced by adding linear regularization constraints to the problem (Zhao 1996; chapter 3). We have explored this only in an *ad hoc* way, merely to be able to assess the effect of smoothness constraints on the resulting fit to the data. A model is defined as a set of masses $m(R_c, \eta, w)$ in integral space. For each point that is not on the boundary of the (R_c, η, w) grid, we measure the smoothness of the model (Press *et al.* 1992; eq. [18.5.10]) through the second order divided differences (in each of the three variables, assuming for simplicity that the distances between adjacent grid points are equal in all directions) of the function $m(R_c, \eta, w)/m_0(R_c)$. The function $m_0(R_c)$ is a rough approximation to the energy dependence of the model, obtained, e.g., by studying the spherical isotropic limit of the given mass density. The regularization constraints are then that the divided differences should equal $0 \pm \Delta$, where the ‘error’ Δ determines the amount of smoothing. Models with $\Delta \rightarrow \infty$ have no smoothing, while models with $\Delta \rightarrow 0$ force $m(R_c, \eta, w)/m_0(R_c)$ to be a linear function on the (R_c, η, w) grid.

The second and third rows in Figure 4.9 show the integral space for model C with the addition of either a modest ($\Delta = 5$) or a large ($\Delta = 0.2$) amount of regularization in the NNLS fit, respectively. As the bottom panels show, the price paid for the increased smoothness is a somewhat poorer fit to the data. However, the fits are still quite good. This demonstrates that the good fits to the data shown in Figure 4.7 are not primarily the result of the use of implausible distributions in integral space. These distributions result from the numerical properties of the problem, but there also exist smooth solutions which provide similar fits.

4.5.4 Dynamical Structure

Figure 4.10 shows the components of the second velocity moment tensor as function of radius, for the edge-on models A–D and the $i = 55^\circ$ models E–H. Table 4.1 lists for two radial ranges the average RMS velocities in km s^{-1} in each of the spherical coordinate directions, and also includes the azimuthal dispersion $\sigma_\phi \equiv [\langle v_\phi^2 \rangle - \langle v_\phi \rangle^2]^{1/2}$.

TABLE 4.1— Properties of selected models for M32

Model (1)	i deg (2)	M_{BH} $10^6 M_\odot$ (3)	Υ (4)	$0.07'' \leq r \leq 0.9''$				$0.9'' \leq r \leq 12''$			
				$\langle v_\phi^2 \rangle^{1/2}$ km s^{-1} (5)	σ_ϕ km s^{-1} (6)	σ_θ km s^{-1} (7)	σ_r km s^{-1} (8)	$\langle v_\phi^2 \rangle^{1/2}$ km s^{-1} (9)	σ_ϕ km s^{-1} (10)	σ_θ km s^{-1} (11)	σ_r km s^{-1} (12)
A	90	0.0	3.0	80	64	81	93	78	61	66	66
B	90	1.9	2.4	91	74	83	81	73	55	63	60
C	90	3.4	2.1	98	79	86	80	71	54	60	58
D	90	5.4	1.7	109	87	94	82	71	53	59	52
E	55	0.0	3.1	91	65	80	91	84	65	65	66
F	55	1.9	2.4	98	70	77	81	82	59	60	57
G	55	3.2	2.0	112	83	82	89	88	67	63	48
H	55	5.4	1.7	114	86	80	88	78	55	56	53

Column (1) lists the model label as indicated in Figure 4.6. Column (2) lists the inclination, column (3) the dark point mass M_{BH} , and column (4) the average mass-to-light ratio Υ of the stellar population (in solar V-band units). These parameters determine the gravitational potential of each model. For each potential, orbit-superposition was used to determine the dynamical structure that best fits the combined HST and ground-based data for M32. Columns (5)–(12) summarize the velocity ellipsoid shapes of the resulting models for two radial ranges. Columns (5)–(8) list the mass-weighted average values of $\langle v_\phi^2 \rangle^{1/2}$, σ_ϕ , σ_θ and σ_r in the radial range $0.07'' \leq r \leq 0.9''$. Columns (9)–(12) list the same quantities for the radial range $0.9'' \leq r \leq 12''$. The azimuthal dispersion is defined as $\sigma_\phi \equiv [\langle v_\phi^2 \rangle - \langle v_\phi \rangle^2]^{1/2}$. As discussed in the text, the models A and E are ruled out by the M32 data, while the best fits are models C and G.

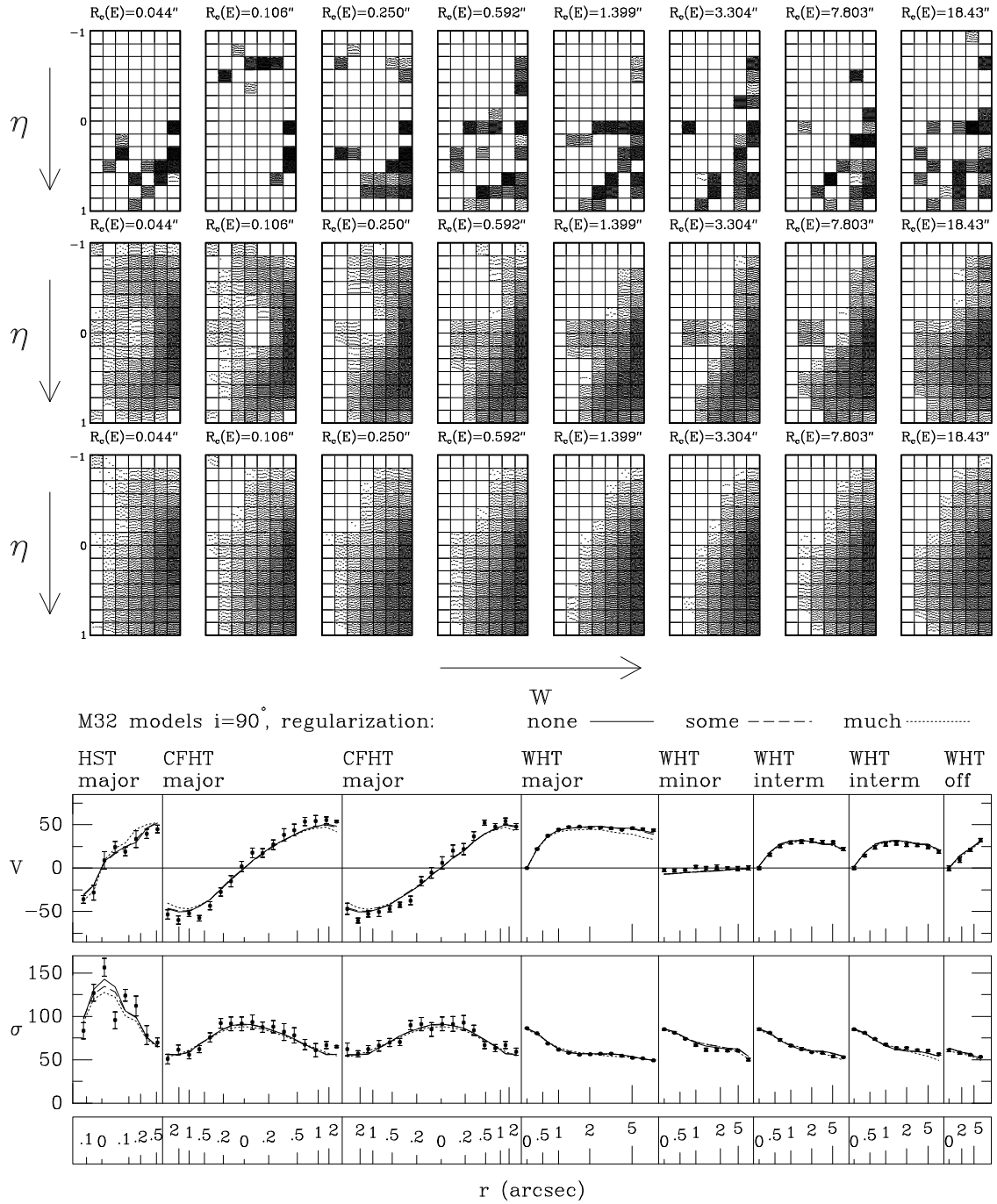


FIGURE 4.9— The top row shows the (η, w) integral space (defined as in Figure 4.5) for a selected set of energies, for the best-fitting edge-on orbit-superposition model C defined in Figure 4.6 and Table 4.1. Each square in each panel represents an orbit. The (logarithmic) grey-scale shows the fraction of the mass at the given energy that was assigned to each orbit by the NNLS fit. Smoother solutions are obtained by adding regularization constraints to the NNLS fit. The second and third rows show the integral space for the same model with a modest and a large amount of regularization, respectively. Most of the mass resides at $\eta > 0$ (i.e., $L_z > 0$), which is obviously required to fit the observed rotation of M32. The bottom panels show the fits of the models to the observed rotation velocities and velocity dispersions. The model without regularization provides the best fit, but even for the smoothest model the fits are still quite good.

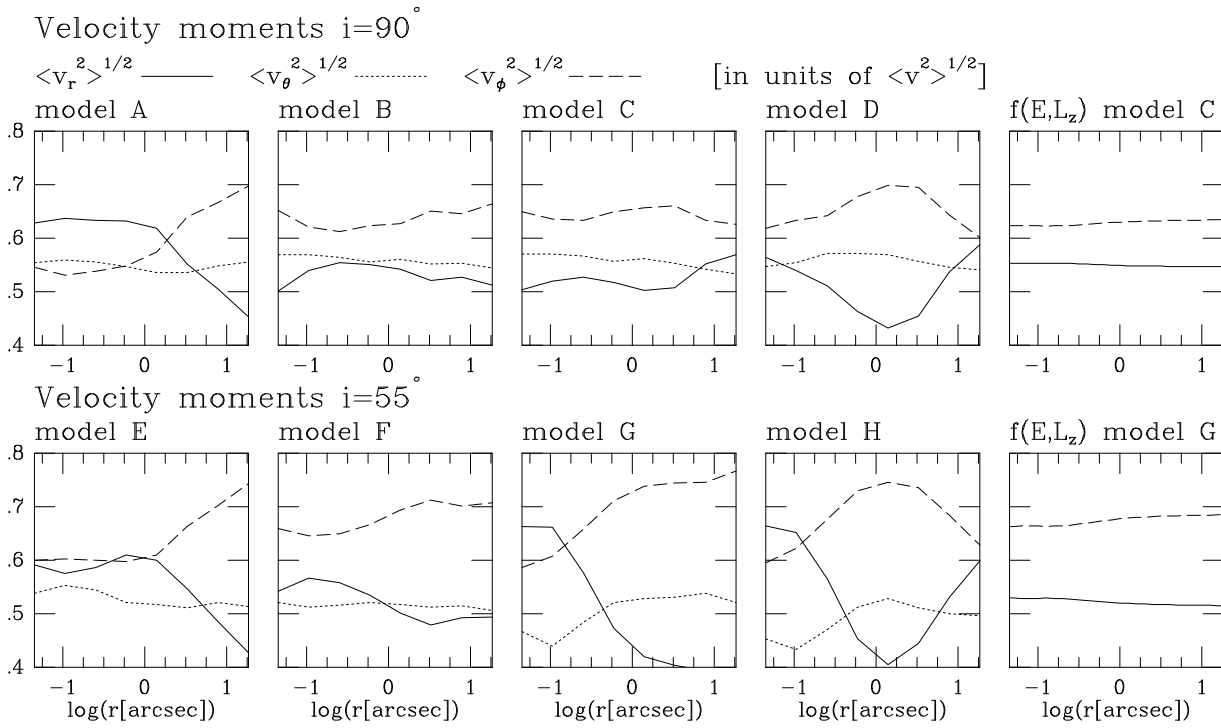


FIGURE 4.10— The four leftmost panels show the dynamical structure as function of radius for the orbit-superposition models A–D and E–H defined in Figure 4.6, averaged over spherical shells. The displayed results were obtained with a modest amount of regularization in the NNLS fit, to obtain smoother results. Only those radii are shown for which the dynamical structure of the models is meaningfully constrained by kinematical data. The curves show the RMS velocities in each of the spherical coordinate directions, normalized at each radius by the total RMS velocity. As a companion to this figure, Table 4.1 lists the average RMS velocities in km/s for two radial ranges, and also includes the azimuthal dispersion $\sigma_\phi \equiv [\langle v_\phi^2 \rangle - \langle v_\phi \rangle^2]^{1/2}$. Models A&E have no BH and invoke as much radial motion as possible (under the constraint that the rotation curve is fit) to produce a peak in the observed velocity dispersions. Nonetheless, they cannot fit the data (cf. Figure 4.8). The models with a BH are all dominated by azimuthal motion at most radii. The rightmost panels show the predictions (obtained by solving the Jeans equations) for models that have the same gravitational potential as the orbit-superposition models C&G, but which have a DF of the form $f(E, L_z)$. These models are also dominated by azimuthal motion, and therefore provide a useful approximation to the dynamics of M32. However, the $f(E, L_z)$ models have $\langle v_\theta^2 \rangle \equiv \langle v_r^2 \rangle$, which is not the case for the models that best fit the data.

By contrast to the models with a BH, the models A&E without a BH invoke a large amount of radial motion in the central arcsec to produce a peak in the observed velocity dispersions (cf. Binney & Mamon 1982). The maximum allowed radial anisotropy in the models is determined by the observed rotation velocities in M32, because dynamical models predict lower values of V/σ when they are more radially anisotropic (Richstone, Bower & Dressler 1990; de Bruijne, van der Marel & de Zeeuw 1996). Figure 4.8 shows that the allowed radial anisotropy is by far insufficient to fit the observed peak in the velocity dispersion profile without invoking a BH.

The models B–D and F–H, which represent the best fits for different potentials, all have a similar dynamical structure: the second velocity moment tensor is dominated by azimuthal motion. This is most pronounced for the intrinsically flatter $i = 55^\circ$ models. The rightmost panels in Figure 4.10 show the velocity moments for $f(E, L_z)$ models with the same gravitational potentials as models C&G. The $f(E, L_z)$ models are similar to the best-fitting three-integral models, in that they have an excess of azimuthal motion. This is why they have been so successful in fitting ground-based data, including available VP shape parameters, and it shows that they provide a

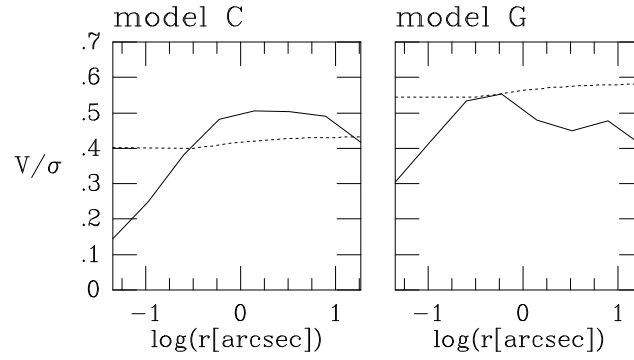


FIGURE 4.11— The quantity $\langle v_\phi \rangle / \langle v^2 \rangle^{1/2}$ as function of radius in the equatorial plane, for the models that best fit the data, i.e., the models C and G defined in Figure 4.6 and Table 4.1. Solid curves show the predictions for the orbit superposition models. As in Figure 4.10, a modest amount of regularization was used in the NNLS fit, and only those radii are shown for which the dynamical structure of the models is meaningfully constrained by kinematical data. Dotted curves show the predictions for oblate isotropic rotator models, obtained by solving the Jeans equations. The inclined, and intrinsically flatter, model G rotates faster than the edge-on model C.

useful low-order approximation to the dynamical structure of M32. However, the $f = f(E, L_z)$ models have $\langle v_r^2 \rangle = \langle v_\theta^2 \rangle$ by definition, in contrast to the inequality between $\langle v_r^2 \rangle$ and $\langle v_\theta^2 \rangle$ seen in the best-fitting three-integral models. This explains the finding of Section 4.3.2 that $f(E, L_z)$ models cannot successfully reproduce *all* observed features of the kinematical data.

The second velocity moments $\langle v_r^2 \rangle$ and $\langle v_\theta^2 \rangle$ can be combined with the mixed moment $\langle v_r v_\theta \rangle$ to determine the tilt of the velocity ellipsoid in the meridional plane. For all the models A–D and E–H we found $\langle v_r v_\theta \rangle$ to be small, and the velocity ellipsoids are more closely aligned with spherical coordinate axes than with cylindrical coordinate axes (but they are not perfectly aligned with either). This is not uncommon in three-integral models for axisymmetric systems (e.g., Dejonghe & de Zeeuw 1988; Dehnen & Gerhard 1993; de Zeeuw, Evans & Schwarzschild 1996).

Figure 4.11 shows the quantity $\langle v_\phi \rangle / \langle v^2 \rangle^{1/2}$ in the equatorial plane, for the best-fitting models C and G. The inclined, and intrinsically flatter, model G rotates faster than the edge-on model C. The predictions for oblate isotropic rotator models ($\sigma_r = \sigma_\theta = \sigma_\phi$) are shown for comparison. The edge-on model C rotates faster than an oblate isotropic rotator model for radii $r \gtrsim 0.5''$; the inclined model G rotates slower than the oblate isotropic rotator model. However, overall the rotation rate is not too dissimilar from that for an oblate isotropic rotator model.

The combined results that the dynamical structure of M32 is not too far from that of an $f(E, L_z)$ model, and that its rotation rate is not too far from that of an oblate isotropic rotator model, imply that the velocity dispersions (but not the RMS velocities) in M32 are not far from isotropic. This is confirmed by the velocity dispersion values listed in Table 4.1. This explains why the value of M_{BH} determined here is similar to that determined from previous studies: these assumed either isotropic velocity dispersions (e.g., Tonry 1987), a range of possible anisotropies that bracket isotropic models (e.g., Richstone, Bower & Dressler 1990), or an $f(E, L_z)$ DF (e.g., van der Marel 1994b). It is an interesting question what physical process could have caused M32 to have the particular dynamical structure inferred here, in particular because the DFs of (giant) elliptical galaxies are in general not well described by the $f(E, L_z)$ form (e.g., van der Marel 1991; Bender, Saglia & Gerhard 1994). The answer to this question is currently unknown.

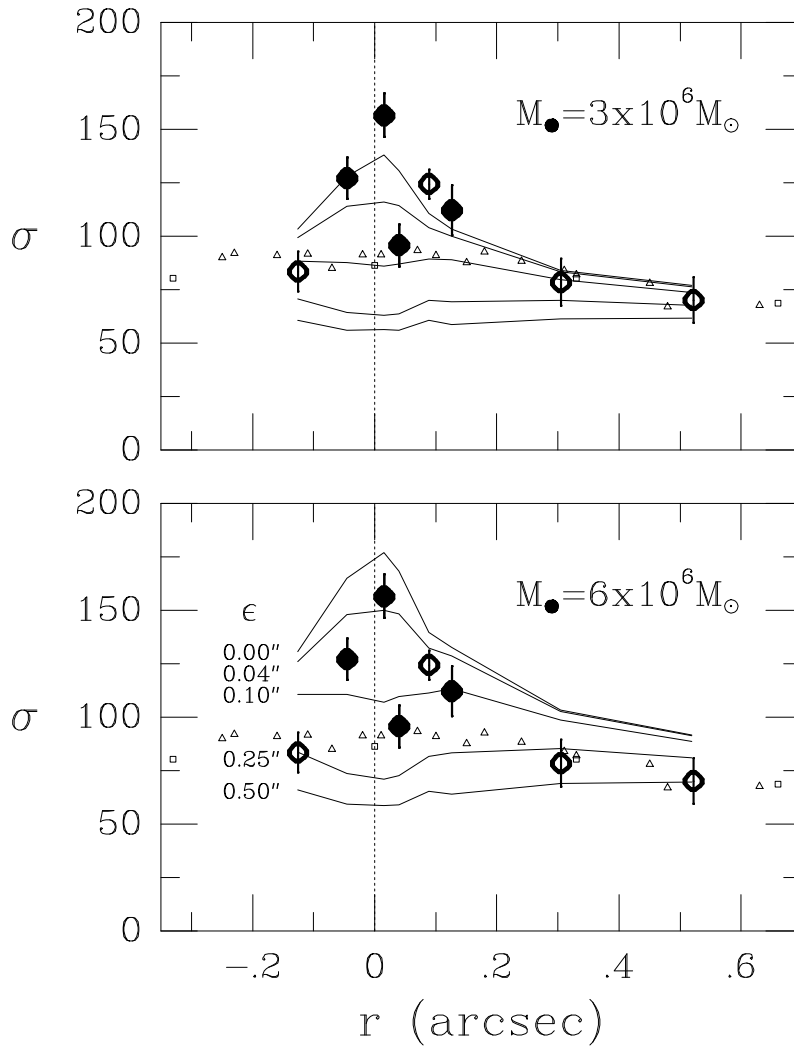


FIGURE 4.12— Solid curves show the velocity dispersions predicted for the HST setup by edge-on $f(E, L_z)$ models with an extended nuclear dark mass. Data points are as in Figure 4.2. The models in the top panel have $M_{\text{BH}} = 3 \times 10^6 M_{\odot}$, those in the bottom panel have $M_{\text{BH}} = 6 \times 10^6 M_{\odot}$. The models have nuclear dark objects with scale radii of $\epsilon = 0, 0.04'', 0.1'', 0.25''$ and $0.5''$, as indicated. The models with the smallest ϵ best fit the data.

We have not attempted to derive confidence bands on the dynamical structure of M32. This is a much more difficult problem than the derivation of confidence bands on the model parameters (M_{BH}, Y), and is beyond the scope of the present paper.

4.6 Models with an extended dark nuclear object

The results in the previous section demonstrate that M32 must have a massive dark object in its nucleus. To obtain a limit on the size of this dark object, we have studied models in which it has a finite size $\sim \epsilon$ (cf. eq. [4.2]). Searching the parameter space of three-integral models with different ϵ is extremely computer-intensive. We have therefore restricted ourselves to $f(E, L_z)$ models with extended dark objects. This is not likely to bias our conclusions, because the best-fitting three-integral models found in Section 4.5 are similar to two-integral models.

Figure 4.12 shows the predictions of edge-on $f(E, L_z)$ models with extended nuclear dark objects, for two representative values of M_{BH} . As in the models of Section 4.3.1, the fraction

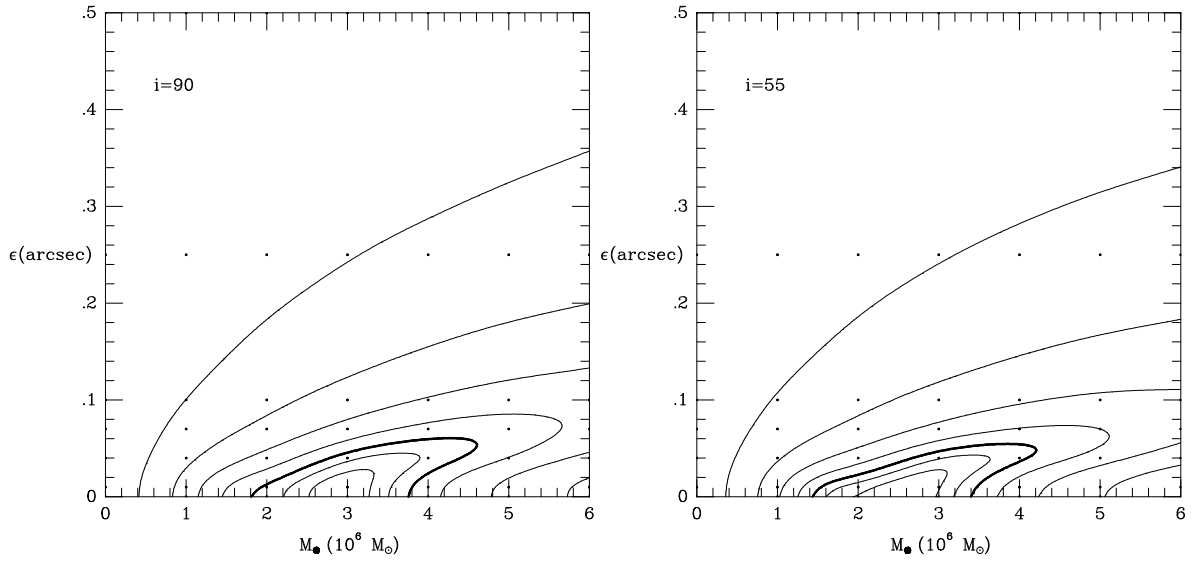


FIGURE 4.13— Contour plots of the quantity χ_σ^2 that measures the quality of the model fit to the HST velocity dispersion measurements, for $f(E, L_z)$ models with an extended dark nuclear object and an inclination of $i = 90^\circ$ (edge-on) or $i = 55^\circ$, respectively. The model parameters along the abscissa and ordinate are the nuclear dark mass M_{BH} and its scale radius ϵ , respectively. The contours are defined as in Figure 4.6. The results show that the nuclear dark mass in M32 must be less extended than $\epsilon = 0.06''$, independent of the inclination.

F of stars with $L_z > 0$ in each model was chosen to best fit the rotation curve. By adjusting F , adequate fits to the rotation velocities can be obtained for most relevant models. Hence, only the velocity dispersions are shown in the figure. The predicted dispersions are typically constant or decreasing towards the center within the scale radius ϵ . The HST data show a much higher velocity dispersion in the center than further out. Hence, the extension of any possible dark nuclear cluster cannot be large; Figure 4.12 suggests $\epsilon \lesssim 0.1''$. Figure 4.13 shows a contour plot of the quantity χ_σ^2 (eq. [4.3]), measuring the quality of the model fit to the HST dispersion measurements, both for the edge-on case and for $i = 55^\circ$. The best fit is obtained for $\epsilon = 0$ (the point mass case discussed previously). The formal 99.73 per cent confidence level (assuming Gaussian formal errors) rules out all models with $\epsilon \gtrsim 0.06''$, independent of the inclination. The models with the largest ϵ must have a total mass of at least $M_{\text{BH}} \approx 4 \times 10^6 M_\odot$.

At the distance of M32, $1'' = 3.39$ pc. Hence, the upper limit on the scale radius corresponds to $\epsilon = 0.20$ pc. Combined with a total mass in the cluster of $M_{\text{BH}} \approx 4 \times 10^6 M_\odot$, this implies a central mass density of at least $\rho_0 = 1.1 \times 10^8 M_\odot \text{pc}^{-3}$. The half-mass radius of a Plummer model is $r_h = 1.30\epsilon$. Hence, there must be $\sim 2 \times 10^6 M_\odot$ inside $r \lesssim 0.078''$. The total V-band luminosity inside this radius⁷ is $1 \times 10^5 L_\odot$, implying a luminous mass of $2.5 \times 10^5 M_\odot$. Hence, the ratio of the total mass to luminosity inside this radius must be $\gtrsim 22.5$.

The observed kinematics constrain only the amount of mass in the system, not whether this mass is luminous or dark. One can therefore fit the data equally well with models in which the average mass-to-light ratio Y of the stellar population increases towards the nucleus, and in which there is no dark mass. We have not explicitly constructed such models, but it is clear from the preceding discussion that in such models Y must rise from ~ 2 in the main body of the galaxy to $\gtrsim 20$ at $r \lesssim 0.1''$. Such a drastic variation in mass-to-light ratio would imply a strong

⁷This quantity does not depend sensitively on the assumed density cusp slope at very small radii. Gebhardt *et al.* (1996) infer a somewhat steeper slope for M32 than used here, but their model has only 10 per cent more luminous mass inside $r \lesssim 0.078''$ than ours.

change in the stellar population, accompanied by broad-band color gradients. The size of these gradients depends on the actual stellar population mix, which is unknown. However, one may use the properties of main sequence stars as a guideline. For these, a change in Y from 2 to 20 implies color changes $\Delta(U - B) \approx 0.9$, $\Delta(B - V) \approx 0.6$, and $\Delta(V - I) \approx 0.8$ (using the tables of stellar properties in Allen 1973). Such variations between $0.1''$ and $1''$ should have been obvious in photometric observations. However, neither subarcsec resolution ground-based imaging (Lugger *et al.* 1992) nor pre-refurbishment HST imaging (Crane *et al.* 1993) have revealed any significant color gradients in the central arcsec of M32. Post-refurbishment HST observations (Lauer *et al.*, private communication) also do not show strong color gradients. Thus, the nuclear mass concentration in M32 cannot be due merely to a change in the mix of ordinary stars in the nuclear region.

The absence of observed color gradients does not exclude the possibility of nuclear concentrations of brown dwarfs, white dwarfs, neutron stars, or stellar-mass BHs in M32. However, at high densities, such clusters of dark objects are not stable over a Hubble time. This was discussed by Goodman & Lee (1989), and their arguments were updated and extended in van der Marel *et al.* (1997a). The latter paper shows that the new HST limit on the density of dark material in M32 rules out all but the most implausible clusters, leaving a single massive BH as the most likely interpretation of the data.

The kinematical predictions of our models depend on the assumed Plummer form of the extended dark object. If it is a cluster of collapsed objects, this distributed dark mass may itself be cusped (e.g., Gerhard 1994). However, the limit on ϵ results from the fact that the dispersion of stars with mass density (4.1) in a Plummer potential does not have a (strong) central peak. This property is common to many alternative types of models, such as those of King (e.g., Binney & Tremaine 1987), Hernquist (1990) and Jaffe (1983). These all produce constant or decreasing dispersions inside their scale radius, as does the Plummer model. Hence, the upper bound on ϵ derived here is likely to be generic to most plausible density profiles for the extended dark object. In addition, King, Jaffe and Hernquist models are more centrally condensed than Plummer models, and would therefore require dark clusters of even higher central densities to fit the data.

4.7 Conclusions and discussion

4.7.1 Summary of results

The main bottlenecks in proving the presence of nuclear BHs in quiescent galaxies from stellar kinematical data have long been: (i) the restricted spatial resolution of ground-based data; and (ii) lack of sufficiently general dynamical models to rule out constant mass-to-light ratio models beyond doubt. The HST now provides spectra of superior spatial resolution. To fully exploit the potential of these new data it is imperative to improve the modeling techniques that have been used in the past decade. The situation is considerably more complicated than for gas disks in (active) galaxies, where the assumption of simple circular orbits is often adequate. Interpretation of stellar kinematical data for flattened elliptical galaxies ideally requires axisymmetric (or even better, triaxial) dynamical models with completely general three-integral distribution functions. Such models have not previously been constructed for any stellar kinematical BH candidate galaxy. We therefore developed a technique for the construction of such axisymmetric models, and used it to interpret our HST data for M32.

To guide the construction and interpretation of the three-integral models we first compared the new HST data to the predictions of $f(E, L_z)$ models, which have been used extensively to interpret ground-based M32 data. Such models have the advantage that the DF can be cal-

culated semi-analytically, but have the disadvantage of having a special dynamical structure, with $\sigma_r = \sigma_\theta$ everywhere. There is no a priori reason why any galaxy should have this property. However, the fact that $f(E, L_z)$ models fit the observed VP shapes inferred from ground-based data to within $\sim 2\%$ (in terms of deviations from a Gaussian), suggested that the M32 DF might in fact be close to the form $f(E, L_z)$. We find here that $f(E, L_z)$ models for M32 can also fit the new HST data, and that this requires the presence of a nuclear dark mass, as was the case for the ground-based data. However, the best fitting dark mass of $M_{\text{BH}} = (2.5\text{--}4.5) \times 10^6 M_\odot$ is larger than the $M_{\text{BH}} = (2\text{--}3) \times 10^6 M_\odot$ that best fits the $\sim 0.5''$ spatial resolution data from the CFHT, and is even more different from the $M_{\text{BH}} = (1.5\text{--}2) \times 10^6 M_\odot$ that best-fits the $\sim 0.9''$ spatial resolution data from the WHT. Thus, under the assumption of an $f(E, L_z)$ DF, the different data sets cannot be fit with the same M_{BH} . This indicates that the M32 DF is not of the form $f(E, L_z)$, although it might be close to it.

To obtain a model-independent estimate of the best-fitting M_{BH} , and to firmly rule out models without any dark mass, it is necessary to study more general three-integral models. We have made such models for M32, both with and without central BHs, and for various possible values of the average mass-to-light ratio Υ of the stellar population. The models were constructed to fit all available kinematical HST, CFHT and WHT data, and the acceptability of each model was assessed through the χ^2 of its fit to the data. The models demonstrate explicitly for the first time that there is no axisymmetric constant mass-to-light ratio model that can fit the kinematical data without invoking the presence of a nuclear dark mass, independent of the dynamical structure of M32. A nuclear dark point mass of $M_{\text{BH}} = (3.4 \pm 1.6) \times 10^6 M_\odot$ is required (with 1σ and 3σ error bars of $0.7 \times 10^6 M_\odot$ and $1.6 \times 10^6 M_\odot$, respectively, which includes the possible effect of small numerical errors in the models). This mass is similar to that quoted by most previous papers, but the confidence on the detection of a nuclear dark mass in M32 is now much higher. Constant mass-to-light ratio models still come very close to fitting the ground-based data, and only the new HST data make the case for a nuclear dark mass clear-cut.

The inclination of M32 cannot be inferred from the available surface photometry, and is therefore a free parameter in the modeling. Ideally one would like to construct dynamical models for all possible inclinations (which would be very computer-intensive) and determine the inclination that best fits the kinematical data. Here we have taken the more modest approach of constructing models for only two representative inclinations: $i = 90^\circ$ (edge-on) and $i = 55^\circ$. The intrinsic axial ratios for these inclinations are $q = 0.73$ and $q = 0.55$, respectively. The three-integral $i = 55^\circ$ models provide a better fit than the edge-on models, which suggests that M32 is not seen edge-on. However, the allowed range for M_{BH} does not depend sensitively on the assumed inclination: models with no central dark mass are firmly ruled out for both inclinations. So even though a more detailed study of the full inclination range for M32 would improve our knowledge of the true inclination and intrinsic axial ratio of M32, it would probably not change significantly the constraints on the central dark mass.

The best-fitting three-integral models are similar to $f(E, L_z)$ models in that they have an excess of azimuthal motion. This is why they have been so successful in fitting ground-based data, including available VP shape parameters, and it confirms that they provide a useful low-order approximation to the dynamical structure of M32. However, $f(E, L_z)$ models do have $\sigma_r \equiv \sigma_\theta$. This does not reproduce the inequality between σ_r and σ_θ , nor the modest tilt of the velocity ellipsoid indicated by the small $\langle v_r v_\theta \rangle$ term, seen in the best-fitting three-integral models. This is why $f(E, L_z)$ models cannot successfully explain *all* observed features of the kinematical data.

To constrain the size of the dark object in M32 we have constructed $f(E, L_z)$ models with an extended dark nuclear object. These show that the HST data put an upper limit of $0.08'' =$

0.26 pc on the half-mass radius of the nuclear dark object, implying a central density exceeding $1 \times 10^8 \text{ M}_\odot \text{ pc}^{-3}$. This limit on the density of dark material in M32 essentially rules out nuclear clusters of planets, brown dwarfs, white dwarfs, neutron stars, or smaller mass BHs (van der Marel 1997a). The absence of color gradients in the central arcsec of M32 implies that the nuclear mass concentration can also not be attributed to a stellar population gradient. A single massive nuclear BH therefore provides the most plausible interpretation of the data.

4.7.2 Dynamical stability

Axisymmetric dynamical models with a nuclear BH provide an excellent fit to all available kinematical data for M32. However, to be physically meaningful, the models must also be dynamically stable. In van der Marel, Sigurdsson & Hernquist (1997) we presented N-body simulations of the $f(E, L_z)$ models for M32. The models were found to be completely stable, both for $i = 90^\circ$ and for $i = 55^\circ$. This shows that dynamical stability is not a problem for the models, and that the inclination of M32 cannot be meaningfully constrained through stability arguments. We have not evolved N-body models for the best-fitting three-integral models, but we expect these models to be stable as well, given their similarity to $f(E, L_z)$ models.

4.7.3 Dynamical relaxation

The two-body relaxation time in M32 can be estimated as in, e.g., Binney & Tremaine (1987; eq. [8-71]). Using the relevant quantities for our best-fitting dynamical model, we find for solar mass stars in the central cusp ($r \lesssim 0.5''$) that $t_{\text{relax}} \approx 3 \times 10^9 (r/0.1'')^{-0.065} \text{ yr}$. The time scale for ‘resonant relaxation’ (Rauch & Tremaine 1996) is of the same order. The central cusp must therefore be evolving secularly over a Hubble time. However, the diffusion of stars in phase space is slow enough that one may assume the evolution to be through a sequence of quasi-equilibrium models. This justifies our approach of modeling M32 as a collisionless equilibrium system. Studies of the secular evolution of the M32 cusp will be interesting, but will not change the need for a nuclear dark object. In fact, the process of dynamical relaxation supports the presence of a dark object: without a dark object the relaxation would proceed at a much more rapid rate that is difficult to reconcile with observations (Lauer *et al.* 1992).

4.7.4 Triaxiality

One remaining uncertainty in our dynamical modeling is the possibility of triaxiality. After the step from spherical models to axisymmetric models, triaxial models are the obvious next step. However, there are several reasons to believe that for M32 this additional step will be less important. First, M32 is known not to be spherical, but there is no reason why it cannot be axisymmetric. There is no significant isophote twisting in M32, and no minor axis rotation. This does not mean that M32 cannot be triaxial (we might be observing it from one of the principal planes), but it also does not mean that M32 needs to be triaxial. Second, spherical constant mass-to-light ratio models (without a nuclear dark mass) for ground-based M32 data failed to fit by only a few kms^{-1} , and it was quite conceivable that axisymmetry could fix this (which it did, cf. Figure 4.15 below). However, axisymmetric constant mass-to-light ratio models for the new HST data fail to fit the nuclear velocity dispersion by $> 50 \text{ km/s}$, and this cannot likely be fixed through triaxiality. Third, theoretical arguments suggest that strongly triaxial models with density cusps as steep as in M32 may not be stable, owing to the fact that regular box-orbits are replaced by boxlets and irregular orbits that may not be able to sustain a triaxial shape (Binney & Gerhard 1985; Merritt & Fridman 1996; Merritt & Valluri 1996; see also the review by de Zeeuw 1996). Rapidly-rotating low-luminosity elliptical galaxies like M32 always have steep

power-law cusps (Faber *et al.* 1997), and may therefore be axisymmetric as a class (de Zeeuw & Carollo 1996). This is consistent with statistical studies of their intrinsic shapes (e.g., Merritt & Tremblay 1996). So, apart from the fact that triaxiality is unlikely to remove the need for a central dark object in M32, it may even be so that M32 cannot be significantly triaxial.

4.7.5 Adiabatic black hole growth

The growth of a black hole into a stellar system is adiabatic if it occurs over a time scale that is ‘long’ (see Sigurdsson, Quinlan, & Hernquist 1995 for a quantitative discussion) compared to the typical orbital period of the stars. For the case of M32, the black hole formation can be considered adiabatic if it took at least 10^6 yr. Young (1980) studied the adiabatic growth of BHs in spherical isothermal models with central density ρ_0 and core radius r_0 . The BH growth leaves the mass density at large radii unchanged, but induces a central cusp $\rho \propto r^{-1.5}$ for $r \rightarrow 0$. The form of the density profile at intermediate radii is determined by the dimensionless parameter $\overline{M}_\bullet \equiv M_{\text{BH}} / [\frac{4}{3}\pi\rho_0 r_0^3]$, which measures the ratio of the BH mass to the initial core mass. Lauer *et al.* (1992) showed that the shape of the M32 brightness profile measured with HST can be well fit with $\overline{M}_\bullet = 0.33 \pm 0.11$. The radial and density normalization implied by the data are then $r_0 = 3.0$ pc and $\rho_0 = (4.2 \times 10^4) Y M_\odot \text{ pc}^{-3}$. This *photometric* model therefore implies that $M_{\text{BH}}/Y = (1.6 \pm 0.5) \times 10^6 M_\odot$. Although this result depends somewhat on the assumed isothermality of the initial distribution (Quinlan, Hernquist & Sigurdsson 1995), it is quite remarkable that our best-fitting *dynamical* models have exactly $M_{\text{BH}}/Y = 1.6 \times 10^6 M_\odot$, for both inclinations that we studied. The M32 data are therefore fully consistent with the presence of a BH that grew adiabatically into a pre-existing core. This is similar to the situation for M87 (cf. Young *et al.* 1978; Harms *et al.* 1994).

Lee & Goodman (1989) extended Young’s calculations to the case of rotating models. For the value of \overline{M}_\bullet implied by the photometry, their models predict a profile of $\langle v_\phi \rangle / \langle v^2 \rangle^{1/2}$ that is approximately flat with radius (with amplitude fixed by the axial ratio of the system). However, this result depends very sensitively on the assumed rotation law of the initial model. The radial variations in $\langle v_\phi \rangle / \langle v^2 \rangle^{1/2}$ seen in our best-fitting models (Figure 4.11) are probably equally consistent with the adiabatic growth hypothesis.

4.7.6 Tidal disruption of stars

A star of mass m_\star and radius r_\star on a circular orbit of radius r will be tidally disrupted if $r \lesssim r_t \equiv (2M_{\text{BH}}/m_\star)^{1/3} r_\star$ (e.g., Binney & Petit 1988). Thus, disruption of a solar type star by the BH in M32 will occur inside $r_t = 4.2 \times 10^{-6}$ pc = 1.2×10^{-6} arcsec. A disruption event will be highly luminous, but is not predicted to occur more often than once every 10^4 yr (Rees 1988). The minimum pericenter distance for a star with given (E, L_z) in a Kepler potential is $r_{\text{p,min}} = R_c(E)(1 - \sqrt{1 - \eta^2})$, where as before, R_c is the radius of the circular orbit at the given energy and $\eta \equiv L_z/L_{\text{max}}(E)$. The kinematical data for M32 only meaningfully constrain the DF for energies with $R_c(E) \gtrsim 0.1''$. For $R_c(E) = 0.1''$, only stars with $|\eta| < 5 \times 10^{-3}$ have $r_{\text{p,min}} < r_t$. The data do not constrain variations in the DF over such a small range in η , and our dynamical models therefore cannot address the existence and properties of the so-called ‘loss cone’ (Frank & Rees 1976; Lightman & Shapiro 1977). For the η -grid that we have employed, all solar type stars on orbits with $R_c(E) > 0.1''$ have $r_{\text{p,min}} > 2 \times 10^2 r_t$. Even giants with $r_\star \approx 10^2 r_\odot$ have $r_{\text{p,min}} > r_t$. This justifies our neglect of tidal disruption in the orbit calculations.

4.7.7 Accretion onto the black hole

An interesting question is why BHs in quiescent galaxies aren't more luminous (e.g., Kormendy & Richstone 1995). For M32, the total X-ray luminosity is $L_X \approx 10^{38} \text{ erg s}^{-1}$ (Eskridge, White & Davis 1996), the far infrared luminosity is $L_{\text{FIR}} < 3 \times 10^{36} \text{ erg s}^{-1}$ (Knapp *et al.* 1989), and for the 6 cm radio emission $\nu L_\nu < 3 \times 10^{33} \text{ erg s}^{-1}$ (Roberts *et al.* 1991). Part or all of the observed X-ray emission may be due to low-mass X-ray binaries, so the total luminosity due to accretion onto the BH in M32 is $L_{\text{acc}} < 10^{38} \text{ erg s}^{-1}$. By contrast, the Eddington luminosity of the BH is $L_{\text{Edd}} = 4.3 \times 10^{44} \text{ erg s}^{-1}$. For a canonical mass-loss rate of $1.5 M_\odot (10^{11} L_\odot)^{-1} \text{ yr}^{-1}$ (Faber & Gallagher 1976), the stars that are bound to the BH in M32 shed $1 \times 10^{-4} M_\odot \text{ yr}^{-1}$ of gas as a result of normal stellar evolution. If a fraction f of this gas is steadily accreted with efficiency ϵ , it produces a luminosity $L_{\text{acc}} = \epsilon f (6.7 \times 10^{42}) \text{ erg s}^{-1}$. Thus either the accretion fraction f or the accretion efficiency ϵ must be very small in M32. Thin disk accretion with $\epsilon \approx 0.1$ requires $f < 1.5 \times 10^{-4}$, which is possible (the accretion fraction is difficult to predict theoretically, because it depends on the hydrodynamics of the stellar winds that shed the gas), but may be implausibly low. Instead, it appears more likely that ϵ is small, since there is a family of 'advection dominated' accretion solutions that naturally predict such low efficiencies. Models of this type successfully explain the 'micro-activity' of the BH (Sgr A*) in our own Galaxy (Narayan, Yi & Mahadevan 1995). In a typical accretion model of this type (Narayan & Yi 1995, their Fig. 11), $f \lesssim 0.16$ suffices to explain the upper bound on L_{acc} for M32.

4.7.8 Forthcoming observations

Future observations of M32 will include spectra with the new long-slit HST spectrograph STIS. These will provide significantly better sky coverage than our FOS data, but the spatial resolution will be similar. The high-resolution HST data can be complemented with that from fully two-dimensional ground-based spectrographs, such as OASIS on the CFHT and SAURON on the WHT. These combined data will yield improved constraints on the BH mass, on the orbital structure and inclination of M32, and on possible deviations from axisymmetry.

Acknowledgments

We dedicate this paper, and its companions Rix *et al.* (1997) and Cretton *et al.* (1999), to the memory of Martin Schwarzschild, who pioneered the modeling technique employed here. Martin's sense of purpose, his exceptional clarity of thinking, his transparent personal integrity and, most of all, his genuine warm interest and support, remain a great source of inspiration to us. Support for this work was provided by NASA through grant number #GO-05847.01-94A, and through a Hubble Fellowship #HF-1065.01-94A awarded to RPvdM, both from the Space Telescope Science Institute which is operated by the Association of Universities for Research in Astronomy, Incorporated, under NASA contract NAS5-26555. NC acknowledges financial support from the Swiss government (Etat du Valais) and NUFFIC, and the hospitality of Steward Observatory, the MPA Garching and Geneva Observatory. TdZ is grateful for the generous hospitality of the Institute for Advanced Study. Both he and NC received financial support from the Leids Kerkhoven Bosscha Fonds.

4.A χ^2 topology for orbit-superposition models

In this Appendix we discuss the topology of the $\chi^2(\Upsilon, M_{\text{BH}})$ contours for the edge-on orbit-superposition models. The top panels of Figure 4.14 show the χ^2 contours when only (subsets of) the ground-based WHT data are included in the fit. These panels can be compared to Figure 4.14d, which shows the contours for the case in which all WHT, CFHT and HST data are

included. Figure 4.14a shows the χ^2 contours when only the major axis V and σ WHT measurements are fit. Binney & Mamon (1982) showed that a large range of gravitational potentials can fit any given observed velocity dispersion profile. The valley seen in the χ^2 contours is a consequence of this: it outlines a one-parameter family of models that can fit the data with different velocity dispersion anisotropy. For a non-rotating spherical system, only models that require negative second velocity moments are ruled out. For a rotating system like M32, the observed rotation rate sets additional limits on the allowed radial anisotropy. For the case of the major axis V and σ WHT measurements, a no-BH model is just marginally acceptable at 99.73 per cent confidence, cf. Figure 4.14a. For the lower-spatial resolution major axis V and σ measurements of Dressler & Richstone (1988) such a model is entirely acceptable. Figure 4.15 compares the predictions of the best-fit axisymmetric orbit-superposition model without a BH to their data. Richstone, Bower & Dressler (1990) concluded that these data could not be fit by any *spherical* model without a BH. This is because spherical models allow less rotation, and therefore failed to fit the observed rotation velocities. This underscores the importance of making axisymmetric models for flattened galaxies like M32.

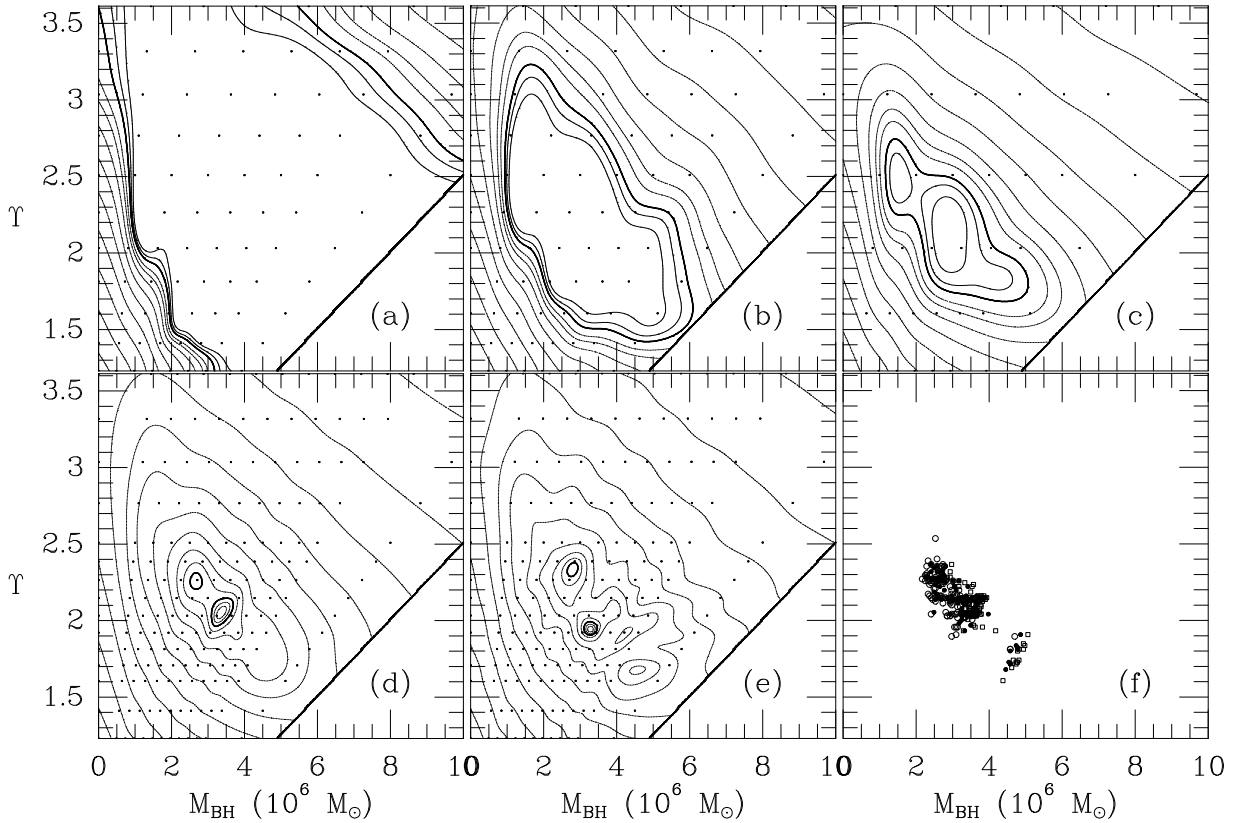


FIGURE 4.14— Contour plots of $\chi^2(M_{\text{BH}}, \Upsilon)$ for edge-on orbit superposition models constructed to fit: (a) the major axis WHT V and σ measurements; (b) all major axis WHT measurements, including VP shapes; (c) all WHT measurements, including various position angles; (d) all WHT, CFHT and HST data (same as Figure 4.6). Contours are defined as in Figure 4.6. Heavy contours show the formal 99.73% confidence regions. Panel (e) shows how the contours in panel (d) are modified if random errors are added to the predictions for each data point, to simulate numerical errors in the models. Panel (f) shows for 100 simulations as in panel (e) the position of the χ^2 minimum (solid symbols), and the lowest and highest M_{BH} that fall within a 99.73% confidence contour (open symbols). These simulations show that numerical errors cannot be responsible for the fact that models with either $M_{\text{BH}} < 1.8 \times 10^6 M_{\odot}$ or $M_{\text{BH}} > 5.0 \times 10^6 M_{\odot}$ fail to fit the data at this confidence level.

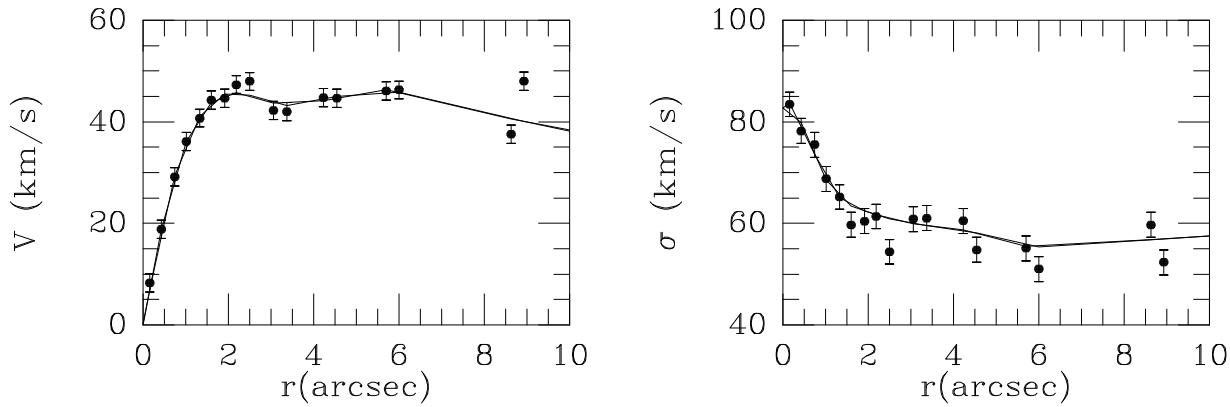


FIGURE 4.15— Rotation velocities and velocity dispersions for the data of Dressler & Richstone (1988). The data were measured from their figure 1; their two separate exposures were averaged. The seeing was $1.04''$ FWHM, the pixel size $0.585''$ and the slit width $1.0''$. The curves show the predictions of the edge-on axisymmetric orbit superposition model without a BH that provides the best fit. This model has $\Upsilon = 3.6$. As in Section 4.5.2, the orbit superposition was done without regularization constraints. The model has $\langle v_\phi^2 \rangle^{1/2} : \sigma_\phi : \sigma_\theta : \sigma_r = 1 : 0.82 : 1.02 : 1.16$ averaged over the radial range $0.07'' \leq r \leq 0.9''$, and $1 : 0.91 : 0.86 : 0.90$ averaged over the radial range $0.9'' \leq r \leq 12''$. Thus, the velocity dispersions are (mildly) radially anisotropic in the central arcsec, and close to isotropic outside the central arcsec. The model provides an adequate fit. These data could not be fit by any spherical model without a BH (Richstone, Bower & Dressler 1990), which illustrates the importance of making axisymmetric models for flattened galaxies like M32.

VP shape measurements provide independent constraints on the velocity dispersion anisotropy. Figure 4.14b shows the χ^2 contours for edge-on orbit-superposition models when not only the WHT major axis V and σ measurements are fit, but also the major axis VP shape measurements. With the inclusion of the VP shapes, models without a BH are ruled out. Figure 4.14c shows the χ^2 contours when also the WHT measurements along other position angles are included, which contracts the allowed M_{BH} range to $(1.1\text{--}5.1) \times 10^6 M_\odot$ at the formal 99.73 per cent confidence level. The WHT data by themselves therefore rule out axisymmetric models without a BH. However, the models without a BH still come very close to fitting the data, and, e.g., fail to fit the central velocity dispersion by only $1\text{--}2 \text{ km s}^{-1}$ (cf. Figure 4.8). So one cannot make a particularly strong claim for a BH on the basis of the WHT data alone, because it is conceivable that the fit could be improved with, e.g., only a minor amount of triaxiality. The same holds for the CFHT data, but the new HST data do make the case for a BH in M32 clear-cut.

The contours for the case in which all the available WHT, CFHT and HST kinematical data are included in the fit (Figure 4.14d) show one global χ^2 minimum, and a second local minimum. The presence of a global minimum does not necessarily imply that the combined data constrain a single best-fit potential. It might be that there is a small range of potentials that all fit equally well, but that such a range of constant χ^2 would not be evident due to the finite numerical accuracy of our technique. In Figure 4.14e we show explicitly how the topology of the χ^2 contours might have been influenced by the possibility of small numerical errors in our models. It was obtained from Figure 4.14d by recalculating the χ^2 contours after adding a random error $\Delta V, \sigma \in [-2, 2] \text{ km/s}$ and $\Delta h_i \in [-0.01, 0.01]$ (cf. Section 4.4) to the prediction for each data point, for each $(\Upsilon, M_{\text{BH}})$ combination. The results show that numerical errors can indeed influence the χ^2 contours near the χ^2 minimum. Thus, the second minimum in Figure 4.14d might be the result of numerical inaccuracies in our technique. However, the numerical errors are small enough that they only have a negligible effect on the overall χ^2 topology. In particular, models without a dark mass remain firmly ruled out.

To assess the possible effect of numerical errors on the confidence bands for M_{BH} , we constructed 100 figures like Figure 4.14e using different random realizations. For each we determined the position of the χ^2 minimum, and the minimum and maximum M_{BH} for which there is an Υ such that the model with $(M_{\text{BH}}, \Upsilon)$ falls within the 99.73 per cent confidence region. The results are plotted in Figure 4.14f. All allowed M_{BH} values fall in the range $M_{\text{BH}} = (1.8\text{--}5.0) \times 10^6 M_{\odot}$. Thus, $M_{\text{BH}} = (3.4 \pm 1.6) \times 10^6 M_{\odot}$ at 99.73 per cent confidence. Similar experiments show that $M_{\text{BH}} = (3.4 \pm 0.7) \times 10^6 M_{\odot}$ at 68.3 per cent confidence. Experiments for $i = 55^\circ$ produced similar results, and mass ranges that were either the same or slightly smaller. Thus, we conclude that the 1σ and 3σ errors on the estimated $M_{\text{BH}} = 3.4 \times 10^6 M_{\odot}$, are 0.7 and $1.6 \times 10^6 M_{\odot}$, respectively.

References

- Allen, C. W. 1973, *Astrophysical Quantities* (London: The Athlone Press)
- Bender, R., Kormendy, J., & Dehnen, W. 1996, *ApJ*, 464, L123
- Bender, R., Saglia, R.P., & Gerhard, O.E. 1994, *MNRAS*, 269, 785
- Binney, J. J., & Gerhard O. E. 1985, *MNRAS*, 215, 469
- Binney, J., & Mamon, G. A. 1982, *MNRAS*, 200, 361
- Binney, J., & Petit, J.-M. 1988, in *Dynamics of Dense Stellar Systems*, ed. Merritt, D. R. (Cambridge: Cambridge University Press), 43
- Binney, J., & Tremaine, S. 1987, *Galactic Dynamics* (Princeton: Princeton University Press)
- Crane, P., et al. 1993, *ApJ*, 106, 1371
- Cretton, N., de Zeeuw, P. T., van der Marel, R. P., & Rix H-W. 1999, *ApJ Supplements*, in press, chapter 3 of this thesis
- de Bruijne, J. H. J., van der Marel, R. P., & de Zeeuw P. T. 1996, *MNRAS*, 282, 909
- Dehnen, W. 1995, *MNRAS*, 274, 919
- Dehnen, W., & Gerhard, O. E. 1993, *MNRAS*, 261, 311
- Dejonghe, H., & de Zeeuw, P. T. 1988, *ApJ*, 333, 90
- de Zeeuw, P. T. 1996, in *Gravitational Dynamics*, eds. Lahav, O., Terlevich, E., & Terlevich, R. J. (Cambridge: Cambridge University Press), 1
- de Zeeuw, P. T., & Carollo, C. M. 1996, in *New Light on Galaxy Evolution*, IAU Symposium 171, eds. Bender, R., & Davies, R. L. (Dordrecht: Kluwer Academic Publishers), 47
- de Zeeuw, P. T., Evans N. W., & Schwarzschild, M. 1996, *MNRAS*, 280, 903
- de Zeeuw, P. T. 1997, in *The Nature of Elliptical Galaxies*, *Proceedings of the Second Stromlo Symposium*, eds. Arnaboldi, M., da Costa G., & Saha, P., (San Francisco: ASP), 44
- Dressler, A. 1984, *ApJ*, 286, 97
- Dressler, A., & Richstone, D. O. 1988, *ApJ*, 324, 701
- Eskridge, P. B., White, R. E., & Davis, D. S. 1996, *ApJ*, 463, L59
- Faber, S. M., & Gallagher, J. S. 1976, *ApJ*, 204, 365
- Faber, S. M., et al. 1997, *AJ*, 114, 1771
- Frank, J., & Rees, M. J. 1976, *MNRAS*, 176, 633
- Gebhardt, K., et al. 1996, *AJ*, 112, 105
- Gerhard, O. E. 1994, in *The nuclei of normal galaxies: lessons from the Galactic Center*, eds. Genzel, R., & Harris, A. I., (Dordrecht: Kluwer Academic Publishers), 267
- Goodman, J., & Lee, H. M. 1989, *ApJ*, 337, 84
- Harms, R. J. et al. 1994, *ApJ*, 435, L35
- Hernquist, L. 1990, *ApJ*, 356, 359
- Jaffe, W. 1983, *MNRAS*, 202, 995
- Kent, S. M. 1987, *AJ*, 94, 306

- Knapp, G. R., Guhathakurta, P., Kim, D.-W., & Jura, M. 1989, *ApJS*, 70, 329
- Kormendy, J., & Richstone D. 1995, *ARA&A*, 33, 581
- Lauer, T. R., et al. 1992, *AJ*, 104, 552
- Lawson, C. L., & Hanson, R. J. 1974, *Solving Least Squares Problems* (Englewood Cliffs, New Jersey: Prentice-Hall)
- Lee, M. H., & Goodman, J. 1989, *ApJ*, 343, 594
- Lightman, A. P., & Shapiro, S. L. 1977, *ApJ*, 211, 244
- Lugger, P. M., Cohn, H. N., Cederbloom, S. E., Lauer, T. R., & McClure, R. D. 1992, *AJ*, 104, 83
- Lynden-Bell, D. 1996, *MNRAS*, 279, 389
- Merritt, D. R., & Fridman, T. 1996, *ApJ*, 460, 136
- Merritt, D. R., & Tremblay, B. 1994, *AJ*, 108, 514
- Merritt, D. R., & Tremblay, B. 1996, *AJ*, 111, 2243
- Merritt, D. R., & Valluri, M. 1996, *ApJ*, 471, 82
- Narayan, R., & Yi, I. 1995, *ApJ*, 452, 710
- Narayan, R., Yi, I., & Mahadevan, R. 1995, *Nature*, 374, 623
- Ollongren, A. 1962, *Bull. Astr. Inst. Netherlands*, 16, 241
- Peletier, R. F. 1993, *A&A*, 271, 51
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, *Numerical Recipes* (Cambridge: Cambridge University Press)
- Qian, E. E., de Zeeuw, P. T., van der Marel, R. P., & Hunter, C. 1995, *MNRAS*, 274, 602
- Quinlan, G. D., Hernquist, L., & Sigurdsson, S. 1995, *ApJ*, 440, 554
- Rauch, K. P., & Tremaine, S. 1996, *New Astron.*, 1, 149
- Rees, M. 1988, *Nature*, 333, 523
- Rees, M. 1996, in *Gravitational Dynamics*, eds. Lahav, O., Terlevich, E., & Terlevich, R. J. (Cambridge: Cambridge University Press)
- Richstone, D. O. 1982, *ApJ*, 252, 496
- Richstone, D. O., & Tremaine, S. 1988, *ApJ*, 327, 82
- Richstone, D. O., & Bower, G., & Dressler, A. 1990, *ApJ*, 353, 118
- Rix, H.-W., de Zeeuw, P. T., Cretton, N., & van der Marel, R. P., Carollo, C. M. C., 1997, *ApJ*, 488, 702, chapter 2 of this thesis
- Roberts, M. S., Hogg, D. E., Bregman, J. N., Forman, W. R., & Jones, C. 1991, *ApJS*, 75, 751
- Schwarzschild, M. 1979, *ApJ*, 232, 236
- Sigurdsson, S., Hernquist, L., & Quinlan, G. D. 1995, *ApJ*, 446, 75
- Tonry, J. L. 1984, *ApJ*, 283, L27
- Tonry, J. L. 1987, *ApJ*, 322, 632
- van den Bosch, F. C. 1997, *MNRAS*, 287, 543
- van der Marel, R. P. 1991, *MNRAS*, 253, 710
- van der Marel, R. P., Rix, H.-W., Carter, D., Franx, M., White, S. D. M., & de Zeeuw, P. T. 1994a, *MNRAS*, 268, 521
- van der Marel, R. P., Evans, N. W., Rix, H.-W., White, S. D. M., & de Zeeuw, P. T. 1994b, *MNRAS* 271, 99
- van der Marel, R. P., de Zeeuw, P. T., Rix H.-W., & Quinlan, G. D. 1997a, *Nature*, 385, 610
- van der Marel, R. P., de Zeeuw, P. T., & Rix, H.-W. 1997b, *ApJ*, 488, 119 (Paper I)
- van der Marel, R. P., Sigurdsson, S., & Hernquist, L. 1997, *ApJ*, 487, 153
- Young, P. 1980, *ApJ*, 242, 1232
- Young, P., Westphal, J. A., Kristian, J., Wilson, C. P., & Landauer, F. P. 1978, *ApJ*, 221, 721
- Zhao, H. S. 1996, *MNRAS*, 283, 149

Chapter 5

A super-massive black hole in the S0 galaxy NGC 4342

Cretton, N. & van den Bosch, F. C.
1999, *ApJ*, 514, 704

We present axisymmetric dynamical models of the edge-on S0 galaxy NGC 4342. This small low-luminosity galaxy harbors, in addition to its outer disk, a bright nuclear stellar disk. A combination of observations from the ground and with the Hubble Space Telescope (HST) has shown that NGC 4342 rotates rapidly and has a strong central increase in velocity dispersion.

We construct simple two-integral Jeans models as well as fully general, three-integral models. The latter are built using a modified version of Schwarzschild's orbit-superposition technique developed by Rix *et al.* and Cretton *et al.* (chapters 2 and 3 of this thesis). These models allow us to reproduce the full line-of-sight velocity distributions, or 'velocity profiles' (VPs), which we parameterize by a Gauss-Hermite series. The modeling takes seeing convolution and pixel binning into account.

The two-integral Jeans models suggest a black hole (BH) mass between 3 and $6 \times 10^8 M_\odot$, depending on the data set used to constrain the model, but they fail to fit the details of the observed kinematics. The three-integral models can fit all ground-based and HST data simultaneously, but only when a central BH is included. Models without BH are ruled out to a confidence level better than 99.73 per cent. We determine a BH mass of $3.0^{+1.7}_{-1.0} \times 10^8 M_\odot$, where the errors are the formal 68.3 per cent confidence levels. This corresponds to 2.6 per cent of the total mass of the bulge, making NGC 4342 one of the galaxies with the highest BH mass to bulge mass ratio currently known.

The models that best fit the data do not have a two-integral phase-space distribution function. They have rather complex dynamical structures: the velocity anisotropies are strong functions of radius reflecting the multi-component structure of this galaxy.

When no central BH is included the best fit model tries to fit the high central velocity dispersion by placing stars on radial orbits. The high rotation velocities measured, however, restrict the amount of radial anisotropy such that the central velocity dispersion measured with the HST can only be fit when a massive BH is included in the models.

SEVERAL lines of evidence suggest that active galactic nuclei (AGNs) are powered by accretion onto a super-massive black hole (BH) (Lynden-Bell 1969; Rees 1984). The much higher volume number density of AGNs observed at redshift $z \approx 2$ than at $z = 0$, suggests that many quiescent (or 'normal') galaxies today must have gone through an active phase in the past, and therefore harbor a massive BH as well. Such a BH will significantly influence the dynamics of the galaxy inside a radius of influence, $r_{\text{inf}} = GM_{\text{BH}}/\sigma^2$, where σ is a characteristic velocity dispersion of the stars in the center. In particular, hydrostatic equilibrium requires that the rms velocities of the stars surrounding a massive BH follow an $r^{-1/2}$ power-law (Bahcall & Wolf 1976; Young 1980).

Since the late 70s, combined imaging and spectroscopy of the central regions of galaxies has suggested that massive BHs should be present in a number of early-type galaxies (see Kormendy & Richstone 1995 for a review). Conclusive dynamical evidence for the presence of a

central BH requires that a model with a BH can fit all observations (photometric and kinematic), and that no model without a BH can provide an equally good fit. Such conclusive evidence can only be inferred from observations that probe well inside the radius where the BH dominates the dynamics. Up to a few years ago, most claimed BH detections were based on observations with spatial resolutions of similar size as the radii of influence of the inferred BH masses (Rix 1993). This, together with the limited amount of freedom in the models used to interpret the data, has hampered an unambiguous proof for the presence of these BHs (i.e., the observed kinematics could not be confronted with all possible dynamical configurations without a BH). Often spherical models were used even when the observed flattening was significant. If the models were axisymmetric, the distribution function (hereafter DF) was often assumed to depend only on the two classical integrals of motion, energy and vertical angular momentum; $f = f(E, L_z)$. This implies that the velocity dispersions in the radial and vertical directions are equal (i.e., $\sigma_R = \sigma_z$). It is well-known that strong radial anisotropy in the center of a galaxy results in a high central velocity dispersion, mimicking the presence of a massive BH (cf. Binney & Mamon 1982). Conclusive evidence for a BH therefore requires that one can rule out radial anisotropy as the cause of the high velocity dispersions measured, and models must thus be sufficiently general.

Recently two major breakthroughs have initiated a new era in the search for massive BHs in normal galaxies. First of all, we can now obtain kinematics at much higher spatial resolution (down to FWHM $\sim 0.1''$), using specially-designed spectrographs, such as the Subarcsecond Imaging Spectrograph (SIS) on the Canada-France-Hawaii Telescope, or the Faint Object Spectrograph (FOS) and STIS aboard the HST. This allows us to probe the gravitational potential much closer to the center, where the BH dominates the dynamics. Not only has this improved the evidence for massive BHs in several old BH-candidate galaxies (M31, Ford *et al.* 1998; M32, van der Marel *et al.* 1998, chapter 4 of this thesis; M87, Harms *et al.* 1994, Macchetto *et al.* 1997; NGC 3115, Kormendy *et al.* 1996a; NGC 4594, Kormendy *et al.* 1996b), but it has also provided new cases (M84, Bower *et al.* 1998; NGC 3377, Kormendy *et al.* 1998; NGC 3379, Gebhardt *et al.* 1999; NGC 4261, Ferrarese, Ford & Jaffe 1996; NGC 4486B, Kormendy *et al.* 1997; NGC 6251, Ferrarese, & Ford 1999; and NGC 7052, van der Marel & van den Bosch 1998). Secondly, the revolutionary increase in computer power has made it possible to investigate a large number of fully general, three-integral models based on the orbit-superposition method (Schwarzschild 1979). In the past decade, this method has been used to build a variety of spherical, axisymmetric and triaxial models (e.g., Schwarzschild 1982; Pfenniger 1984; Richstone & Tremaine 1984, 1988; Zhao 1996). Levison & Richstone (1985), Richstone & Tremaine (1985), and Pfenniger (1984) showed how to include rotation velocities and velocity dispersions as kinematic constraints. More recently, Rix *et al.* (1997, chapter 2 of this thesis) and Cretton *et al.* (1999, chapter 3 of this thesis) extended this modeling technique even further by fitting to the *entire* velocity profiles (see also Richstone 1997). Van der Marel *et al.* (1997) used this to build fully general, axisymmetric models of M32, and showed convincingly that M32 harbors a massive BH (see also chapter 4). Recent review papers on this rapidly evolving field include Ford *et al.* (1998), Ho (1998), Richstone (1998), and van der Marel (1997).

In many galaxies where the presence of a BH has been suggested, a nuclear disk, seen close to edge-on, is present. These disks are either in gaseous form (M84, M87, NGC 4261, NGC 4594, NGC 6251, NGC 7052), or made up of stars (NGC 3115). It is easier to detect BHs in edge-on systems with disks, where one can use both the measured rotation velocities and the velocity dispersions to determine the central mass density. It is therefore not surprising that BHs have predominantly been found in galaxies with nuclear disks. Furthermore, nuclear disks allow a good determination of the central mass density of their host galaxies. Gaseous disks have

the advantage that their kinematics can be easily measured from emission lines. Since gas in a steady-state disk can only move on non-intersecting orbits, the measured rotation velocities of a settled gas disk, in the equatorial plane of an axisymmetric potential, correspond to the circular velocities, $V_c(R) = \sqrt{R d\Phi/dR}$. The rotation curve of a nuclear gas disk therefore provides a direct measure of the central potential gradient, and thus of the central mass density. However, often the gas disks are not in a steady state; many show a distorted morphology (e.g. M87, see Ford *et al.* 1994), and non-gravitational motion, such as outflow, inflow or turbulence can be present and complicate the dynamical analysis (e.g., NGC 4261, Jaffe *et al.* 1996; NGC 7052, van den Bosch & van der Marel 1995). Nuclear *stellar* disks do not suffer from this, but have the disadvantage that their kinematics are much harder to measure. First of all, the kinematics have to be determined from absorption lines rather than emission lines, and secondly, the line-of-sight velocity distributions, or velocity profiles (VPs), measured are ‘contaminated’ by light from the bulge component. However, van den Bosch & de Zeeuw (1996) showed that with sufficient spatial and spectral resolution one can resolve the VPs in a broad bulge-component and a narrow disk-component. From these VPs the rotation curve of the nuclear disk can be derived, providing an accurate measure for the central mass density. Therefore, galaxies with an embedded nuclear disk (either gaseous or stellar) observed close to edge-on are ideal systems to investigate the presence of massive BHs.

In this paper we discuss the case of NGC 4342; a small, low-luminosity ($M_B = -17.47$) S0 galaxy in the Virgo cluster. The galaxy is listed as IC 3256 in both the Second and Third Reference Catalogues of Bright Galaxies, since in the past it has occasionally been confused with NGC 4341 and NGC 4343 (see Zwicky & Herzog 1966). At a projected distance of $\sim 30''$ SE of NGC 4342, a small galaxy is visible. It is uncertain whether this is a real companion of NGC 4342 or whether it is merely close in projection. HST images of NGC 4342 revealed both an outer disk, as well as a very bright nuclear stellar disk inside $\sim 1''$ (van den Bosch *et al.* 1994; Scorza & van den Bosch 1998). It is a normal galaxy, with no detected ISM (Roberts *et al.* 1991), and with small color gradients (van den Bosch, Jaffe & van der Marel 1998, hereafter BJM98). For its size and luminosity, it does however reveal a remarkably large central velocity dispersion and a very steep rotation curve (see BJM98). Unfortunately, the spectral resolution of the available kinematic data is insufficient to actually resolve the VPs in disk and bulge components. In order to determine the central mass density in NGC 4342, we thus have to construct dynamical models of the entire system: bulge and disk components. Here we present simple two-integral Jeans models as well as fully general three-integral models, and we provide evidence for the presence of a central massive dark object (MDO) of $\sim 3 \times 10^8 M_\odot$. Throughout this paper we assume the MDO to be a BH, but we discuss alternatives in Section 5.6.2.

In Section 5.1 we briefly discuss the data used to constrain the models and in Section 5.2 we describe our mass model. In Section 5.3 we show the results of some simple two-integral modeling, and we discuss its shortcomings. Section 5.4 describes the general outline of the three-integral modeling technique. In Section 5.5 we discuss shortcomings of the velocity profile parameterization used when applied to dynamically cold systems, and present a modified approach. The results of the three-integral modeling are discussed in Section 5.6. Finally, in Section 5.7, we sum up and present our conclusions. Throughout this paper we adopt a distance of 15 Mpc for NGC 4342, consistent with the distance of the Virgo cluster (Jacoby, Ciardullo & Ford 1990).

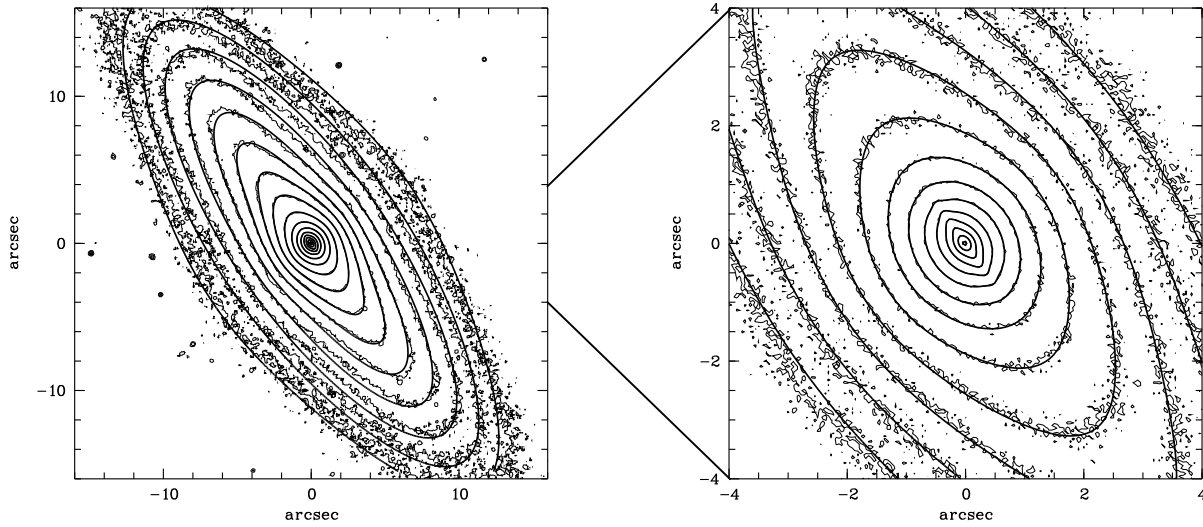


FIGURE 5.1— Contour maps of the WFPC2 *I*-band image of NGC 4342 at two different scales: $32'' \times 32''$ (left-hand panel) and $8'' \times 8''$ (right-hand panel). Superimposed are the contours of the MGE model of the intrinsic surface brightness convolved with the HST PSF (see Section 5.2).

5.1 The data

All data used in this paper are presented and discussed in detail in BJM98. Here we merely summarize.

5.1.1 Photometric data

BJM98 used the Wide Field and Planetary Camera 2 (WFPC2) aboard the HST to obtain *U*, *V* and *I* band photometry of NGC 4342. The spatial resolution of these images is limited by the HST Point Spread Function ($\text{FWHM} \sim 0.1''$) and the size of the pixels ($0.0455'' \times 0.0455''$). The full field-of-view covers about $35'' \times 35''$. Figure 5.1 shows contour plots at two different scales of the *I*-band image. The presence of the nuclear disk is evident from the highly flattened, disk-like isophotes inside $1.0''$.

5.1.2 Kinematic data

Using the ISIS spectrograph mounted at the 4.2m William Herschel Telescope (WHT) at La Palma, BJM98 obtained long-slit spectra of NGC 4342 along both the major and the minor axis. The spectra have a resolution of $\sigma_{\text{instr}} = 9 \text{ km s}^{-1}$, and were obtained with a slit width of $1.0''$ under good seeing conditions with a PSF FWHM of $0.80''$ (major axis) and $0.95''$ (minor axis). After standard reduction, the parameters $(\gamma, V, \sigma, h_3, h_4)$ that best fit the VPs were determined using the method described in van der Marel (1994). These parameters quantify the Gauss-Hermite (GH) expansion of the velocity profile $\mathcal{L}(v)$ as introduced by van der Marel & Franx (1993):

$$\mathcal{L}(v) = \frac{\gamma}{\sigma} \alpha(w) \left(1 + \sum_{j=3}^4 h_j H_j(w) \right), \quad (5.1)$$

where

$$w \equiv (v - V)/\sigma, \quad (5.2)$$

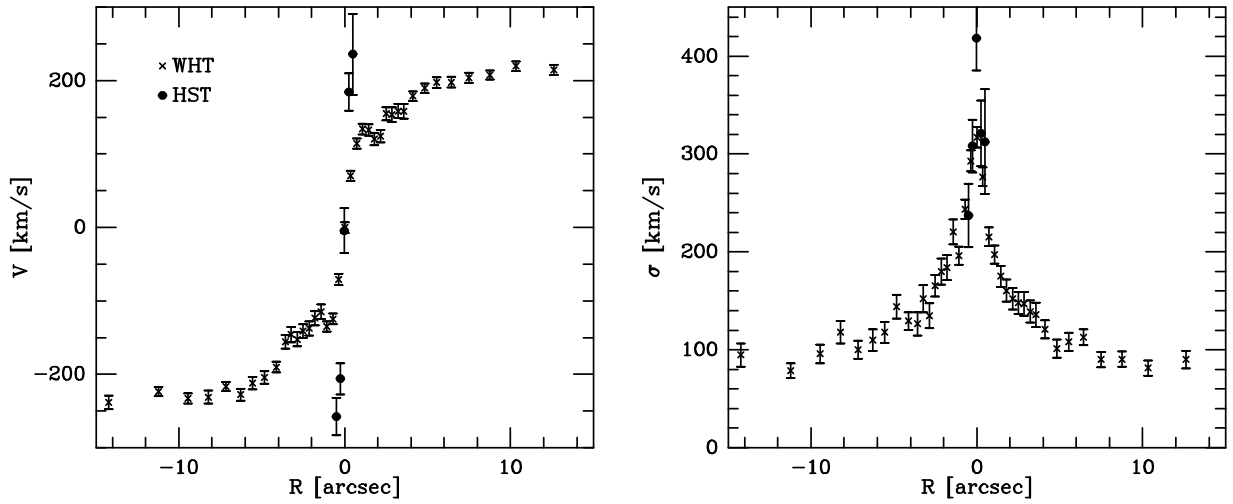


FIGURE 5.2— Observed rotation velocities V and velocity dispersions σ (as determined from the best-fitting Gaussian, see text) along the major axis of NGC 4342 obtained with the WHT (crosses) and the FOS (solid dots). The gradient of the rotation velocity and the central velocity dispersion increase considerably going to the four times higher spatial resolution of the FOS. See BJM98 for details on the data.

and

$$\alpha(w) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}w^2}. \quad (5.3)$$

Here v is the line-of-sight velocity, H_j are the Hermite polynomials of degree j , and h_j are the Gauss-Hermite coefficients. The first term in equation (5.1) represents a Gaussian with line strength γ , mean radial velocity V , and velocity dispersion σ . The even GH-coefficients quantify symmetric deviations of the VP from the best-fitting Gaussian, and the odd coefficients quantify the asymmetric deviations.

We have averaged the kinematic WHT data at positive and negative radii. In this way we obtain sets of (V, σ, h_3, h_4) at 19 different positions along the major axis and 8 along the minor axis.

BJM98 also obtained FOS spectra at 7 different aperture positions, all inside the central $0.5''$ of NGC 4342, using the circular $0.26''$ -diameter aperture (the FOS 0.3 -aperture). Due to the limited signal-to-noise ratio (S/N) of these spectra, only (γ, V, σ) of the best-fitting Gaussian could be determined. Rotation velocities and velocity dispersions along the major axis of NGC 4342 for both the WHT (crosses) and FOS (solid dots) are shown in Figure 5.2. The central rotation gradient, as measured with the FOS, is extremely steep ($V \sim 200 \text{ km s}^{-1}$ at $0.25''$ from the center). In addition, the velocity dispersion increases from $\sim 90 \text{ km s}^{-1}$ at the outside (the ‘cold’ outer disk) to 317 km s^{-1} in the center as measured with the WHT. The central velocity dispersion increases to 418 km s^{-1} , when observed at four times higher spatial resolution with the FOS.

5.2 The mass model

We have used the Multi-Gaussian Expansion (MGE) method developed by Emsellem, Monnet & Bacon (1994, hereafter EMB94) to build a mass model for NGC 4342. The method assumes that both the PSF and the intrinsic surface brightness are described by a sum of Gaussians, each of which has 6 free parameters: the center (x_j, y_j) , the position angle, the flattening q'_j , the central intensity I'_j , and the size of the Gaussian along the major axis, expressed by its standard

deviation a'_j . The best-fitting parameters of the different Gaussians are determined using an iterative approach in which additional components are added until convergence is achieved (see EMB94 for details on the method). This method is well suited for complicated, multi-component galaxies such as NGC 4342.

We fitted the HST *I*-band PSF by a sum of 5 circular (i.e., $q'_j = 1$) Gaussians (see BJM98). Using this model PSF we derived the parameters of the N Gaussians describing the intrinsic surface brightness (i.e., deconvolved for PSF effects) by fitting to the HST *I*-band image of NGC 4342. We forced the N Gaussians to have the same position angle and center, which yields an axisymmetric mass model (see below). Therefore, the model is described by $3N + 3$ free parameters, which are simultaneously fit to the image. We achieved convergence with $N = 11$ Gaussian components. The results of the fit are shown in Figure 5.1, where we show contour plots of the *I*-band image with superimposed contours of the convolved surface brightness of the MGE model. The fit is excellent, except for a small discrepancy at the outside. This is due to slight twisting of the isophotes at large radii (see BJM98). Since our model is axisymmetric, this cannot be modeled. Nevertheless, the discrepancy is small, and is unlikely to affect our conclusions on the dynamics of the central region. The parameters of the different Gaussian components are listed in Table 5.1.

The total luminosity of the MGE model in the *I*-band is $L_I = 3.57 \times 10^9 L_\odot$. This yields $M_I = -19.86$. The absolute blue magnitude of NGC 4342, at a distance of 15 Mpc, is $M_B = -17.47$ (Sandage & Tammann 1981), and we thus find $B - I = 2.39$. This is consistent with the colors of NGC 4342 presented by BJM98. They find $U - V \approx 1.5$ and $V - I \approx 1.3$. We thus derive $B - V \approx 1.09$, in good agreement with the average value for early-type galaxies (Faber *et al.* 1989). If we assume that the luminosity distribution of the bulge corresponds to the Gaussian components rounder than $q'_j = 0.3$, we find that the bulge makes up ~ 52 per cent of the total luminosity of NGC 4342. The outer disk, described by Gaussian components 9 and 10, makes up an additional 46.5 per cent, and the nuclear disk (modeled by Gaussian component 4) adds only about 1.5 per cent to the total luminosity. There is no reason that the mathematical components correspond to actual physical components, but at least this gives an order-of-magnitude description of the luminosities of the bulge and the two disk components. A more accurate

TABLE 5.1— Parameters of MGE model for the deconvolved *I*-band surface brightness.

Index j	I'_j	a'_j	q'_j	$L_{I,j}$
(1)	(2)	(3)	(4)	(5)
1	490833.0	0.032	0.817	1.40×10^7
2	99417.9	0.101	0.865	2.92×10^7
3	67415.3	0.282	0.601	1.07×10^8
4	84108.1	0.343	0.136	4.47×10^7
5	28511.8	0.394	0.856	1.26×10^8
6	15529.3	0.753	0.622	1.82×10^8
7	8490.8	0.756	1.000	1.61×10^8
8	6055.4	1.866	0.665	4.66×10^8
9	2951.4	4.419	0.250	4.79×10^8
10	1572.8	9.229	0.266	1.18×10^9
11	229.9	11.854	0.723	7.76×10^8

Col. (1): Index number of each Gaussian; col. (2) its central surface brightness; col. (3) its standard deviation (which expresses the size of the Gaussian along the major axis); col. (4) its flattening; and col. (5) its total *I*-band luminosity. All Gaussians have the same position angle and the same center.

disk–bulge decomposition, which yields similar results, is discussed in Scorza & van den Bosch (1998).

Assuming that the density is built up from a sum of three–dimensional Gaussians stratified on spheroids, one can, for any inclination angle i , analytically calculate the density distribution from the MGE fit to the intrinsic surface brightness. The mass density of such an MGE model is given by

$$\rho(R, z) = Y \sum_j I_j \exp \left[-\frac{1}{2a_j^2} \left(R^2 + \frac{z^2}{q_j^2} \right) \right], \quad (5.4)$$

where Y is the mass–to–light ratio, and I_j , a_j and q_j are related to I'_j , a'_j , q'_j and i (see EMB94). The potential that corresponds to this density distribution follows from solving the Poisson equation. This yields

$$\Phi(R, z) = -4\pi G Y \sum_j a_j^2 q_j I_j \int_0^1 \exp \left[-\frac{t^2}{2a_j^2} \left(R^2 + \frac{z^2}{1 - e_j^2 t^2} \right) \right] \frac{dt}{\sqrt{1 - e_j^2 t^2}}, \quad (5.5)$$

where $e_j^2 = 1 - q_j^2$.

The inclination angle is well constrained by the thinness of the nuclear disk: $i > 83^\circ$ (Scorza & van den Bosch 1998). Throughout we assume that NGC 4342 is observed edge–on (i.e., $i = 90^\circ$). Given the lower limit on the inclination angle of 83° , this assumption does not significantly influence the conclusions presented in this paper.

5.3 Jeans modeling

5.3.1 Formalism

The three–integral modeling described in the next section requires large amounts of CPU time. We therefore decided to first explore parameter space of the models (i.e., mass–to–light ratio and mass of the possible BH) by solving the Jeans equations and assuming that the phase–space distribution function depends only on the two classical integrals of motion. The Jeans equations for hydrostatic equilibrium are moment equations of the collisionless Boltzmann equation (see Binney & Tremaine 1987). They relate the velocity dispersion tensor $\hat{\sigma}^2$ and the streaming motion \mathbf{v} to the density ρ and potential Φ . For an axisymmetric system with distribution function $f(E, L_z)$ one always has $\sigma_R = \sigma_z$ and $\overline{v_R v_z} = 0$, and the Jeans equations in cylindrical coordinates reduce to

$$\frac{\partial(\rho \sigma_R^2)}{\partial R} + \rho \left(\frac{\sigma_R^2 - \overline{v_\phi^2}}{R} + \frac{\partial \Phi}{\partial R} \right) = 0, \quad (5.6)$$

$$\frac{\partial(\rho \sigma_z^2)}{\partial z} + \rho \frac{\partial \Phi}{\partial z} = 0. \quad (5.7)$$

Here $\overline{}$ denotes the local average over velocities. From equation (5.7) and $\rho(R, z)$ and $\Phi(R, z)$, one can, at every point (R, z) in the meridional plane, calculate $\sigma_R^2 (= \sigma_z^2)$ by simple integration. By rewriting equation (5.6), one can compute $\overline{v_\phi^2} = \overline{v_\phi^2} + \sigma_\phi^2$ without the need of performing a numerical (ill–conditioned) derivative (see e.g. Hunter 1977; Simien, Pellet & Monnet 1979; Binney, Davies & Illingworth 1990).

The expressions for σ_R^2 and $\overline{v_\phi^2}$ for the density distribution of equation (5.4) are given in EMB94 (their equations [42] and [44]). The Jeans equations do not prescribe how $\overline{v_\phi^2}$ splits in

streaming motion $\overline{v_\phi}$ and azimuthal velocity dispersion σ_ϕ . We follow the approach introduced by Satoh (1980), and write

$$\overline{v_\phi} = k\sqrt{\overline{v_\phi^2} - \sigma_R^2}, \quad (5.8)$$

such that we can control the anisotropy σ_ϕ/σ_R by means of the free parameter k . For $k = 1$ the model is fully isotropic with $\sigma_\phi = \sigma_R = \sigma_z$. Once k has been fixed one can project the luminosity-weighted dynamical quantities on the plane of the sky (x', y') to yield the projected rotation velocities

$$V_{\text{rot}}(x', y') = \frac{-1}{S(x', y')} \int_{-\infty}^{\infty} \nu \overline{v_\phi} \sin i \cos \phi \, dz', \quad (5.9)$$

and the rms velocities

$$V_{\text{rms}}^2(x', y') = \frac{1}{S(x', y')} \int_{-\infty}^{\infty} \nu \left(\sigma_z^2 \cos^2 i + \sigma_R^2 \sin^2 \phi \sin^2 i + (\sigma_\phi^2 + \overline{v_\phi^2}) \cos^2 \phi \sin^2 i \right) dz'. \quad (5.10)$$

Here $\nu = \rho/\Upsilon$ is the luminosity density and $S(x', y')$ is the projected surface brightness at position (x', y') . V_{rot} and V_{rms}^2 are the true first and second order moments of the line-of-sight velocity distribution. The projected velocity dispersion $\sigma_p(x', y')$ is simply derived from

$$\sigma_p(x', y') = \sqrt{V_{\text{rms}}^2(x', y') - V_{\text{rot}}^2(x', y')}. \quad (5.11)$$

It is straightforward to include a BH in such a model, by simply adding $-GM_{\text{BH}}/\sqrt{R^2 + z^2}$ to the stellar potential (5.5).

5.3.2 Application to NGC 4342

We use the Jeans equations to calculate the predicted rotation velocities and velocity dispersions for the luminosity distribution of NGC 4342. We assume that the stellar mass-to-light ratio Υ and the anisotropy parameter k are constant throughout the galaxy. We calculate V_{rot} and V_{rms}^2 (using equations [5.9] and [5.10]) on a two-dimensional grid on the sky. The grid is logarithmic in r (in order to properly sample the strong gradients near the center), and linearly sampled in θ . Once V_{rot} and V_{rms}^2 are tabulated, we convolve them with the PSF of the observations, weighted by the surface brightness. After pixel binning, taking the proper slit width into account, these are compared to the observations.

The V and σ determined from the GH fitting to the WHT VPs cannot be compared directly to V_{rot} and σ_p derived from the modeling discussed above: the latter ones correspond to the true moments of first and second order of the VPs, whereas the former ones correspond to the best-fitting Gaussian. We therefore recalculated the VP from V , σ , h_3 and h_4 , from which we then estimate the first and second order moments for direct comparison with the Jeans models.

The results are shown in Figure 5.3, where we plot V_{rot} , V_{rms} and σ_p of the VPs along the major axis for both the WHT and the FOS data. Also plotted are predictions for four models, that only differ in the mass of the central BH (0, 3, 5 and $10 \times 10^8 M_\odot$). All models have $i = 90^\circ$, $\Upsilon_I = 6.2 M_\odot / L_\odot$ and $k = 1$ (i.e., all models are fully isotropic). For this value of k , we obtain the best fit to the observed velocity dispersion outside $\sim 2''$. However, V_{rot} is not very well fitted: the wiggles in the rotation curve are not reproduced by the model. One can alter k , as function of radius, such that we fit these wiggles, but at the cost of introducing them in the velocity dispersion profile. This is due to the poor fit of the Jeans models to the rms velocities (see lower left panel of Figure 5.3). V_{rms}^2 depends only on the sum of σ_ϕ^2 and $\overline{v_\phi^2}$ (see equation [5.10]) and is therefore independent of k . Consequently, the Jeans models cannot simultaneously fit V_{rot} and

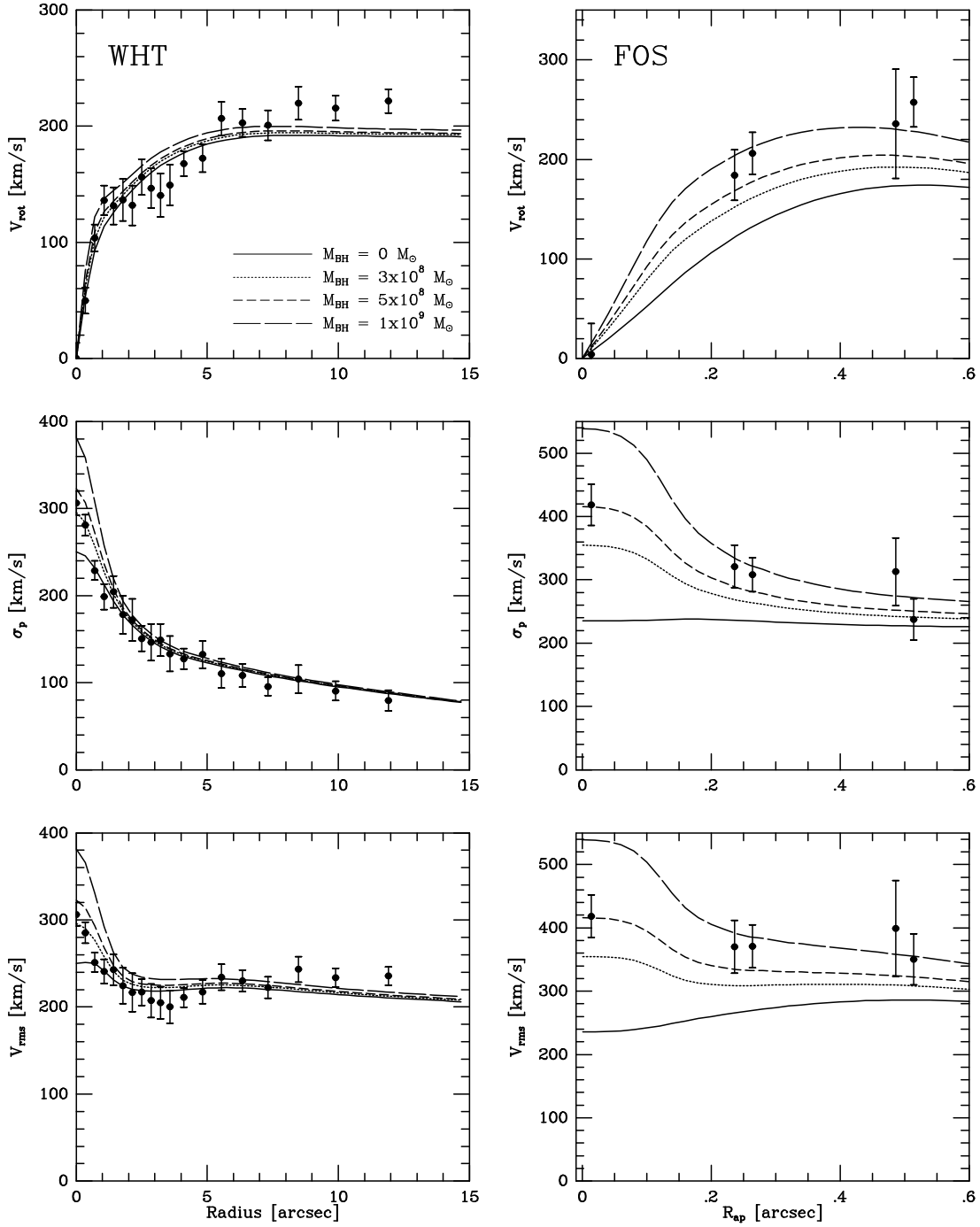


FIGURE 5.3— Results of the Jeans modeling. The solid dots with errorbars indicate the observed rotation velocities and velocity dispersions. Overplotted are four models that differ only in the mass of the central BH ($0, 3, 5$ and $10 \times 10^8 M_{\odot}$). All models have fully isotropic velocity dispersions, and a stellar mass-to-light ratio of $\Upsilon_l = 6.2 M_{\odot} / L_{\odot}$. The three panels on the left show the WHT kinematics and the model predictions for V_{rot} , σ_p and V_{rms} . The model with $M_{\text{BH}} = 3 \times 10^8 M_{\odot}$ provides the best fit to the velocity dispersions. Neither of the four models provides an accurate fit to the rotation velocities (see also Figure 5.4). The panels on the right compare the model predictions with the HST/FOS kinematics. Both the rotation velocities and the velocity dispersions suggest the presence of a BH with a mass of $\sim 6 \times 10^8 M_{\odot}$.

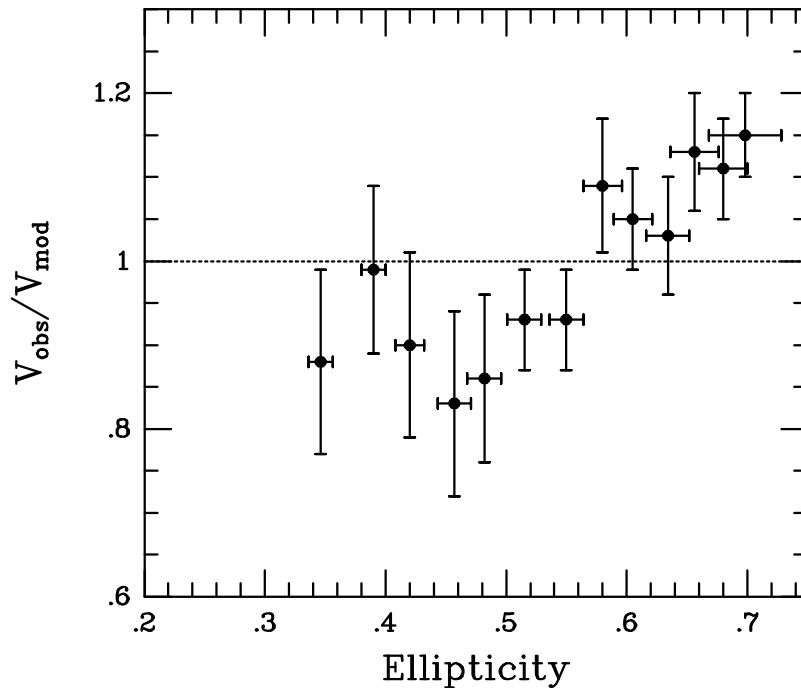


FIGURE 5.4— The ratio $V_{\text{obs}}/V_{\text{mod}}$ of the observed rotation velocities over the rotation velocities predicted by the isotropic Jeans model with $Y_I = 6.2 M_{\odot}/L_{\odot}$ and $M_{\text{BH}} = 3 \times 10^8 M_{\odot}$, as function of the local, projected ellipticity of the isophotes. Only results beyond $2''$ are shown. At smaller radii seeing influences the observed quantities significantly. There is a clear correlation in that the Jeans model underpredicts the rotation velocities in the moderately flattened region, and overpredicts the rotation velocities in the radial interval where the isophotes are strongly flattened, and thus dominated by the light of the outer disk component.

σ_p along the major axis. This suggests that the assumption made, i.e., $f = f(E, L_z)$, is wrong, and that three-integral models are required.

The Jeans models without a central BH clearly underpredict the central velocity dispersion (for both the WHT and the FOS measurements), as well as the central rotation gradient measured with the FOS. Models with a massive BH provide a much better fit. The actual mass of the BH depends on the data set used to constrain the model: the WHT data suggests a BH mass of $\sim 3 \times 10^8 M_{\odot}$, whereas the FOS data are best fitted with $M_{\text{BH}} \sim 6 \times 10^8 M_{\odot}$. So although the two-integral Jeans modeling cannot fit all the observed kinematics, it does suggest that a BH of a few times $10^8 M_{\odot}$ may be present in the center of NGC 4342.

In Figure 5.4 we plot the ratio of the observed rotation velocity over the rotation velocity of the best-fitting isotropic Jeans model, $V_{\text{obs}}/V_{\text{mod}}$ (along the major axis), versus the local observed ellipticity (in I -band) of NGC 4342. There is a clear correlation in the sense that the Jeans models underpredict the rotation velocity in the strongly flattened region, and overpredict V_{rot} in the less flattened region. The ellipticity is a measure of the local disk-to-bulge ratio, and this therefore suggests that disk and bulge have different velocity anisotropies. Since V_{rot} scales with \sqrt{T} , another possibility may be that the disk and bulge are made up of different stellar populations (whose mass-to-light ratios are different by almost a factor two). However, the separate components (bulge, nuclear disk and outer disk) do not stand out as separate entities in either the $U - V$ or the $V - I$ color images (see BJM98), rendering this explanation improbable.

5.4 Three-integral modeling

In order to investigate the presence of a $\sim 3\text{--}6 \times 10^8 M_\odot$ BH as suggested by the Jeans models, we now construct fully general, axisymmetric models of NGC 4342. We use an extension of Schwarzschild's (1979) orbit-superposition method (see de Zeeuw 1997), described in chapter 2 and 3 of this thesis. The main method, but for spherical systems, is outlined in chapter 2. The application to axisymmetric systems is discussed in chapter 3. Here we briefly outline the method, and we refer the interested reader to the relevant chapters for details and tests of this modeling technique.

5.4.1 The method

The first step of the method is to integrate orbits in the combined potential $\Phi_{\text{stars}} + \Phi_{\text{BH}}$. Each orbit is then projected onto the space of observables, taking convolution with the PSF and pixel binning into account. Finally, a non-negative least-squares algorithm is used to determine the distribution of orbit weights that best fits the observational data (taking the observational errors into account), while also reproducing the luminous density distribution of the model.

Throughout we limit ourselves to models with an inclination angle $i = 90^\circ$, and we assume that the stellar population has a constant mass-to-light ratio. Therefore, each model is characterized by only two free parameters: the mass-to-light ratio, Y_I , and the mass of the black hole, M_{BH} . Our aim is to find the set (M_{BH}, Y_I) that best fits the available constraints (surface brightness and velocity profiles).

5.4.2 The orbit library

The motion of a star in an axisymmetric potential, for which E and $L_z = Rv_\phi$ are conserved, can be reduced to motion in the meridional (R, z) -plane, in the effective potential $\Phi_{\text{eff}} = \Phi(R, z) + L_z^2/(2R^2)$. The orbit is constrained within a region bounded by the zero velocity curve (ZVC) defined through $E = \Phi_{\text{eff}}(R, z)$.

Each orbit in an axisymmetric potential $\Phi(R, z)$ admits two integrals of motion: energy $E = \Phi(R, z) + \frac{1}{2}v^2$ and vertical angular momentum $L_z = Rv_\phi$. Regular orbits admit one additional integral, I_3 , which in general is not known analytically. Such an orbit is confined to a sub-space inside the ZVC. We only found a very small fraction of our orbit catalog to be irregular.

The orbit library has to be set up such that one properly samples the full extent of phase space. The sampling has to be sufficiently dense to suppress discreteness noise, but is limited by the amount of available CPU time. After some testing we have chosen to calculate orbits on a $20 \times 20 \times 7$ (E, L_z, I_3)-grid. Each energy is uniquely defined by a circular radius R_c according to

$$E = \frac{1}{2}R_c \left. \frac{\partial \Phi}{\partial R} \right|_{R=R_c} + \Phi(R_c, 0). \quad (5.12)$$

The 20 energies are sampled logarithmically between $R_c(E) = 0.01''$ and $R_c(E) = 60''$. These values were chosen such as to encompass the major fraction of the total mass of the galaxy. The mass inside $R_c = 0.01''$ is only a fraction of 3.25×10^{-5} of the total mass, whereas a fraction of $\sim 2.85 \times 10^{-6}$ of the mass is located outside $R_c = 60''$. For each energy, we calculate the maximum vertical angular momentum

$$L_{z,\text{max}}(E) = \sqrt{R_c^3 \left(\frac{\partial \Phi}{\partial R} \right)_{(R_c, 0)}}, \quad (5.13)$$

which corresponds to the circular orbit with energy E , and sample $|\eta| = |L_z|/L_{z,\max}$ on a linear grid of 10 values between 0.01 and 0.99. Hence, the purely circular and radial orbits are presumed to be represented by their closest neighbors on the grid, but are not explicitly calculated. At each value of $|\eta|$, only the orbit with positive angular momentum has to be integrated, since its counterpart ($L_z = -\eta L_{z,\max}$) is simply a mirror reflection around zero velocity. Since the third integral I_3 can generally not be expressed explicitly in terms of the phase-space coordinates, we use the method suggested by Levison & Richstone (1985), and take the starting point on the ZVC as a numerical representation of the third integral (see also chapter 3). For that purpose we first calculate, for each (E, L_z) -pair, the locus $(R_{\text{tt}}, z_{\text{tt}})$ on the ZVC of the ‘thin tube orbit’. For a certain value of E and L_z , this is the only orbit that touches the ZVC at only one value of R . All other regular orbits touch the ZVC at two different values of R . We sample I_3 by linearly sampling the R -coordinate of the orbit’s starting point on the ZVC (R_{ZVC}) between R_{tt} and R_{\max} , where R_{\max} is the maximum extent of the ZVC in the equatorial plane.

Since the calculation of the potential (equation [5.5]) and the forces all require the evaluation of a numerical quadrature, we have calculated them on a 4000×300 (R, θ) -grid in the meridional plane. The grid is sampled linearly in θ between 0 and $\pi/2$, and logarithmically in R between 10^{-4} and 10^3 arcseconds. Each orbit is integrated for 200 radial periods, using linear interpolation between grid points to evaluate the potential and forces. On average, the energy conservation over 200 radial periods is better than one part in 10^5 , justifying the interpolation scheme adopted. In total $20 \times 10 \times 7 = 1400$ orbits are integrated, resulting in a library of $2 \times 1400 = 2800$ orbits (when doubling for the $-L_z$ orbits).

During the integration of each orbit we project its phase-space coordinates onto the space of observables (x', y', v_{los}) , where (x', y') is the plane on the sky, and v_{los} is the line-of-sight velocity. In the following we use v as shorthand for v_{los} . We adopt a three-dimensional grid in the (x', y', v) -space, i.e., our storage cube, in which we record the fractional time the orbit spends in each of the cells (see chapter 2 for details). Once the orbit integration is finished, we convolve each of the velocity slices (x', y') with the PSF appropriate for the observations. Since we have kinematic constraints obtained with three different instrumental setups and different PSFs (WHT major axis, WHT minor axis, and FOS apertures) we use three separate (x', y', v) -storage cubes each of which is convolved with the respective PSF. We use two cubes with $0.1'' \times 0.1''$ (x', y') -cells for the WHT major and minor axes. For the FOS cube, $0.05'' \times 0.05''$ cells are used to comply with the higher spatial resolution of the HST. For all storage cubes we use 101 velocity bins of 30 km s^{-1} . The final step is to calculate the contribution of the orbit to each of the positions in the plane of the sky where we have photometric and/or kinematic constraints. These positions are in general extended areas (e.g., determined by the pixel size of the CCD, the slit width and the pixel rebinning used to obtain spectra of sufficient S/N). For each constraint position $l = 1, \dots, N_c$ we therefore sum the fractional times over the area of l (see chapter 2 for details). This gives us in the end, for each orbit k and each constraint position l , the properly PSF-convolved velocity profile ‘histogram’ $\text{VP}_{l,v}^k$, integrated over the area of position l . By using *fractional* times we ensure that each orbit is normalized to unity.

5.4.3 The observational constraints

The final step of the orbit-superposition method is to find the set of non-negative weights, γ_k , of each orbit that best matches the kinematic constraints and reproduces the luminous mass density of the model. Since we normalize each orbit to unity, the orbital weights γ_k measure the fraction of the total light of the galaxy that resides on orbit k . We use the following sets of constraints:

The solution γ has to reproduce the luminosity density $\rho(R, z)/Y$ (equation [5.4]). We have

subdivided the first quadrant of the meridional plane in 20×5 (r, θ) -cells, with r and θ the standard spherical coordinates. The grid that encompasses those cells is binned linearly in θ , and logarithmically in r between $0.01''$ and $60''$. For each orbit k we store the fractional time t_n^k spent in cell n . Since we integrate the orbits in the meridional plane, this is similar to the fractional time the orbit spends in the three-dimensional volume obtained by integrating the area of the cell over 2π radians in the ϕ -direction. Therefore we have computed the total luminosity, L_n , of the MGE model inside each volume n . We will refer to these constraints as the ‘self-consistency constraints’.

From the WHT spectra we obtained sets of $(V_l, \sigma_l, h_{3,l}, h_{4,l})$ at 19 positions l along the major axis, and 8 along the minor axis. The quantities are the luminosity weighted averages over the areas of constraint positions $l = 1, \dots, 27$. For each of these 27 positions we have calculated the surface brightness S_l integrated over that area, and convolved with the appropriate PSF. The quantities S_l , V_l , σ_l , $h_{3,l}$ and $h_{4,l}$ parameterize the velocity profiles $\mathcal{L}_l^{\text{obs}}(v)$ with

$$S_l = \int_{-\infty}^{\infty} \mathcal{L}_l^{\text{obs}}(v) dv. \quad (5.14)$$

One can rewrite this parameterization in the form $(S_l, S_l h_{1,l}, S_l h_{2,l}, S_l h_{3,l}, S_l h_{4,l})$, with

$$S_l h_{m,l} = 2\sqrt{\pi} \int_{-\infty}^{\infty} \mathcal{L}_l^{\text{obs}}(v) \alpha(w_l) H_m(w_l) dv, \quad (5.15)$$

where $m = 1, \dots, 4$, α is again the standard Gaussian (see equation [5.3]), and

$$w_l = \frac{v - V_l}{\sigma_l}, \quad (5.16)$$

with V_l and σ_l the measured rotation velocity and velocity dispersion of the VP’s best-fitting Gaussian at constraint position l . Note that with this definition, $h_{1,l}$ and $h_{2,l}$ are zero for all constraint positions. This parameterization has the advantage that the orbit-superposition problem for the orbit weights is linear (see chapter 2).

In addition, we obtained (V_l, σ_l) at 7 FOS aperture positions. Again we have calculated, for each of these positions l , the PSF-convolved, aperture-integrated surface brightness S_l . The PSF used and the aperture diameter adopted are described in BJM98. As for the WHT spectra, we use the parameterization $(S_l, S_l h_{m,l})$ with $m = 1, 2$ as constraints rather than (S_l, V_l, σ_l) .

In total we thus have 100 self-consistency constraints L_n , 34 constraints on projected surface brightness S_l , and 122 kinematic constraints $S_l h_{m,l}$.

5.4.4 Non-negative least squares fitting

At each constraint position l we have a measured velocity profile $\mathcal{L}_l^{\text{obs}}(v)$, which we parameterized by $(S_l, S_l h_{m,l})$. For each of the observational constraints, we have the measurement errors $\Delta S_l, \Delta V_l, \Delta \sigma_l, \Delta h_{3,l}$ and $\Delta h_{4,l}$ from which we can calculate the errors $\Delta(S_l h_{m,l})$. As described in Section 5.4.2 we determined, for each orbit k and each position l , the velocity profile $\text{VP}_{l,v}^k$ at velocity v . We parameterize each orbital VP by $(S_l^k, S_l^k h_{m,l}^k)$ with $m = 1, \dots, 4$ using equations (5.14) and (5.15) and with $\mathcal{L}_l^{\text{obs}}(v)$ replaced by $\text{VP}_{l,v}^k$ and by changing the integration to a summation over all velocity bins. The orbit weights γ_k ($k = 1, \dots, N_{\text{orbits}}$) that result in the best fit to the observations can be determined by minimizing

$$\chi_{\text{obs}}^2 = \sum_l \left(\frac{S_l - \sum_k \gamma_k S_l^k}{\Delta S_l} \right)^2 + \sum_{m=1}^4 \sum_l \left(\frac{S_l h_{m,l} - \sum_k \gamma_k S_l^k h_{m,l}^k}{\Delta(S_l h_{m,l})} \right)^2. \quad (5.17)$$

In addition to minimizing χ_{obs}^2 , we also want the solution to match the luminosity density in the meridional plane, i.e., we also want to minimize

$$\chi_{\text{sc}}^2 = \sum_n \left(\frac{L_n - \sum_k \gamma_k t_n^k}{\Delta L_n} \right)^2, \quad (5.18)$$

where ΔL_n sets the accuracy for reproducing the luminosity density in the meridional plane. Throughout, we set $\Delta S_l = 0.01 S_l$ and $\Delta L_l = 0.01 L_l$, such that we aim for an accuracy of one per cent in reproducing both the projected PSF-convolved surface brightness and the luminosity density in the meridional plane. It is in principle sufficient to fit only the luminous density in the meridional plane: the surface brightness should be fitted automatically. However, because of discretization this is in practice not necessarily so, and we thus include the surface brightness at the constraint positions l as separate constraints.

Minimizing $\chi^2 \equiv \chi_{\text{obs}}^2 + \chi_{\text{sc}}^2$ is a least-squares problem for a matrix equation (see chapter 2 for a detailed description of the matrices involved). It has to be solved under the physical constraint that $\gamma_k \geq 0$. Following Pfenniger (1984), we use the Non-Negative Least Squares (NNLS) algorithm by Lawson & Hanson (1974) to solve for the orbit weights (see also chapters 2, 3 and 4).

5.4.5 Regularization

The NNLS matrix equation solved is numerically rather ‘ill-conditioned’ giving rise to a distribution function with strong oscillatory behavior. Such DFs are unphysical (e.g., Lynden-Bell 1967; Spergel & Hernquist 1992; Merritt 1993). Smoothing in the solution space can be achieved via regularization (e.g., Merritt 1996 and references therein). We follow the scheme used by Zhao (1996), which is based on a minimization, up to a certain degree, of the differences in weights between neighboring orbits. The technique is described in chapters 2 and 3, and we refer the interested reader to those papers for details. The extra regularization constraints result in a less good fit to the data. The amount of smoothing is set such that the regularized model is still compatible with the data in a statistical sense, i.e., such that $\Delta\chi^2 = \chi^2 - \chi_{\text{min}}^2 = 1$, with χ_{min}^2 the value obtained without regularization. Unless mentioned otherwise, we discuss results in which no regularization has been adopted.

5.5 Shortcoming of the Gauss-Hermite parameterization

Although the modeling technique outlined in the previous sections works well for dynamically hot galaxies (e.g., M32, see chapter 4), some problems arise when trying to apply it to dynamically cold systems (such as NGC 4342) because we do not include the zeroth-order moment h_0 in the fit. This quantity measures the normalization of the best-fitting Gaussian to the normalized VP and is observationally inaccessible: it is directly proportional to the unknown difference in line strength between the galaxy spectrum and the template spectrum used to analyze. (In practice one uses the assumption $h_0 = 1$ to estimate the line strength from the observations.) As we now show, excluding h_0 from the GH series and expanding the orbital VPs around the observed VPs up to fourth order only, can lead to artificial counter-rotation in models of cold systems.

Assume we have an observed velocity profile VP_{obs} , at a certain position on the sky, that is perfectly Gaussian. The GH moments h_m with $m > 0$ of such a VP_{obs} will all be zero. In the method described above, we derive the orbital h_m^k ($m = 1, \dots, 4$) from equation (5.15) in which the observed V and σ enter in the weighting function $\alpha(w)H_m(w)$. In the NNLS algorithm we

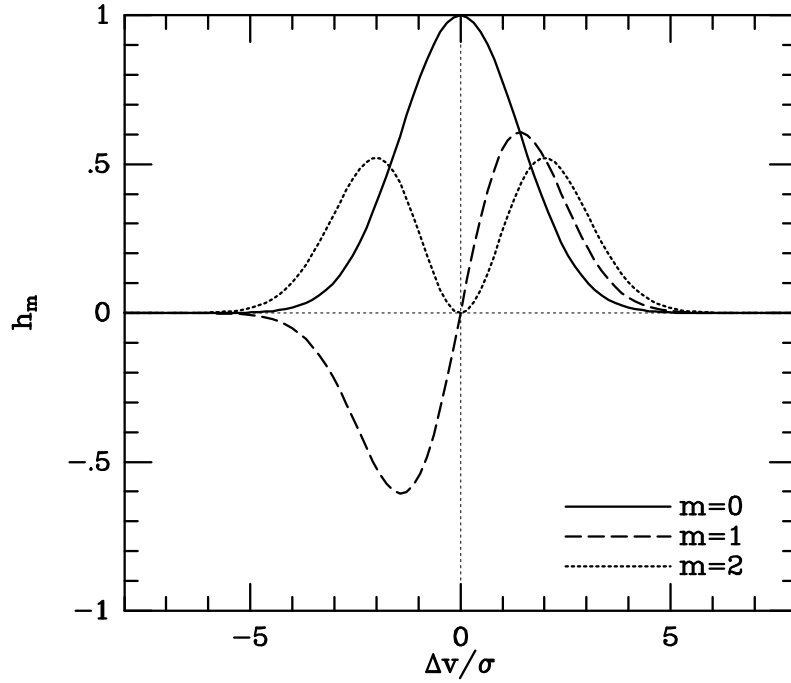


FIGURE 5.5— The Gauss–Hermite moments h_m ($m = 0, 1, 2$) of a Gaussian VP expanded around another Gaussian with the same dispersion σ . Results are plotted as functions of the velocity difference ΔV between the two Gaussians, expressed in units of σ . For $\Delta V = 0$, the two Gaussians are identical and $h_0 = 1$ and $h_m = 0$ ($m = 1, 2, 3, \dots$).

solve for the orbit weights by minimizing the difference between the GH moments of VP_{obs} and VP_{orb} . In principle, the differences between the $S_l h_{m,l}$ and $\sum_k \gamma_k S_l^k h_{m,l}^k$ are minimized, but for simplicity we illustrate the problem with a one-orbit model.

In the ideal case, an orbit whose VP deviates more strongly from the observed VP will be assigned a smaller weight. However, this is not always the case with the VP parameterization described in Section 5.4.3. To illustrate this we calculated the GH moments h_m ($m = 0, \dots, 2$) of a Gaussian expanded around another Gaussian, both with the same dispersion σ . In Figure 5.5 we plot the resulting moments h_m as function of the velocity difference ΔV between the two Gaussians (in units of σ). For $\Delta V = 0$, $h_0 = 1$ and all higher-order moments are zero (i.e., the two Gaussians are identical). In the regime $|\Delta V|/\sigma \lesssim 2$ the higher-order moments increase with increasing velocity difference. For larger values of $|\Delta V|/\sigma$ they start to decrease again to reach approximately zero for $|\Delta V|/\sigma \gtrsim 5$. On the contrary, h_0 equals unity when the VPs are identical, and decreases monotonically for increasing $|\Delta V|/\sigma$.

In NGC 4342, which is a dynamically cold system, the major axis kinematics reach $V_{\text{obs}}/\sigma_{\text{obs}} > 2.5$ at the outside (see Figure 5.6). An orbital VP with approximately the same V and σ as an observed VP with $V/\sigma > 2.5$ will have its h_1 and h_2 close to zero. Consequently, it will likely be given a non-zero weight. The same orbit, but with opposite sense of rotation (i.e., with reversed vertical angular momentum), will also have h_1 and h_2 close to zero, since $|\Delta V|/\sigma > 5$. In other words, the $+L_z$ and $-L_z$ orbits have the same S , h_1 and h_2 and are therefore indistinguishable for the fitting algorithm. The $+L_z$ and $-L_z$ orbits have opposite values of h_3 , but if $|h_3|$ is small, the difference between the two orbits, in terms of the VP parameterization used, remains small. If it would have been possible to include h_0 in the fit, then the $+L_z$ would have h_0 close to unity, and the $-L_z$ orbit would have $h_0 \approx 0$ (see Figure 5.5): the two VPs *would* have been easily distinguishable. In addition, if the observational data had been of sufficient S/N to derive GH

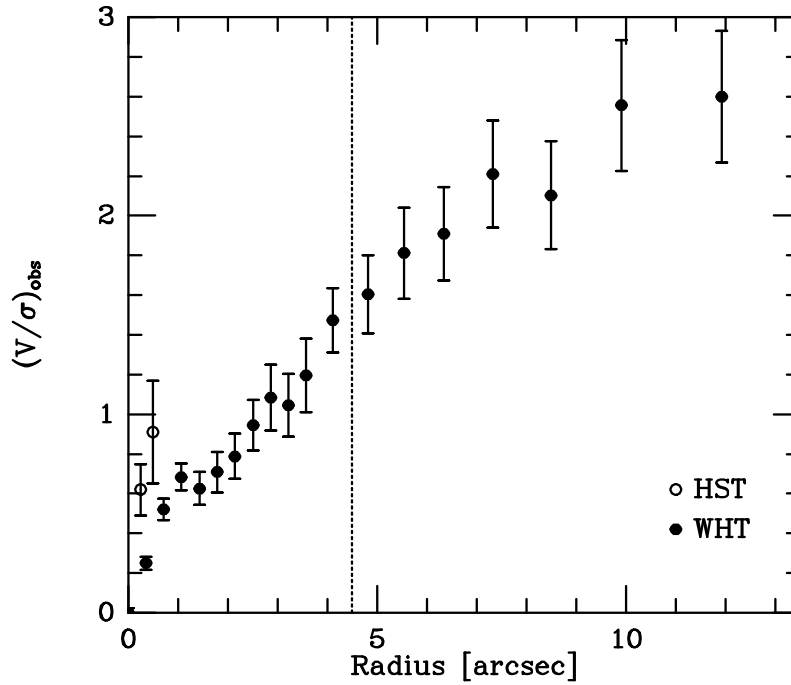


FIGURE 5.6— The ratio V/σ as function of radius of the major axis kinematics of NGC 4342 as measured with the WHT (solid dots) and the HST/FOS (open circles). The dotted line indicates R_{lim} outside which $V/\sigma > 1.5$. In the modified approach, we exclude counter-rotating orbits with circular radii $R_c > R_{\text{lim}}$ from the NNLS fit.

moments up to very high order, it would also have been possible to distinguish the VPs of the $+L_z$ and $-L_z$ orbits. In practice, however, one requires unrealistic high S/N spectra to be able to measure these moments.

We can thus expect that our solutions will have significant amounts of counter-rotation at radii where $V_{\text{obs}}/\sigma_{\text{obs}} \gtrsim 2$. We indeed found solutions in which the reconstructed VPs at large radii along the major axis have two peaks at positive and negative v_{los} (see Section 5.6.1). In chapter 4, we applied the same modeling technique to the dynamically hot system M32, which has $V/\sigma < 1.0$ everywhere, and therefore did not encounter this problem.

To solve the problem outlined above, we use a ‘modified approach’ in which we add some extra constraints to the models: we exclude counter-rotating orbits whose circular radius R_c is larger than a limiting radius R_{lim} . We have chosen $R_{\text{lim}} = 4.5''$, since for larger radii $V_{\text{obs}}/\sigma_{\text{obs}} > 1.5$ (see Figure 5.6) and the VP parameterization used becomes poor in the light of the problem discussed above. As we show in Section 5.6.1, this solves the problem of the artificial counter-rotation at large radii along the major axis, but it has the disadvantage that our models are not fully general any more: we have imposed some a priori constraints on the amount of counter-rotating orbits in NGC 4342. On the other hand, the actual observations do not reveal any counter rotation at large radii. Critically, one may argue that the VP analysis of the spectra is not suitable to detect such counter rotation, since we do not go to sufficient high order in the GH expansion. We have therefore also analyzed our spectra with the unresolved Gaussian decomposition (UGD) method (Kuijken & Merrifield 1993), which is not hampered by this limitation (e.g., Merrifield & Kuijken 1994), and found good agreement with the results of the GH parameterization. This suggests that indeed NGC 4342 has a negligible amount of counter-rotating stars in its outer disk. The additional constraint imposed on the models, when using the modified approach, is thus justified observationally.

5.6 Results and discussion

5.6.1 The black hole and mass-to-light ratio

Based on the BH mass and mass-to-light ratio suggested by the Jeans modeling (Section 5.3.1) we construct three-integral models with M_{BH}/Y_I in the range $0 - 1.5 \times 10^8 M_\odot$ and $4 \leq Y_I \leq 9$. For each value of M_{BH}/Y_I only one orbit library has to be constructed: a change in mass-to-light ratio is equivalent to a scaling of the model velocities proportional to \sqrt{Y} . We therefore calculate 10 different orbit libraries, all with $Y_I = 1$, that differ only in the mass of the central BH. If one scales to another value of Y_I , the mass of the BH changes accordingly to $Y_I M_{\text{BH}}$. We sample the mass-to-light ratio at 16 different values in the interval $Y_I \in [4, 9]$.

For each of the in total 160 different (M_{BH}, Y_I) -models we minimize $\chi^2 = \chi_{\text{obs}}^2 + \chi_{\text{sc}}^2$ (equations [5.17] and [5.18]). We use χ^2 statistics to compare different (M_{BH}, Y_I) -models in a proper statistical way. We determine the measure $\Delta\chi^2 \equiv \chi^2 - \chi_{\text{min}}^2$, where χ_{min}^2 is the overall lowest χ^2 . Under the assumptions that the errors are normally distributed and that there are no numerical errors in the model, one can assign confidence levels to the measure $\Delta\chi^2$. The exact level of confidence depends on the number of degrees of freedom in the models (Press *et al.* 1992), in our case two: M_{BH} and Y_I .

The resulting χ^2 -plots are shown in Figure 5.7. The first three contours show the formal 68.3, 95.4 and 99.73 per cent confidence levels (the latter one is plotted with a thick contour). The solid dots correspond to actual model calculations. Bi-cubic spline interpolation is used to calculate χ^2 at intermediate points. Four different plots are shown, labeled *a* to *d*. Plot *a* (upper left panel) shows the χ^2 plot of the fits, when using only the V and σ measurements of the WHT spectra as kinematic constraints. Clearly, a large range of parameter space gives equally good fits. There is at best only a marginal indication that models with a BH fit the ground-based rotation velocities and velocity dispersions better than without a BH. There is a clear correlation between M_{BH} and Y_I for the best fitting models, which is a reflection of the virial theorem; an increase of M_{BH} requires a decrease of Y_I . Plot *b* (upper right panel) is similar to plot *a*, except that we have now included the h_3 and h_4 measurements of the WHT spectra as constraints on the models. Two changes are evident. First, the allowed range of Y_I at given BH mass is much smaller. Secondly, BHs more massive than $\sim 6 \times 10^8 M_\odot$ can be ruled out at the 99.73 per cent confidence level. Plot *c* (lower left panel) shows the resulting χ^2 -plot when all constraints, including the HST measurements are taken into account. These high spatial-resolution measurements allow us to rule out models without a BH at a confidence level better than 99.73 per cent. Finally, in plot *d* (lower right panel) we show the results when using the modified approach on the entire set of constraints. As can be seen, the exclusion of counter-rotating orbits with circular radii beyond $4.5''$ puts some further limits on the allowed range of acceptable models, consequently contracting slightly the 99.73 per cent confidence region.

Although the solutions we find with the standard approach can result in models with significant counter-rotation (see Section 5.5), one can still meaningfully use the χ^2 statistics to put confidence levels on M_{BH} and Y_I . The main requirement is that orbital VPs and observed VPs are parameterized in exactly the same way, which is the case. The counter rotation that we find is simply a consequence of our particular VP parameterization. Although the modified approach results in a more strict solution space, and in principle is based on additional constraints that are observationally justified, we will nevertheless consider the χ^2 surface of Figure 5.7c as the main result. We merely present the results of the modified approach to show that the counter-rotation in the best-fitting models has no significant influence on the BH mass and mass-to-light ratio of our best-fitting model. Furthermore, Figure 5.7c results in more conservative estimates of the errors on BH mass and mass-to-light ratio.

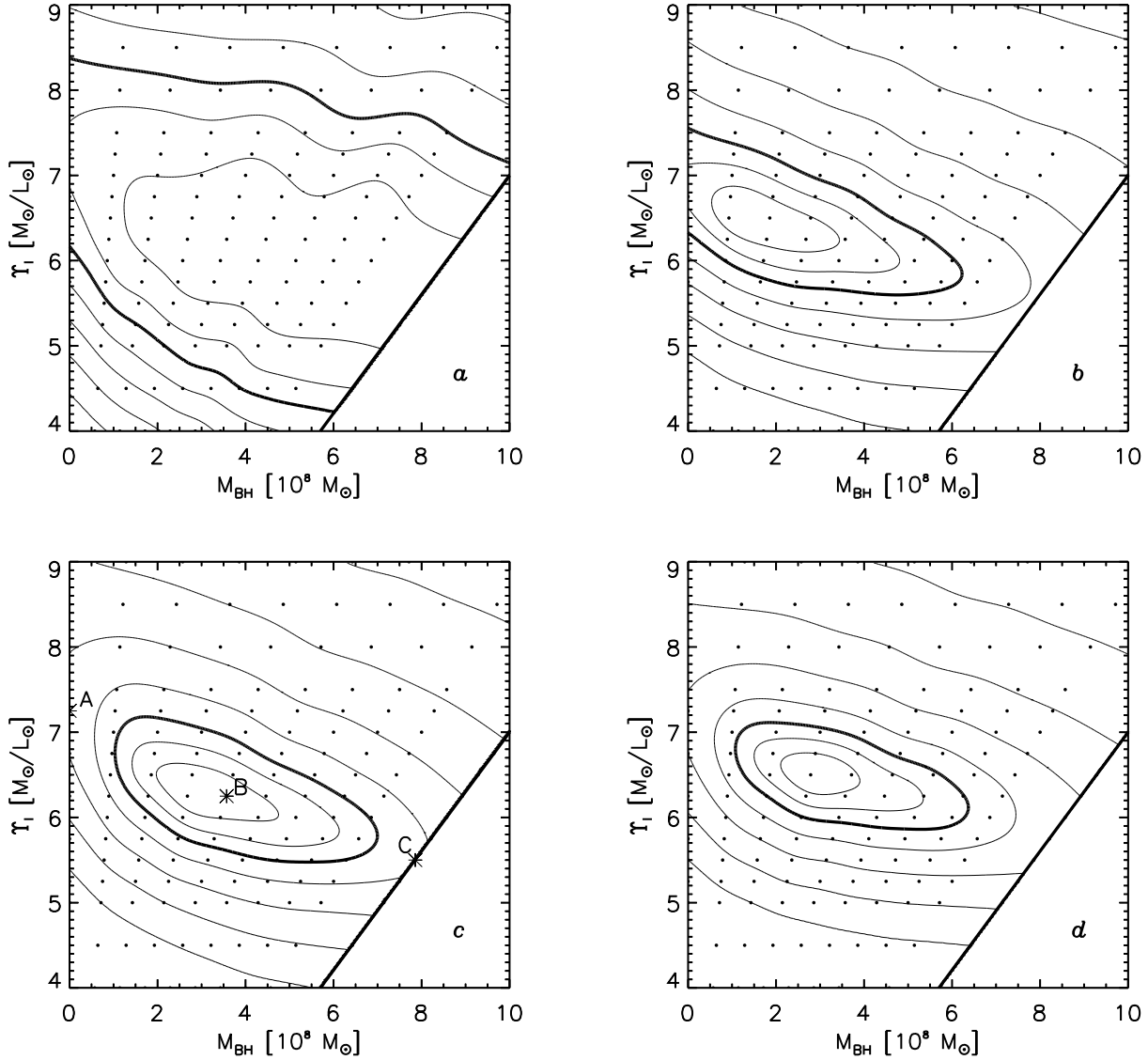


FIGURE 5.7— Contour plots of $\chi^2 = \chi_{\text{obs}}^2 + \chi_{\text{sc}}^2$, which measures the goodness-of-fit to the constraints as function of BH mass, M_{BH} , and I -band mass-to-light ratio, Y_I . The first three contours define the formal 68.3, 95.4 and 99.73 per cent confidence levels (latter one is plotted with a thick contour). The subsequent contours are characterized by a factor two increase in $\Delta\chi^2$. Solid dots indicate actual model calculations. Only ten different orbit libraries have been calculated, all with $Y = 1.0$. Models with different mass-to-light ratios are computed from these models by scaling as described in Section 5.6.1. This explains the inclined streams of points. No models are computed in the lower right corner of each panel. The χ^2 surface of panel *a* results from excluding h_3 and h_4 as well as all HST/FOS measurements from the constraints. In panel *b* we have only excluded the HST/FOS data, and panel *c* shows the resulting χ^2 plot when all constraints are taken into account. The asterisks labeled A, B and C indicate special models discussed in the text. Finally, in panel *d* we have used the modified approach and included the additional constraint that counter-rotating orbits with $R_c > 4.5''$ are not allowed in the solution.

The labeled asterisks in panel *c* indicate three models that will be discussed in more detail below. Model B provides the best overall fit and has $M_{\text{BH}} = 3.6 \times 10^8 M_\odot$ and $Y_I = 6.25 M_\odot/L_\odot$. From the χ^2 statistics we derive $M_{\text{BH}} = 3.0^{+1.7}_{-1.0} \times 10^8 M_\odot$ and $Y_I = 6.3^{+0.5}_{-0.4} M_\odot/L_\odot$. The errors are the formal 68.3 per cent confidence levels. For comparison, the modified approach yields $M_{\text{BH}} = 3.0^{+0.9}_{-0.8} \times 10^8 M_\odot$ and $Y_I = 6.5^{+0.3}_{-0.3} M_\odot/L_\odot$, which is consistent with and slightly more

strict than the results obtained with the standard method.

It is important to realize that the WHT data at the larger radii are essential for the limits we can put on the BH mass. If we exclude kinematic data outside $8''$, the model is less constrained and we find $M_{\text{BH}} = 4.7^{+1.5}_{-1.7} \times 10^8 M_{\odot}$ and $\Upsilon_I = 5.6^{+0.5}_{-0.3} M_{\odot} / L_{\odot}$. At 99.73 per cent confidence level, one can now only rule out BHs more massive than $\sim 8 \times 10^8 M_{\odot}$. Evidently, data at large radii helps in constraining the mass of the BH, despite the fact that the radius of influence of the BH is limited to $\lesssim 1''$. The reason for this counterintuitive aspect is twofold. First of all, the data at large radii are particularly useful in constraining the system's mass-to-light ratio. Because of the virial theorem, which imposes a correlation between M_{BH} and Υ , this helps in constraining the BH mass. Secondly, any orbit that contributes at large radii also contributes to the observed velocity profiles at the center because of the projection on the sky. This is especially important for an edge-on galaxy like NGC 4342. By including outer data points, one not only constrains the dynamics of the system at the outer parts, but in addition one limits the amount of freedom for the more central orbits (which contain information on the BH mass), because of the way the outer orbits contribute to the central VPs.

In chapter 4, we used the same three-integral modeling technique and statistical means to infer that M32 harbors a massive BH of $(3.4 \pm 0.7) \times 10^6 M_{\odot}$. To facilitate direct comparison, the χ^2 surfaces presented here are plotted with the same contour levels as in chapter 4. The χ^2 contours of Figure 5.7 correspond to the goodness-of-fit to *all* constraints. In chapter 4, we plotted contours of χ^2 of the fit to the kinematic constraints only (i.e., the right term in equation [5.17]), but we noted that contour plots of the total χ^2 look similar. Here, we have computed the surfaces of the χ^2 that only measures the fit to the kinematic constraints and found very similar contours: all models can fit the surface brightness and luminosity density to better than one per cent as requested.

In order to compare the kinematical predictions of the models with the actual observations, we compute the velocity profile of the model at each constraint position l , $\text{VP}_{l,v}$, as the weighted sum of all $\text{VP}_{l,v}^k$, using the solution γ for the orbit weights. For these VPs we then compute V_l and σ_l of the best-fitting Gaussian, as well as S_l and the GH moments $h_{3,l}$ and $h_{4,l}$, all of which can be directly compared to the observations. The kinematical predictions of models A, B and C are plotted in Figure 5.8 together with the observations. Although we did not measure the h_3 and h_4 coefficients for the FOS spectra, we plot the model predictions for these quantities. All three models provide equally good fits to the projected surface brightness and the meridional luminosity density (not plotted). The main difference between models A and B is their fit to the central HST velocity dispersion. The model without BH (model A) underpredicts the observed dispersion by $\sim 130 \text{ km s}^{-1}$ ($\sim 4.0\Delta\sigma$). The rapid central rotation in NGC 4342, as measured with the FOS is remarkably well fit even without central BH. The main noticeable difference between models B and C is their fit to the central WHT velocity dispersion along the minor axis, and the fit to the h_3 measurements along the major axis. Although model C fits the central HST velocity dispersion even better than model B, the fit to the higher-order GH coefficients is worse to such an extent that this model can be ruled out to better than 99.73 per cent confidence (cf. panels *a* and *b* of Figure 5.7).

A cautious reader may be surprised that Models A and C, plotted in Figure 5.8, can be ruled out with high confidence despite the fact that their fits to the actual data look almost indistinguishable from that of Model B. This apparent discrepancy arises because χ^2 -by-eye is not a good way to judge statistical significance. In total we use 122 kinematic constraints plus 134 constraints on the density distribution (both intrinsic and projected). With so many points, and with errorbars that are sometimes hardly visible in Figure 5.8 (e.g., for the major axis WHT rotation velocities), it becomes infeasible to assess significance levels by merely looking at the fits.

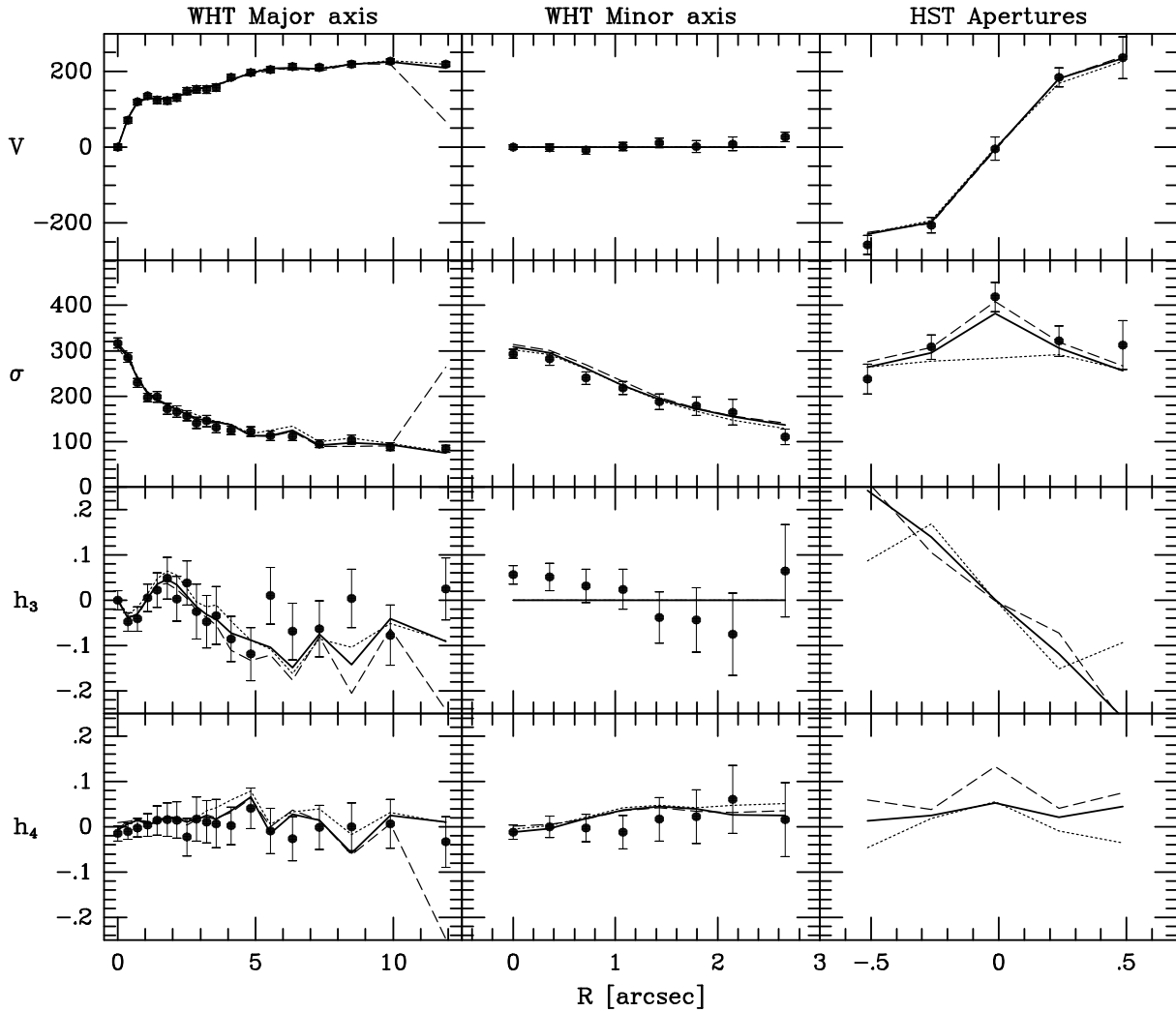


FIGURE 5.8— Kinematics for the three data sets (solid dots with errorbars) compared to the predictions for models A (dotted lines), B (solid lines) and C (dashed lines). Although we did not measure h_3 and h_4 from our HST/FOS spectra, we plot the model predictions in the lower two panels on the right. Model A, which has no BH, strongly underestimates the central velocity dispersion as measured with the FOS. The strong deviation of the predictions of model C for the outermost WHT point along the major axis, is due to the ‘artificial’ counter-rotation of the model (see Figure 5.10). As explained in the text, these deviations do not influence our χ^2 statistics (cf. panels *c* and *d* of Figure 5.7).

In order to exemplify this, we have calculated the cumulative χ^2 , as function of constraint number, for models A, B, and C. The results are shown in Figure 5.9. The constraints are organized as follows: constraints 1 to 19 correspond to the projected surface brightness S_I of the WHT data points along the major axis, constraints 20 to 27 to those of the WHT minor axis, and constraints 28 to 34 to those of the HST data points. The next 34 constraints correspond to $S_I h_{1,l}$, etcetera, up to $S_I h_{4,l}$. Constraints 171 to 270, finally, correspond to the self-consistency constraints, L_n , of the meridional density. Note that here we have included the undetermined $S_I h_{3,l}$ and $S_I h_{4,l}$ of the HST data, which add zero to the cumulative χ^2 . It is evident from Figure 5.9 that models A and C fit the data significantly worse than model B, something not immediately evident from Figure 5.8. A comparison of models B and C shows that there is not a particular point that dominates the $\Delta\chi^2$ between these two models, but rather that $\Delta\chi^2$ accumulates slowly over *all*

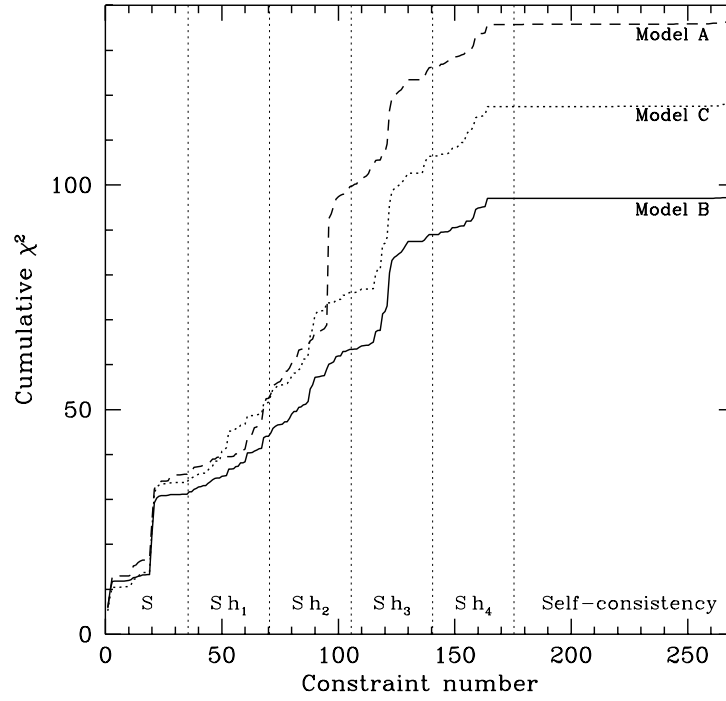


FIGURE 5.9— The cumulative χ^2 as function of constraint number for models A (dashed line), B (solid line), and C (dotted line). The constraints are ordered as follows: constraints 1 to 19 correspond to the projected surface brightness S_i of the WHT data points along the major axis, constraints 20 to 27 to those of the WHT minor axis, and constraints 28 to 34 to those of the HST data points. The next 34 constraints correspond to $S_i h_{1,i}$, etcetera, up to $S_i h_{4,i}$. Constraints 171 to 270, finally, correspond to the self-consistency constraints, L_n , of the meridional density. Note that none of the models accumulates a significant contribution to their total χ^2 from the self-consistency constraints; they all fit the meridional density to a high degree of accuracy. Furthermore, it is evident that models A and C accumulate a significant larger χ^2 than model B, and they do so slowly over *all* constraints. This explains why a simple χ^2 -by-eye assessment of Figure 5.8 seems to be inconsistent with the high levels of confidence we find for excluding models A and C in favor of model B.

points. The $\Delta\chi^2$ between models A and B is clearly dominated by a single point (the central HST velocity dispersion), but again the remainder of the total $\Delta\chi^2$ slowly accumulates over all points. Note finally that the self-consistency constraints do not add significantly to χ^2 for any of the models; all models fit the meridional density to a high degree of accuracy.

The large differences between the predictions of model C at the outermost point along the major axis (WHT measurements) and the actual observed values is due to the problem with the counter-rotation. This is illustrated in Figure 5.10, where we plot VPs at four different radii along the major axis. Solid dots represent the ‘observed’ velocity profiles, reconstructed from the measured GH-parameters (assuming that all GH coefficients of order five and higher are zero). The solid lines represent the reconstructed VPs of model C, and the dashed lines correspond to the Gauss-Hermite series fitted to these model VPs. The upper four panels show the VPs of model C. The lower panels show the VPs for a model with the same BH mass and mass-to-light ratio as model C, but for which the modified approach is used. For $R = 0.0''$ and $R = 4.83''$ the model VPs of the two different approaches are almost identical. However, for the larger radii, the VPs of model C clearly reveal large amounts of counter rotation. The VP of the upper right panel has such a large counter-rotating component, that the best-fitting Gaussian no longer corresponds to the peak at positive velocities: it is very broad and centered around $V \approx 0 \text{ km s}^{-1}$. This causes the large difference between the model predictions and the

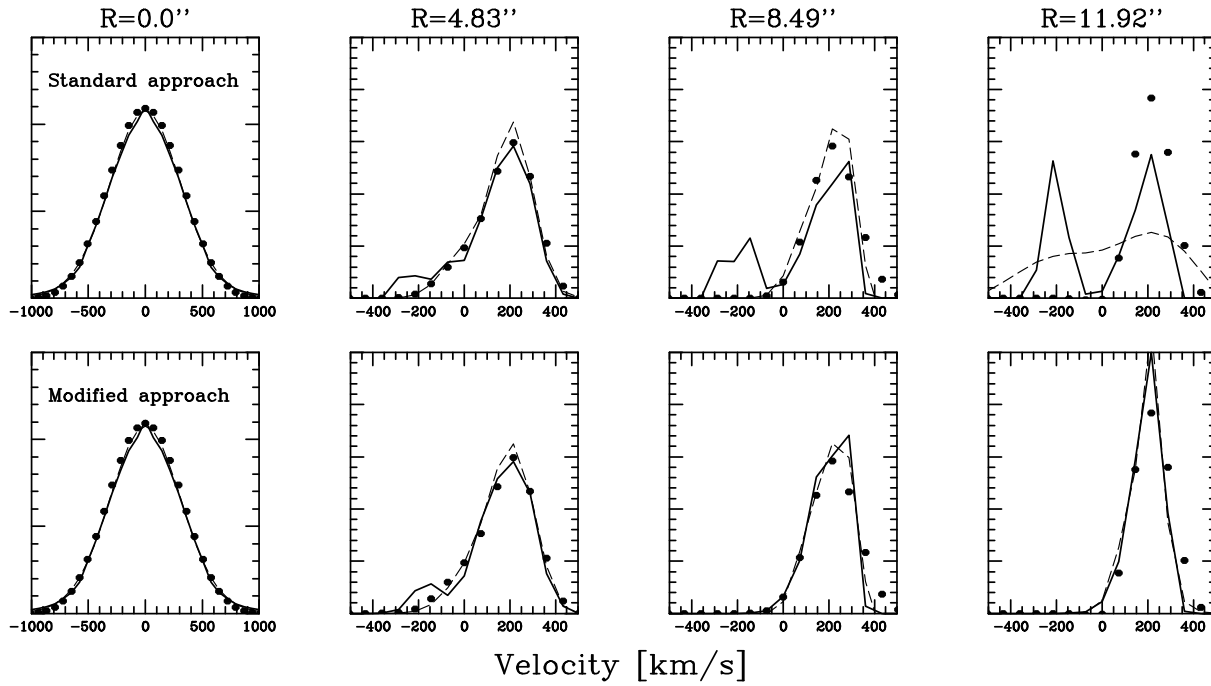


FIGURE 5.10— Velocity profiles at four different radii along the major axis of NGC 4342. The observed VPs, reconstructed from the measured V , σ , h_3 and h_4 while assuming $h_m = 0$ for $m \geq 5$, are plotted as solid dots. The solid lines correspond to the model VPs, and the dashed lines to the Gauss-Hermite series (up to order 4) fitted to these model VPs. Upper panels correspond to model C, whereas lower panels correspond to a model with the same BH mass and mass-to-light ratio, but whose orbit solution was computed with the modified approach. The upper panels clearly reveal that the model VPs at larger radii are double peaked. This can yield strange Gaussian fits, as evident from the upper right panel. The modified approach excludes counter-rotating orbits with $R_c > 4.5''$ from the orbit library, and does not reveal these double peaked VPs.

observed values at the outermost WHT point in Figure 5.8. Although the reconstructed V and σ may differ strongly from the observed values, the h_1 and h_2 values are very similar. Since these are the values that enter into the computation of χ^2 , this discrepancy does not affect the χ^2 statistics.

It is interesting that the central velocity dispersions, as measured with the WHT, can be fit without the requirement of a central BH. This contradicts the conclusions reached from the Jeans modeling, which suggests, on the basis of the WHT measurements alone, that a BH of $\sim 3 \times 10^8 M_\odot$ is required. The three-integral modeling, however, shows that only the HST measurements are of sufficient spatial resolution to discriminate strongly between models with and without a massive BH (cf. panels *b* and *c* of Figure 5.7). Clearly, without these high spatial resolution kinematics, the case for a BH is only marginal: the χ^2 plot in panel *b* of Figure 5.7 suggests a BH mass of $\sim 2.0 \times 10^8 M_\odot$, but cannot rule out models without a BH to a significant confidence level.

5.6.2 Alternatives to a black hole

A dense cluster

Although the dynamical evidence for the presence of a MDO of a few $10^8 M_\odot$ in NGC 4342 is compelling, it does not automatically imply evidence for a BH. Alternatives to a point mass, such as a cluster of brown dwarfs or stellar remnants are not ruled out by the modeling presented above. Any of these alternatives is only viable if its lifetime is not significantly smaller

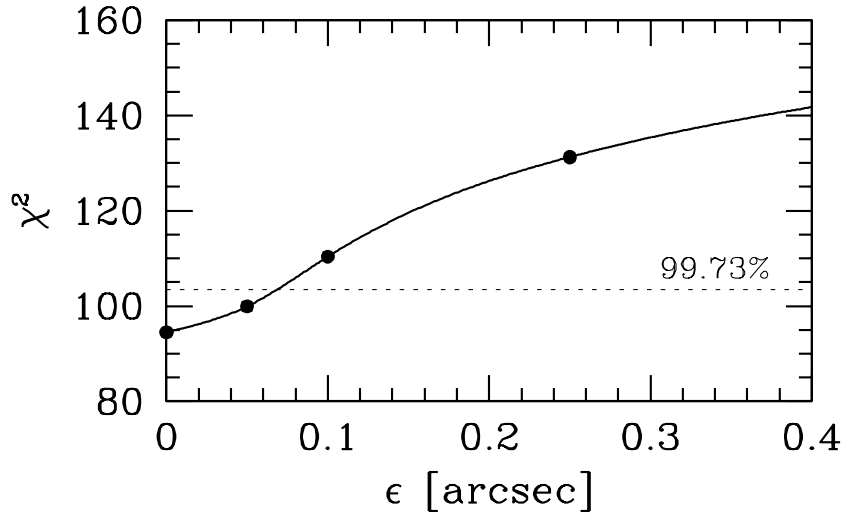


FIGURE 5.11— The value of $\chi^2 = \chi_{\text{obs}}^2 + \chi_{\text{sc}}^2$, which measures the goodness-of-fit to the constraints, as function of the scale length ϵ (in arcsec) of the Plummer potential with a total mass of $3.6 \times 10^8 M_\odot$ representing a cluster of dark objects in the center of NGC 4342. All models have a stellar mass-to-light ratio of 6.25. The best fit to the data is achieved in the limit $\epsilon \rightarrow 0$, which corresponds to our model with a BH rather than a dark cluster. We can rule out models with $\epsilon > 0.07''$ at the 99.73 per cent confidence level.

than the age of the galaxy (typically ~ 10 Gyr). There are three main processes that determine the evolution of a dense cluster: (i) core collapse, (ii) evaporation due to weak gravitational scattering, and (iii) physical collisions between the objects comprising the cluster. The latter of these processes is likely to ultimately lead to the formation of a single dense object, probably a massive BH. After a timescale τ_{coll} , each object has physically collided with another (in a statistical sense). τ_{coll} strongly depends on the mass and density of the cluster, as well as on the mass and size of the constituents (see Maoz 1997 and references therein). The characteristic timescale for evaporation of a cluster which consists of equal mass constituents is $\tau_{\text{evap}} \approx 300 \tau_{\text{relax}}$ (Spitzer & Thuan 1972), whereas typically after $\tau_{\text{cc}} \approx 16 \tau_{\text{relax}}$, a Plummer sphere of equal mass objects undergoes core collapse (Cohn 1980). Here τ_{relax} is the median relaxation time (see Spitzer & Hart 1971). By choosing a Plummer model for the dark cluster we are conservative, in that more concentrated clusters have shorter collapse times, and are thus less likely as alternatives for a massive BH.

In order to constrain the size of a dark cluster, we construct models in which we replace the point-mass potential of the BH by a Plummer potential with a scale length ϵ . We consider model B for which we replace the $3.6 \times 10^8 M_\odot$ BH by a Plummer potential with the same mass, but with different values of ϵ . The stellar mass-to-light ratio is kept constant at 6.25. Figure 5.11 shows the resulting χ^2 as function of ϵ . The best fit is obtained for $\epsilon = 0.0$ (model B), and the fit deteriorates with increasing scalelength of the Plummer potential. The dotted line indicates the formal 99.73 per cent confidence level, and at this level of confidence we can rule out dark clusters with $\epsilon > 0.07''$. This upper limit on the scale length of the Plummer sphere corresponds to 5.1 pc at the assumed distance of 15 Mpc, implying a central density of the cluster $> 6.7 \times 10^5 M_\odot \text{pc}^{-3}$.

We calculate the relaxation timescale τ_{relax} , and the collision timescale τ_{coll} in which we adopt the mass-radius relation for *non-collapsed* objects used by Goodman & Lee (1989). In Figure 5.12 we plot the characteristic timescales for core-collapse and collisional destruction of our dark cluster, as function of the mass m of the cluster's constituents. As can be seen, the timescales for

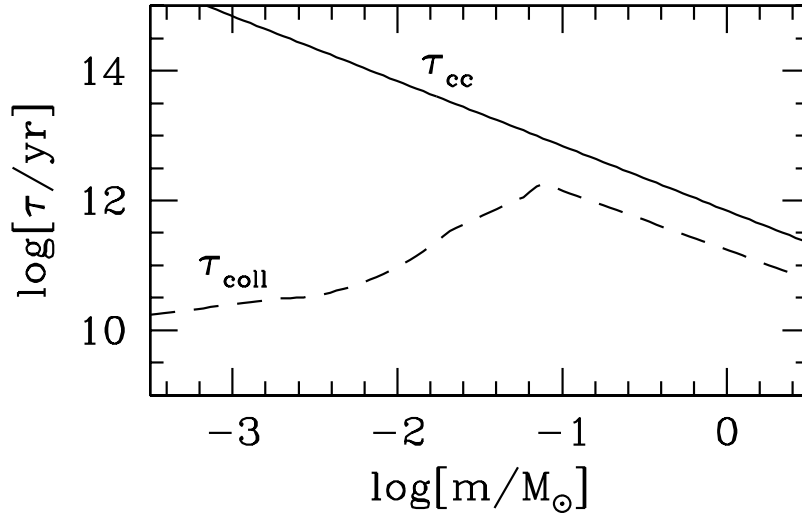


FIGURE 5.12— The timescales for core collapse (τ_{cc}) and collisional destruction (τ_{coll}) of a $3.6 \times 10^8 M_\odot$ Plummer cluster of non-collapsed objects with a central density of $6.7 \times 10^5 M_\odot \text{pc}^{-3}$ as function of the mass m of the constituents. If any of these timescales is less than $\sim 10^{10}$ years, it makes such cluster an unlikely alternative to a BH. Unfortunately, current data does not allow us to rule out any of the dark clusters as an alternative MDO in NGC 4342.

core collapse ($\tau_{cc} \approx 7 \times 10^{11} (m/M_\odot)^{-1}$ yr) and evaporation ($\tau_{evap} \sim 20\tau_{cc}$) for the inferred dark cluster in NGC 4342 exceed the Hubble time for $m \lesssim \text{few } M_\odot$, and neither of these processes thus allows us to rule out a dark cluster as an alternative to a BH. The collision timescale τ_{coll} is the most restricting, and we are close to being able to rule out dark clusters of non-collapsed objects with masses less than $\sim 0.001 M_\odot$. However, for clusters of brown dwarfs with masses of $\sim 0.08 M_\odot$, τ_{coll} is still of the order of 10^{12} yr, and such clusters can clearly not be ruled out by current observations. Clusters of *collapsed* stellar remnants, such as white dwarfs or neutron stars, have collision timescales that are even much larger than the value of τ_{coll} plotted in Figure 5.12. This is due to the much smaller collisional cross-sections of these collapsed objects. In fact, the collapse time for a dark cluster can be made arbitrarily long by giving its objects an arbitrarily small mass, and the problems with collisions and mergers can be avoided by assuming the cluster to be a collisionless gas of elementary particles. However, another effect might allow one to rule out such a dark cluster: the trapping of stars by the cluster due to dynamical friction. If the time scale for this process is smaller than the age of the galaxy, enough stars will get trapped such that the cluster can no longer be considered dark. Alternatively, these trapped stars can merge and form a massive BH, thus spoiling the goal for which the cluster was introduced. The time scale for the trapping to occur is independent of the mass m of the cluster constituents, but does depend on the mass m^* of the stars being captured, and is given by $(m/m^*)\tau_{cc}$ (as long as $m^* \gg m$, see Quinlan 1996). For NGC 4342 we thus find that this capture time scale is only smaller than $\sim 10^{10}$ yr for stars with $m^* \gtrsim 70 M_\odot$. We can thus not use this trapping mechanism to put an appreciable constraint on the nature of a possible dark cluster in NGC 4342.

In conclusion, it is clear that we are still a long way from being able to rule out dark clusters as viable alternatives for a massive BH in the center of NGC 4342.

An end-on bar

It has been argued that a bar observed end-on may mimic the presence of a BH (Gerhard 1988). In particular, the nuclear disk in NGC 4342 may in fact be a (very thin) nuclear bar. However,

there are several reasons why this interpretation is unlikely. First of all, axisymmetric models without BH fail to fit the central velocity dispersion, as measured with the HST/FOS, by $\sim 130 \text{ km s}^{-1}$. It seems unlikely that the elongated orbits in a bar potential can fix this. Furthermore, there are no indications that (the center of) NGC 4342 is triaxial. Although van den Bosch & Emsellem (1998) have provided evidence that NGC 4570, a galaxy with a double disk morphology similar to NGC 4342, has been shaped under the influence of a rapidly tumbling bar potential, none of the characteristics found in NGC 4570 that led to this conclusion are apparent in NGC 4342. In addition, no minor axis rotation is found, and the small amount of isophote twist observed (see BJM98) is limited to the outer region of the galaxy, and is more likely to be associated with a small warping of the outer disk, probably induced by the small companion at $\sim 30''$ SE. In the end-on bar hypothesis, there has to be a preferred orientation of NGC 4342 with respect to our line-of-sight, something clearly not in favor of this model. Finally, the steep cusp of the bulge of NGC 4342 makes the bar hypothesis unlikely, since the pressure support from this strongly cusped bulge probably assures stability against bar formation; the steep rotation curve implies the presence of an inner Lindblad resonance, which is effective in inhibiting the formation of a bar (e.g., Toomre 1981; Sellwood 1989; van den Bosch & de Zeeuw 1996).

Although none of these arguments by themselves are very convincing, and we can thus not exclude this possibility with a high degree of confidence, we consider an end-on nuclear bar in NGC 4342 unlikely given the various arguments listed above.

5.6.3 The dynamical structure of NGC 4342

During the orbit integrations we store, in addition to the fractional time spent by orbit k in meridional cell n , t_n^k , the time-weighted first and second order velocity moments averaged over cell n . With the solution for the orbit weights γ we can compute the internal dynamical structure of the model averaged over each cell n . These are given by

$$\langle v_a \rangle_n = \frac{\sum_k \gamma_k t_n^k \langle v_a \rangle_n^k}{\sum_k \gamma_k t_n^k}, \quad (5.19)$$

and

$$\langle v_a^2 \rangle_n = \frac{\sum_k \gamma_k t_n^k \langle v_a^2 \rangle_n^k}{\sum_k \gamma_k t_n^k}. \quad (5.20)$$

Here we use the subscript a to indicate either of the cylindrical coordinates R , ϕ or z . The cell-averaged velocity dispersions can be computed according to $\langle \sigma_a \rangle_n^2 = \langle v_a^2 \rangle_n - \langle v_a \rangle_n^2$.

In order to suppress noise, we average $\langle v_a \rangle_n$ and $\langle v_a^2 \rangle_n$ over cells n that have the same radius r but different θ (r and θ being the standard spherical coordinates). Given the strongly flattened shape of NGC 4342 and its multi-component structure, we decided to split the galaxy in two parts: we determine the internal dynamical structure in two cones; one with half-opening angle of 30° centered around the equatorial plane, including the nuclear and outer disks, and the other one with half-opening angle of 60° centered around the minor axis, representing the bulge of NGC 4342.

We investigate the dynamical structure of models A, B, and C using the modified approach (see Section 5.5) and regularization as described in Section 5.4.5 in order to suppress noise. As a check, we compare the dynamical structures before and after regularization (with the amount of smoothing chosen such that $\Delta\chi^2 = 1$). The regularized model is found to have a smoother dynamical structure, but the main features are similar for both cases indicating that our results are not too sensitive to the particular method and amount of regularization adopted.

In Figure 5.13 we plot the dynamics in the ‘equatorial cone’ as function of radius. The upper panels show the rms velocities for models A, B and C in the range between $r = 0.1''$ and $r = 12''$

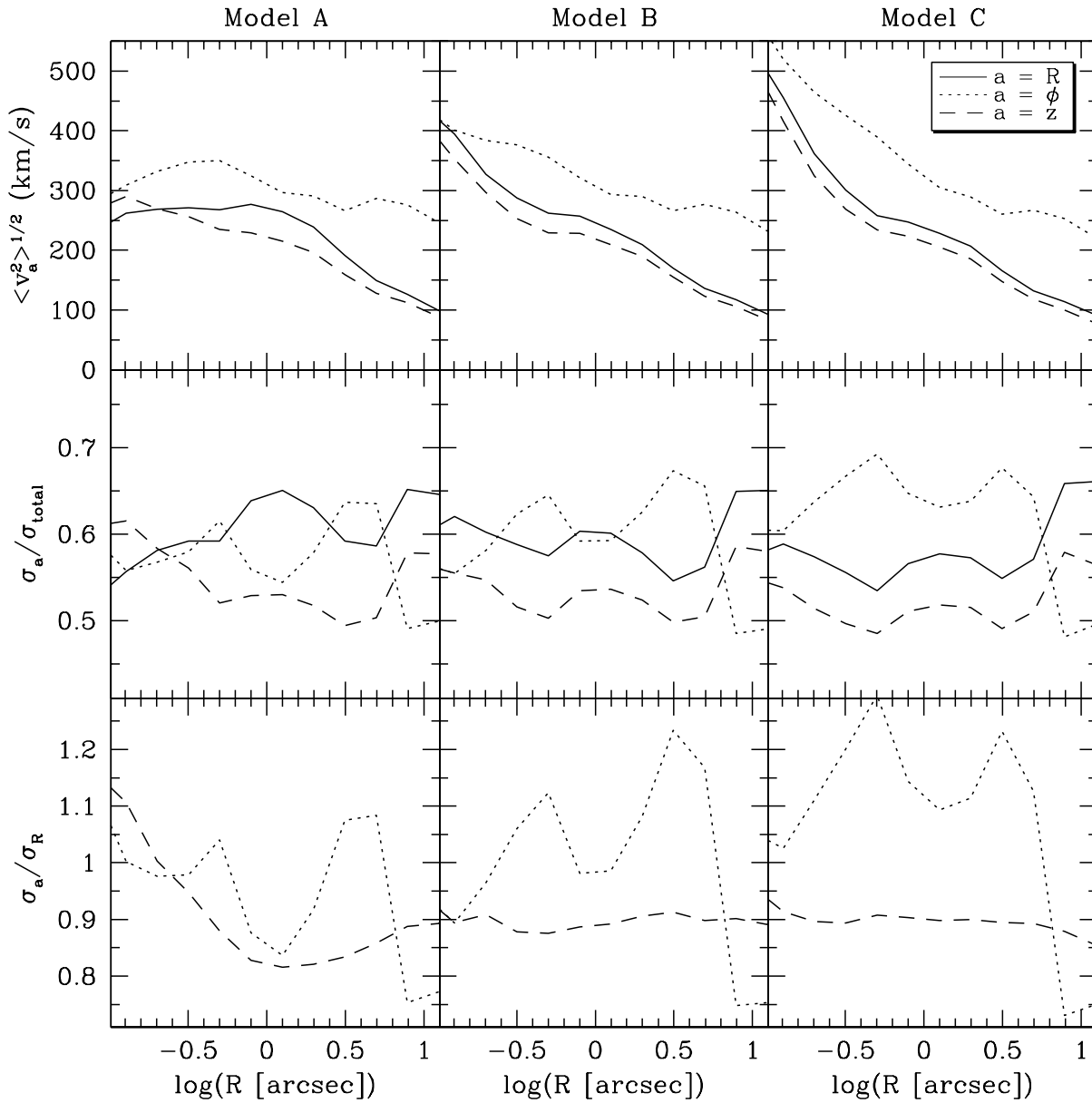


FIGURE 5.13— Dynamical structure of models A, B and C averaged over a cone with half-opening angle of 30° centered on the equatorial plane. Upper panels show the rms velocities $\langle v_a^2 \rangle^{1/2}$ in km s^{-1} , middle panel the normalized velocity dispersions $\sigma_a / \sigma_{\text{total}}$, and lower panels the ratios σ_a / σ_R . Solid curves are for the radial component ($a = R$), dotted curves for the azimuthal component ($a = \phi$), and dashed curves for the vertical component ($a = z$). Results are plotted over the radial interval where we have kinematic constraints along the major axis, i.e., $0.1'' < R < 12''$.

(corresponding to the regime where we have kinematic constraints along the major axis). The middle panels display the ratio $\sigma_a / \sigma_{\text{total}}$ for the same radial interval ($\sigma_{\text{total}}^2 = \sum_a \sigma_a^2$). Since there is no streaming motion in the radial and vertical directions, $\langle \sigma_R \rangle$ and $\langle \sigma_z \rangle$ are equal to their respective rms velocities. The difference between $\langle v_\phi^2 \rangle$ and σ_ϕ reflects the streaming motion of the model. The lower panels show the ratios σ_z / σ_R and σ_ϕ / σ_R . The three models differ predominantly in the inner $\sim 3.0''$, reflecting the differences in BH mass, but are very similar outside of $3.0''$: clearly $\langle v_\phi^2 \rangle$ dominates the dynamics outside this radius in accordance with the

rapid rotation of the outer disk. The rapid change in dynamical structure (from azimuthally anisotropic to radially anisotropic) going from $3''$ to $12''$ reflects the strong increase of $(V/\sigma)_{\text{obs}}$ over this radial interval (see Figure 5.6): the dynamically cold outer disk, mainly built up of close-to-circular orbits with low σ_ϕ , becomes the dominant mass component. Over the same radial interval, the projected ellipticity increases from ~ 0.4 to ~ 0.7 , and this thus explains the observed correlation between ellipticity and $V_{\text{obs}}/V_{\text{mod}}$ of the best fitting isotropic Jeans model plotted in Figure 5.4. At $R \gtrsim 8''$, $\sigma_\phi/\sigma_R \sim 0.75$, not too different from the value in the solar neighborhood where $\sigma_\phi/\sigma_R = 0.6$ (Dehnen 1998).

Models B and C have σ_z/σ_R remarkably constant at ~ 0.9 , and are thus not too different from two-integral models (for which this ratio is exactly 1.0). This is very different from the case without BH (model A) for which σ_z/σ_R is a strong function of radius R . This is the reason that the two-integral, isotropic Jeans model discussed in Section 5.3, could not fit the HST rotation velocities without BH, whereas model A can.

The main change going from model A to model C, is a strong increase of σ_ϕ/σ_R in the inner $\sim 3''$. In these inner regions the circular velocities increase strongly with increasing BH mass. Nevertheless, all three models provide an almost equally good fit to the observed rotation velocities. We checked our solutions and found that going from model A to C, the ratio of $-L_z$ -orbits over $+L_z$ -orbits increases in the radial interval $0.5'' < R < 3.0''$. This causes the net streaming motions of all three models to be roughly similar despite the large differences in circular velocities. Furthermore, it explains the strong increase in σ_ϕ observed when going from model A to model C.

The increase of azimuthal anisotropy (at around one arcsec) with increasing BH mass is, at least qualitatively, in agreement with predictions for the velocity anisotropy resulting from the adiabatic growth of a BH (Goodman & Binney 1984; Quinlan, Hernquist & Sigurdsson 1995; Sigurdsson, Hernquist & Quinlan 1995; Merritt & Quinlan 1998). In all of these studies, the mass distribution was taken to be spherical, or, in the case of Merritt & Quinlan, became spherical due to the BH growth. Because of the presence of the rapidly rotating nuclear disk in NGC 4342, it however remains unclear whether a comparison of our results with those of the adiabatic BH growth models has any direct relevance, i.e., the central region of NGC 4342 is far from spherical.

The dynamics of the bulge (represented by the cone with half-opening angle of 60° centered around the minor axis) is shown in Figure 5.14. We only plot the results out to $R = 3.0''$, corresponding to the radial interval where we have kinematic constraints along the minor axis. At $R \gtrsim 1''$ the models are again similar, being dominated by rotation, albeit to a much lesser extent than along the major axis. Model A is radially anisotropic in the central region ($R \lesssim 1''$), and σ_R decreases steadily with BH mass going from model A to model C. The radial anisotropy of the central region of the bulge of model A is required to fit the high central velocity dispersions observed as good as possible. Close to the equatorial plane, the steep rotation curve observed prevents model A from being too radially anisotropic (see also Section 5.6.4). Model B is again remarkably close to two-integral form, having $\sigma_z/\sigma_R \sim 1.0$.

The fact that our best fitting model has a DF that is close to two-integral form is particularly intriguing in the light of some recent results. As pointed out by Merritt (1998, and references therein), the presence of a massive BH has a smoothing effect on the orbital population of a triaxial galaxy: the box orbits that support the triaxiality are strongly influenced by a central BH, since stars on box orbits pass arbitrarily close to the center after a long enough time. Chaotic mixing (Kandrup & Mahon 1994; Merritt & Valluri 1996) occurs, which smoothes out the phase-space density. The smoothest DF for an axisymmetric system is the one that is constant with respect to the third integral, i.e., $f = f(E, L_z)$. In chapter 4, we used the same three-integral

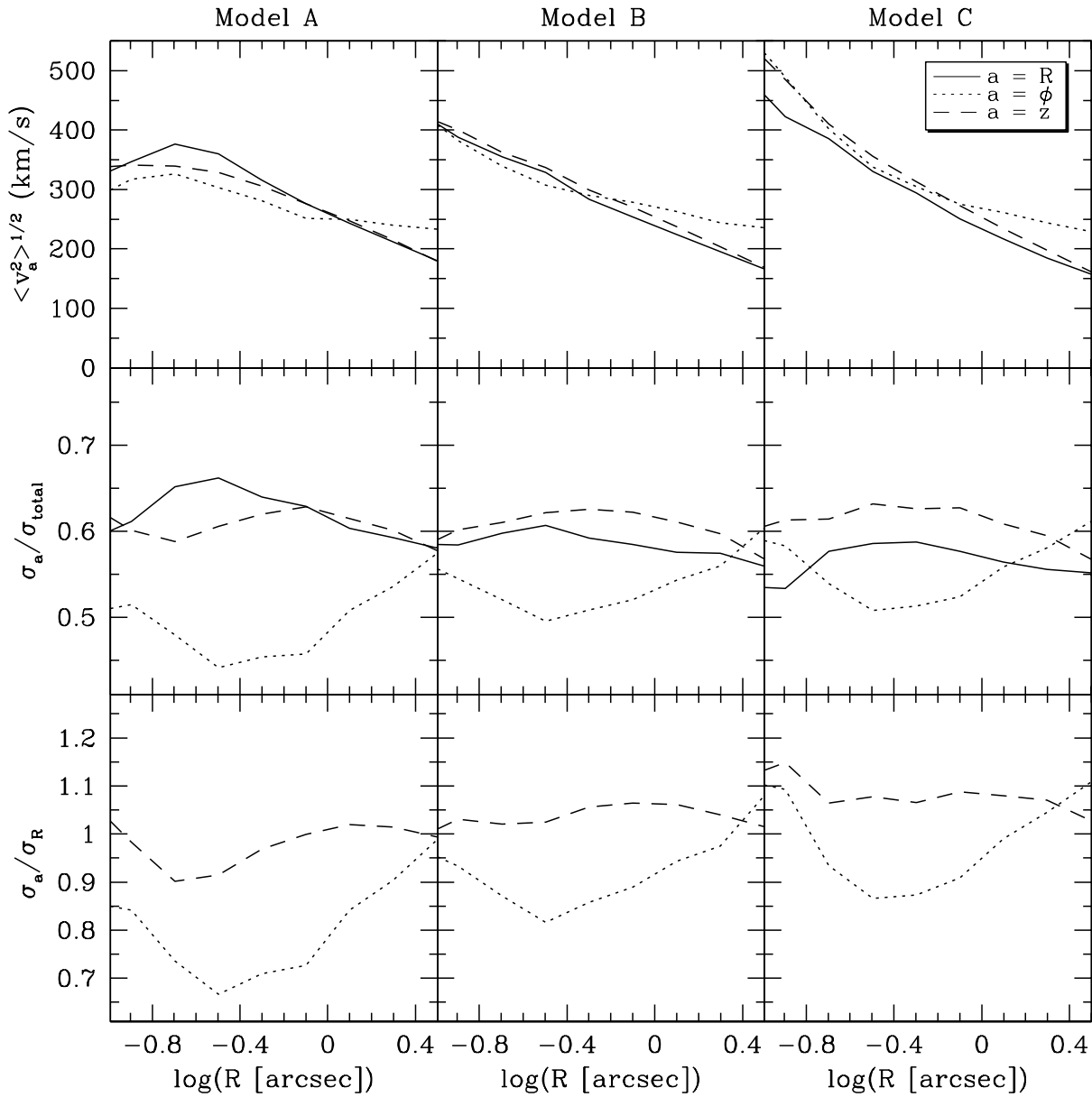


FIGURE 5.14— Same as Figure 5.13, except that here the quantities are averaged over a cone with half-opening angle of 60° centered on the symmetry axis $R = 0$ of the models.

modeling technique as employed here and found the best fitting model for M32 to not only harbor a massive BH, but also to have a DF that is remarkably close to two-integral form, i.e., similar to the results for NGC 4342 presented here. These results are thus consistent with a scenario in which M32 and NGC 4342 have evolved towards axisymmetric structures of two-integral form because of chaotic mixing induced by a massive BH.

5.6.4 The influence of central radial anisotropy

We now investigate to what extent radial anisotropy can influence the central kinematics of NGC 4342. We solve the orbit weights for the model with $\Upsilon_I = 7.25 M_\odot / L_\odot$ and $M_{\text{BH}} = 0$ using two different sets of constraints. The first one uses all constraints except the *rotation velocities*

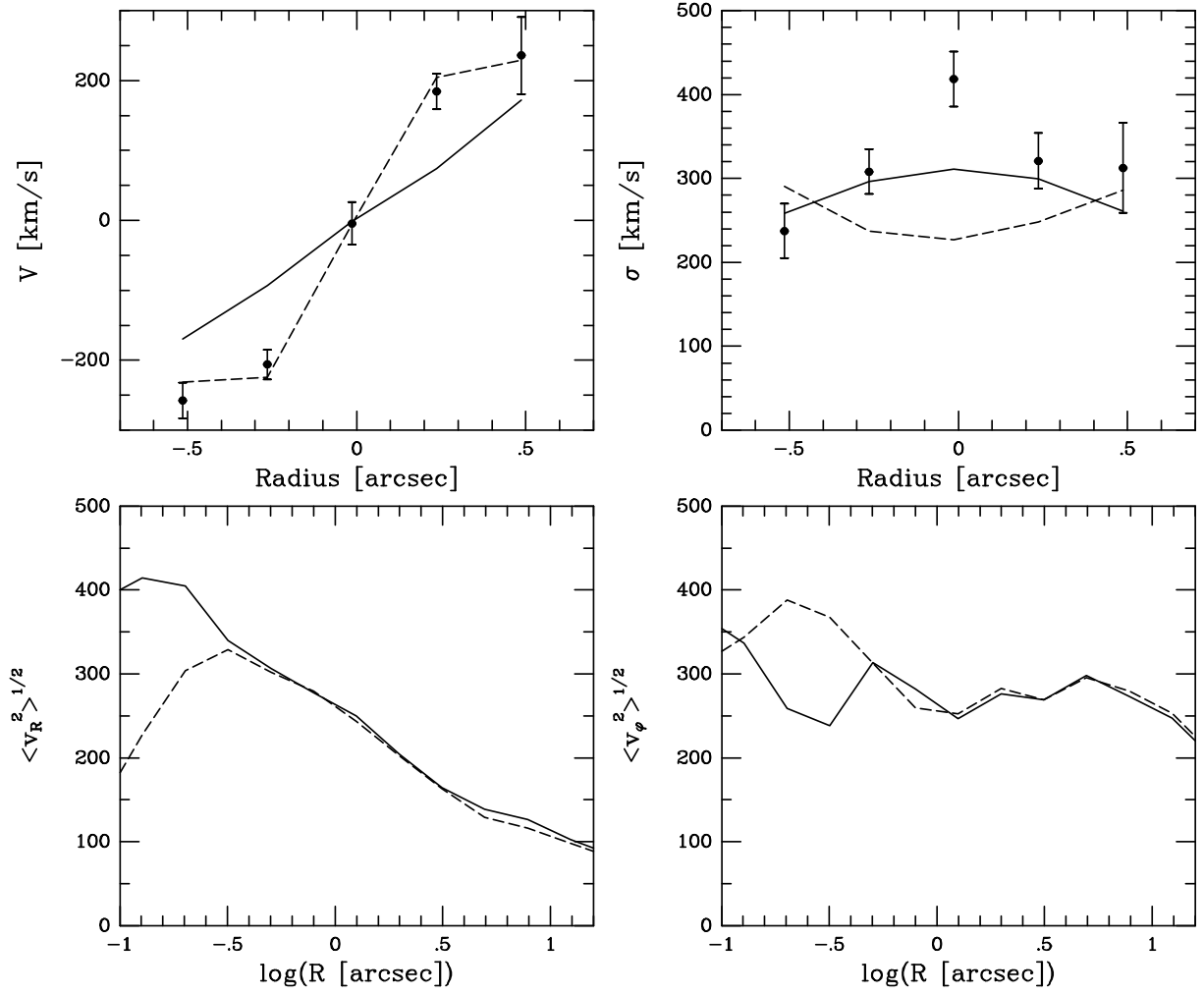


FIGURE 5.15— Radial velocity anisotropy at work: Solid lines are for a model with $Y_I = 7.25 M_\odot / L_\odot$ and $M_{\text{BH}} = 0$, where we have neglected the FOS measurements of the rotation velocities. The dashed lines correspond to the same model, but for which we have neglected the FOS velocity dispersion measurements. The two lower panels show the intrinsic dynamics, $\langle v_R^2 \rangle^{1/2}$ (lower left panel) and $\langle v_\phi^2 \rangle^{1/2}$ (lower right panel), of the two models.

measured with the FOS. The second one excludes only the FOS *velocity dispersions* from the constraints. The results for both models are shown in Figure 5.15, where we plot the predictions for the rotation velocities and velocity dispersions of both models. We only plot the predictions for the FOS kinematics: both models have almost indistinguishable WHT kinematics. The lower two panels in Figure 5.15 show the dynamics of the resulting models; we plot both $\langle v_R^2 \rangle^{1/2}$ and $\langle v_\phi^2 \rangle^{1/2}$ as function of radius. These quantities are averaged between $\theta = 0^\circ$ and $\theta = 90^\circ$.

The model with the FOS rotation velocities excluded from the constraints (solid lines) reveals the highest central velocity dispersion achievable without a BH, while still fitting the WHT measurements. The model predicts a central rotation curve that is too shallow to fit the observed rotation velocities, and is strongly radially anisotropic in the center. The model excluding the FOS velocity dispersions (dashed lines) shows a very steep central rotation curve. The model is strongly azimuthally anisotropic in the center, resulting in a much smaller central velocity dispersion. Clearly, a model without a central BH cannot simultaneously fit the rotation velocities and velocity dispersions measured with the FOS.

5.6.5 The density distribution in the outer region of NGC 4342

The best-fitting mass-to-light ratio of $Y_I = 6.3M_\odot/L_\odot$ is unusually large for the stellar population of an early-type galaxy (e.g., van der Marel 1991). The mass-to-light ratio is mainly set by the large rotation velocities measured at the outside of NGC 4342 (along the major axis). The large value found may indicate that NGC 4342 is embedded in a massive dark halo. We did not attempt to include such a component for two reasons: First, we are mainly interested in trying to fit the central kinematics and to examine the mass of a possible BH. The characteristics of a possible dark halo will not affect the central dynamics, and therefore will not alter our main conclusion that NGC 4342 harbors a massive BH. It may however result in a different dynamical structure of the outer bulge and disk. Secondly, since we can fit all kinematics without having to infer a dark halo, including such a component in the models will merely lead to a large range of different models that can fit the observed kinematics equally well. Only kinematics at much larger radii can constrain the presence and characteristics of such a possible dark halo (cf. chapter 2).

Since mass-to-light ratio scales with distance d as d^{-1} , another explanation for the large value of Y_I derived might be an underestimate of the distance. In order to bring the inferred mass-to-light ratio in accordance with the average value for early-type galaxies, NGC 4342 has to be at about twice the assumed distance of 15 Mpc. This is however hard to reconcile with the observed heliocentric velocity of NGC 4342 of only 714 km s^{-1} (de Vaucouleurs *et al.* 1976). Galaxies out to 15 Mpc from the center of the Virgo cluster can still experience a Virgocentric infall of a few hundred km s^{-1} : our local group falls towards Virgo with $\sim 200 \text{ km s}^{-1}$ (Tammann & Sandage 1985). A distance of 30 Mpc for NGC 4342, however, would imply a Virgocentric infall velocity of $\sim 1500 \text{ km s}^{-1}$ (where we have taken Virgo to be at 1100 km s^{-1}), which seems too large. So, the small recession velocity of NGC 4342 implies that it can not be too far behind Virgo such as to have an appreciable effect on the inferred mass-to-light ratio.

Another problem related to the outer density distribution is that the mass model used (the MGE model) falls off as $\exp(-r^2)$ at large radii. This is probably not correct, but since we only have photometry that is limited to a small radial extent, we cannot examine this in more detail. The main requirement is that we have a proper mass model for the major part of the galaxy. Our MGE model has a total luminosity of $L_I = 3.57 \times 10^9 L_\odot$. We have shown in Section 5.2, that for this luminosity we derive $B - V = 1.09$, in good agreement with the average value for early-type galaxies. This therefore suggests that our MGE model covers the major part of the galaxy, and that the galaxy does not extend far beyond the radial extent of our MGE model. In addition, as is the case with a dark halo, a change in the luminosity density profile in the outer parts of NGC 4342 will not affect our main conclusions regarding the presence of a central BH.

5.6.6 Comparison with other BH detections

There are now about a dozen BH candidates known. The best cases are the center of the MW, where a proper motion study has revealed a $2.5 \times 10^6 M_\odot$ BH (Eckart & Genzel 1997), and NGC 4258, where water masers are found to be in Keplerian rotation around a $3.6 \times 10^7 M_\odot$ BH (Miyoshi *et al.* 1995). Other strong evidence for the existence of massive BHs comes from the observation of the Fe K α line at 6.4 keV in a number of Seyfert 1 nuclei: the X-ray emission line exhibits relativistic motions, interpreted as arising from the accretion disk surrounding a massive BH (Tanaka *et al.* 1995; Nandra *et al.* 1997).

Most other BH detections (or confirmations) are based on observations with the HST. Observations of nuclear gas disks with the HST has provided strong cases for massive BHs in M84, M87, NGC 4261, NGC 6251, and NGC 7052. All other cases are based on stellar dynamical

evidence: M31, M32, NGC 3115, NGC 3377, NGC 3379, and NGC4486B. Of these only M 32, NGC 3379 and NGC 4342 discussed here have been confronted with three–integral modeling thus far. As discussed in the introduction, one can only hope to properly detect a BH if one can observe its kinematics inside r_{inf} . All currently detected BHs have $r_{\text{inf}} > 0.1''$, and since all these galaxies, except the nucleus of the Milky Way, have now been observed with the HST, all current BH–candidates fulfill this requirement. Remarkable enough, most of these galaxies (M31, M32, M87, NGC 3115, and NGC 4594) were already considered BH candidates when the angular resolution of the observations was still an order of magnitude larger, and similar to the angular size of the radius of influence of the inferred BH (Rix 1993). The BH in NGC 4342 has a relatively small radius of influence, with an angular size of $\sim 0.4''$. Indeed we have shown that only the HST/FOS observations are of sufficient spatial resolution to rule out models without a central BH.

Kormendy & Richstone (1995) have suggested a relation between the mass of the BH and that of the bulge of its parent galaxy: $\langle M_{\text{BH}}/M_{\text{bulge}} \rangle = 0.22^{+0.14}_{-0.09} \times 10^{-2}$. For more detailed discussions on this relation, including the results for NGC 4342 presented here as well as other recent BH detections see van der Marel (1997), Kormendy *et al.* (1998), Ho (1998), Merritt (1998), Richstone (1998), and van der Marel & van den Bosch (1998). NGC 4342 is the galaxy with currently the second–highest ratio of BH mass over bulge mass of $2.6^{+1.5}_{-0.9} \times 10^{-2}$, and is only superseded by the peculiar galaxy NGC 4486B, for which the case of a massive BH is less strong (Kormendy *et al.* 1997). NGC 3115 ($M_{\text{BH}}/M_{\text{bulge}} = 2.4 \times 10^{-2}$) and the Milky Way ($M_{\text{BH}}/M_{\text{bulge}} = 0.017 \times 10^{-2}$) are other clear deviants. Clearly, the scatter around the suggested relation is considerable, seemingly as large as nearly two orders of magnitude.

5.7 Conclusions

Spectra obtained with the WHT and HST/FOS of the edge–on S0 galaxy NGC 4342 have revealed a very steep central rotation curve and a strong central increase in velocity dispersion. These data suggest a large central mass concentration. In this chapter we presented detailed dynamical models of NGC 4342 used to investigate whether its nucleus harbors a massive BH.

We model the luminous density distribution of NGC 4342 with multiple Gaussian components. After projection and PSF convolution this model provides an excellent fit to the HST *I*–band surface brightness distribution. The parameters of this model were derived with the MGE method.

Simple isotropic Jeans models suggest that NGC 4342 harbors a massive BH of a few times $10^8 M_{\odot}$. The actual mass of the BH depends on the dataset fitted: the WHT data suggest $M_{\text{BH}} \sim 3 \times 10^8 M_{\odot}$; the HST/FOS data suggest a somewhat larger BH mass of $\sim 6 \times 10^8 M_{\odot}$. This discrepancy already suggests that the assumptions underlying the Jeans models, i.e., $f = f(E, L_z)$ and therefore $\sigma_R = \sigma_z$, are incorrect. This is also evident from the fact that the Jeans models cannot accurately fit the major axis rms velocities measured with the WHT. These rms velocities are independent of the freedom in the anisotropy σ_{ϕ}/σ_R allowed in the Jeans modeling. We find that for a mass–to–light ratio $\Upsilon_I = 6.2 M_{\odot}/L_{\odot}$ and an isotropic velocity distribution ($\sigma_R = \sigma_z = \sigma_{\phi}$) the Jeans model with $M_{\text{BH}} = 3 \times 10^8 M_{\odot}$ provides the best fit to the observed WHT velocity dispersions along the major axis. However, the rotation velocities are not very well fitted and we find a correlation between $V_{\text{obs}}/V_{\text{mod}}$ and the local ellipticity of the projected surface brightness, such that the model underpredicts the rotation velocities in the highly flattened regions (dominated by the disk light) and overpredicts them in the less flattened region (dominated by the bulge light). This suggests that the different components in NGC 4342 have different velocity anisotropies.

We thus constructed three-integral axisymmetric models of NGC 4342 in order to examine the mass of a possible BH and the dynamical structure of the different components. The modeling technique is an extension of Schwarzschild's orbit-superposition technique, and is based on finding the ensemble of orbits that best fits the observations. These models make no assumption about the dynamical structure and are fully general. This technique (see chapters 2 and 3) has previously been used to prove the existence of a massive BH of $(3.4 \pm 0.7) \times 10^6 M_\odot$ in the compact elliptical M32 (chapter 4). We have constructed a range of dynamical (M_{BH}, Y_I) -models of NGC 4342 to determine a central BH mass of $3.0^{+1.7}_{-1.0} \times 10^8 M_\odot$ and an *I*-band mass-to-light ratio of $6.3^{+0.5}_{-0.4} M_\odot / L_\odot$. The high spatial resolution of the HST/FOS data allow us to rule out models without a BH to a confidence level better than 99.73 per cent. With a similar confidence we can rule out models with a BH more massive than $7 \times 10^8 M_\odot$. This upper limit on the BH mass is mainly due to the VP shape parameters h_3 and h_4 . With the current data we can not rule out alternatives to a massive BH, such as a cluster of brown dwarfs or stellar remnants. Nevertheless, the QSO paradigm, together with the fact that the presence of massive BHs in galactic nuclei has unambiguously been demonstrated in a few galaxies where the inferred central densities are high enough to rule out dark clusters as alternatives (see discussion in Maoz 1997), make the interpretation of the inferred MDO in NGC 4342 in terms of a massive BH the most likely.

We computed the intrinsic mean velocities and velocity dispersions of the three-integral models. The dynamical structures of the best fitting models vary strongly with radius, reflecting the multi-component structure of NGC 4342. Between $2''$ and $12''$ in the equatorial plane the best fitting models change from azimuthally anisotropic to radially anisotropic, while $\sigma_z / \sigma_R \approx 0.9$. This explains the correlation between the projected ellipticity and the failure of the isotropic Jeans models to fit the observed rotation velocities along the major axis. The bulge in the best fitting model without BH is radially anisotropic. However, we have shown that even without the constraints of the measured HST/FOS rotation velocities, models without BH cannot fit the central HST/FOS velocity dispersion. The rotation velocities measured from the ground already constrain the amount of central radial anisotropy such that models without a BH cannot fit the high central velocity dispersion measured with the HST/FOS.

The BH mass thus derived contributes a fraction of $2.6^{+1.5}_{-0.9}$ per cent to the total mass of the bulge ($1.2 \times 10^{10} M_\odot$). With this BH mass, NGC 4342 has one of the highest ratios of BH mass over bulge mass. Currently, the BH in our own galaxy has, with 0.02 per cent, the lowest BH mass to bulge mass ratio known: the scatter in the M_{BH} vs. M_{bulge} relation seems to be as large as two orders of magnitude. Extremely high spatial resolution is required in order to investigate if other galaxies have even lower values of $M_{\text{BH}}/M_{\text{bulge}}$. In conclusion, current data are consistent with a relation between bulge mass and BH mass, but the scatter is very large, and it is likely that the current $M_{\text{BH}}/M_{\text{bulge}}$ ratios found are biased towards an upper limit. Although the newly installed Space Telescope Imaging Spectrograph (STIS) is likely to detect many more BH cases in the coming years, detection of BHs with masses of the order of a 0.02 per cent of the bulge mass or less in galaxies in Virgo or beyond, will probably have to await a next generation space telescope.

Recently, Merritt & Quinlan (1998) have proposed a self-regulating mechanism that can explain the relation between bulge mass and BH mass. In this theory, it is assumed that the growth of BHs is dependent on the global morphology of their host galaxies: box orbits, present in non-axisymmetric structures, are believed to be efficient in driving large amounts of gas into the nucleus, thus allowing the BH to grow. As already mentioned in Section 5.6.3, these box orbits are in turn strongly influenced by a central point mass: a triaxial bulge containing a massive BH evolves towards rounder or at least more axisymmetric shapes. Merritt & Quinlan

found, using N -body simulations, rapid evolution to nearly complete axisymmetry once the BH mass exceeded ~ 2.5 percent of the mass of the bulge. This ratio is interestingly close to that of NGC 4342 and NGC 3115, the current record holders (if we exclude NGC 4486B). This agreement suggests that $M_{\text{BH}}/M_{\text{bulge}}$ may indeed be limited by a feedback mechanism that turns off the supply of fuel once the BH has become massive enough to destroy the box orbits and thereby evolve to an axisymmetric shape.

Acknowledgments

Part of the calculations presented in this paper were performed on two UltraSparcs, kindly made available by the Lorentz Center. Many thanks are due to Eric Emsellem, for his help with the MGE analysis, and to Tim de Zeeuw, Walter Jaffe, HongSheng Zhao and Roeland van der Marel for many fruitful discussions. We are grateful to the referee for his helpful comments that improved the manuscript. FvdB was supported by a Hubble Fellowship, #HF-01102.11-97A, awarded by STScI. NC acknowledges the hospitality of Steward Observatory where part of this work was done.

References

- Bahcall, J. N., & Wolf, R.A. 1976, *ApJ*, 209, 214
 Binney, J. J., & Mamon, G. A. 1982, *MNRAS*, 200, 361
 Binney, J. J., & Tremaine, S. D. 1987, *Galactic Dynamics*. (Princeton: Princeton University Press)
 Binney, J. J., Davies, R. L., & Illingworth, G. D. 1990, *ApJ*, 361, 78
 Bower, G. A., et al. 1998, *ApJ*, 492, L111
 Cretton, N., de Zeeuw, P. T., van der Marel, R. P., & Rix, H.-W. 1999, accepted in *ApJ Supplements*, chapter 3 of this thesis
 Cohn, H. 1980, *ApJ*, 150, 163
 Dehnen, W. 1998, *AJ*, 115, 2384
 de Vaucouleurs, G., de Vaucouleurs, A., & Corwin, H. C. 1976, *Second Reference Catalogue of Bright Galaxies*. (University of Texas, Austin)
 de Zeeuw, P. T. 1997, in *The Nature of Elliptical Galaxies*, eds., M. Arnaboldi, G.S. Da Costa, P. Saha. ASP. Conference Series 116, p.44
 Eckart, A., & Genzel, R. 1997, *MNRAS*, 284, 576
 Emsellem, E., Monnet, G., & Bacon, R. 1994, *A&A*, 285, 723 (EMB94)
 Faber, S. M., Wegner, G., Burstein, D., Davies, R. L., Dressler, A., Lynden-Bell, D., & Terlevich, R. J. 1989, *ApJS*, 69, 763
 Ferrarese, L., Ford, H. C., & Jaffe, W. 1996, *ApJ*, 470, 444
 Ferrarese, L., Ford, H. C. 1999, *ApJ*, 515, 583
 Ford, H. C., et al. 1994, *ApJ*, 435, L27
 Ford, H. C., Tsvetanov, Z. I., Ferrarese, L., & Jaffe, W. 1998, in *IAU Symp. 184, The Central Regions of the Galaxy and Galaxies*, ed. Y. Sofue (Dordrecht: Kluwer), in press
 Gebhardt, K., et al. 1999, *AJ*, in press
 Gerhard, O. E. 1988, *MNRAS*, 232, 13P
 Goodman, J. & Binney, J. 1984, *MNRAS*, 207, 511
 Goodman, J. & Lee, H. M. 1989, *ApJ*, 337, 84
 Harms, R. J., et al. 1994, *ApJ*, 435, L35
 Ho, L. C. 1998, in *Observational Evidence for Black Holes in the Universe*, ed. S. K. Chakrabarti (Dordrecht: Kluwer), (astro-ph/9803307)
 Hunter, C. 1977, *AJ*, 82, 271

- Jacoby, G. H., Ciardullo, R., & Ford, H. C. 1990, *ApJ*, 356, 332
- Jaffe, W., Ford, H. C., Ferrarese, L., van den Bosch, F. C., & O'Connell, R. W. 1996, *ApJ*, 460, 214
- Kandrup, H. E., & Mahon, M. E. 1994, *J. Phys. E*, 49, 3735
- Kormendy, J., & Richstone, D. 1995, *ARA&A*, 33, 581
- Kormendy, J., et al. 1996a, *ApJ*, 459, L57
- Kormendy, J., et al. 1996b, *ApJ*, 473, L91
- Kormendy, J., et al. 1997, *ApJ*, 482, L139
- Kormendy, J., Bender, R., Evans, A. S., & Richstone, D. 1998, *AJ*, 115, 1823
- Kuijken, K., & Merrifield, M. R. 1993, *MNRAS*, 264, 712
- Lawson, C. L., & Hanson, R. J. 1974, *Solving Least Squares Problems*. (Englewood Cliffs: New Jersey)
- Levison, H. F., & Richstone, D. O. 1985, *ApJ*, 295, 349
- Lynden-Bell, D. 1967, *MNRAS*, 136, 101
- Lynden-Bell, D. 1969, *Nature*, 223, 690
- Macchetto, F., Marconi, A., Axon, D. J., Capetti, A., Sparks, W. B., & Crane, P. 1997, *ApJ*, 489, 579
- Maoz, E. 1998, *ApJ*, 494, L181
- Merrifield, M. R., & Kuijken, K. 1994, *ApJ*, 432, 575
- Merritt, D. 1993, *ApJ*, 413, 79
- Merritt, D. 1996, *AJ*, 112, 1085
- Merritt, D., & Valluri, M. 1996, *ApJ*, 471, 82
- Merritt, D. 1998, *Comments in Astrophysics*, Vol. 19 (astro-ph/9803211)
- Merritt, D., & Quinlan, G. D. 1998, *ApJ*, 498, 625
- Miyoshi, M., et al. 1995, *Nature*, 373, 127
- Nandra, K., George, I. M., Mushotzky, R. F., Turner, T. J., & Yaqoob, T. 1997, *ApJ*, 477, 602
- Pfenniger, D. 1984, *A&A*, 141, 171
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, *Numerical Recipes*. (Cambridge University Press: Cambridge)
- Quinlan, G. D., Hernquist, L., & Sigurdsson, S. 1995, *ApJ*, 440, 554
- Quinlan, G. D. 1996, *New Astronomy*, 1, 255
- Rees, M. J. 1984, *ARA&A*, 22, 471
- Richstone, D. O., & Tremaine, S. D. 1984, *ApJ*, 286, 27
- Richstone, D. O., & Tremaine, S. D. 1985, *ApJ*, 296, 370
- Richstone, D. O., & Tremaine, S. D. 1988, *ApJ*, 327, 82
- Richstone, D. O. 1997, in *The Nature of Elliptical Galaxies*, eds., M. Arnaboldi, G.S. Da Costa, P. Saha. ASP. Conference Series 116, p.123
- Richstone, D. O. 1998, in *IAU Symp. 184, The Central Regions of the Galaxy and Galaxies*, ed. Y. Sofue (Dordrecht: Kluwer)
- Rix, H.-W. 1993, in *IAU Symp. 153, Galactic Bulges*, eds., H. Dejonghe, & H.J. Habing. (Dordrecht: Kluwer), p. 423
- Rix, H.-W., de Zeeuw, P. T., Cretton, N., van der Marel, R. P., & Carollo, C. M. 1997, *ApJ*, 488, 702, chapter 2 of this thesis
- Roberts, M. S., Hogg, D. E., Bregman, J. L., Forman, W. R., & Jones, C. 1991, *ApJS*, 75, 751
- Sandage, A., & Tammann, G. A. 1981, *A Revised Shapley-Ames Catalogue of Bright Galaxies*. (Carnegie Institute of Washington: Washington DC)
- Satoh, C. 1980. *PASJ*, 32, 41
- Schwarzschild, M. 1979, *ApJ*, 232, 236
- Schwarzschild, M. 1982, *ApJ*, 263, 599
- Scorza, C., & van den Bosch, F. C. 1998, *MNRAS*, 300, 469

- Sellwood, J. A. 1989, *MNRAS*, 238, 115
- Sigurdsson, S., Hernquist, L., & Quinlan, G. D. 1995, *ApJ*, 446, 75
- Simien, F., Pellet, A., & Monnet, G. 1979, *A&A*, 72, 12
- Spergel, D. N., & Hernquist, L. 1992, *ApJ*, 397, L78
- Spitzer, L., & Hart, M. H. 1971, *ApJ*, 164, 399
- Spitzer, L., & Thuan, T. X. 1972, *ApJ*, 175, 31
- Tammann, G. A., & Sandage, A. 1985, *ApJ*, 294, 81
- Tanaka, Y., et al. 1995, *Nature*, 375, 659
- Toomre, A. 1981, in *Structure and Evolution of Normal Galaxies*, eds. S. M. Fall & D. Lynden-Bell (Cambridge: Cambridge University Press), p. 111
- van den Bosch, F. C., Ferrarese, L., Jaffe, W., Ford, H. C., & O'Connell, R. W. 1994, *AJ*, 108, 1579
- van den Bosch, F. C., & van der Marel, R. P. 1995, *MNRAS*, 274, 884
- van den Bosch, F. C., & de Zeeuw, P. T. 1996, *MNRAS*, 283, 381
- van den Bosch, F. C., & Emsellem, E. 1998, *MNRAS*, 298, 267
- van den Bosch, F. C., Jaffe, W., & van der Marel, R. P. 1998, *MNRAS*, 293, 343 (BJM98)
- van der Marel, R. P. 1991, *MNRAS*, 253, 710
- van der Marel, R. P. 1994, *MNRAS*, 270, 271
- van der Marel, R. P. 1997, in *IAU Symp. 186, Galaxy Interactions at low and high Redshift*, eds. D. B. Sanders & J. Barnes (Dordrecht: Kluwer)
- van der Marel, R. P., & Franx, M. 1993, *ApJ*, 407, 525
- van der Marel, R. P., & de Zeeuw, P. T., Rix, H.-W., & Quinlan, G. D. 1997, *Nature*, 385, 610
- van der Marel, R. P., Cretton, N., de Zeeuw, P. T., & Rix, H.-W., 1998, *ApJ*, 493, 613, chapter 4 of this thesis
- van der Marel, R. P., & van den Bosch, F. C. 1998, *AJ*, 116, 2220
- Young, P. J. 1980, *ApJ*, 242, 1232
- Zhao, H. S. 1996, *MNRAS*, 283, 149
- Zwicky, F., & Herzog, E. 1966, *Catalogue of Galaxies and of Clusters of Galaxies*, Vol. III. California Institute of Technology

Chapter 6

The Distribution of Stellar Orbits in NGC 2320

Cretton, N., Rix, H.-W., de Zeeuw P. T.
1999, to be submitted to MNRAS

We present direct observational constraints on the internal velocity distribution of the giant elliptical NGC 2320. Spectroscopic data within one effective radius along multiple position angles is used to derive the line-of-sight velocity distribution, quantified by the Gauss-Hermite moments. In addition, the gas rotation curve and dispersion profile are measured from the [OIII] emission lines. After correcting for the asymmetric drift, we calculate the circular velocity of the gas, which provides an independent constraint on the normalization of the gravitational potential.

To interpret the stellar motions, we build axisymmetric three-integral dynamical models by means of an extension of the Schwarzschild orbit-superposition technique. We construct models in which the mass follows the light (i.e., no dark matter) and models with a logarithmic gravitational potential.

Using χ^2 -statistics, we combine the stellar and gas data to constrain the value of the mass-to-light ratio Y_V (in the V-band). We find $Y_V = 15.0 \pm 0.6 h_{75}$ for the mass-follows-light models and $Y_V = 17.05 \pm 0.7 h_{75}$ for the logarithmic models. For the latter, Y_V is defined as the mass enclosed within $15''$ divided by the total amount of light in the same volume.

Radially constant Y_V models (without dark matter) and logarithmic models (with dark matter) provide comparable good fits to the data and possess similar dynamical structure. For the full range of Y_V permitted by the observational constraints, the models are radially anisotropic along the major axis in the interval meaningfully constrained by the kinematical data ($1'' \lesssim r \lesssim 40''$). Along the true minor axis, they are more nearly isotropic. The best fitting model has $\sigma_r/\sigma_{\text{total}} \simeq 0.7$, $\sigma_\phi/\sigma_{\text{total}} \simeq 0.5 - 0.6$ and $\sigma_\theta/\sigma_{\text{total}} \simeq 0.5$ in the equatorial plane. We explore the uncertainties associated with the anisotropy profiles by examining the formal 3σ -range of models around the best fitting Y_V .

THE orbital distribution of stars, often quantified by the (an)-isotropy of the local velocity dispersion, is a useful measure of the dynamical state of a galaxy. It provides constraints on galaxy formation and helps discriminating between N-body merger remnants. Furthermore, it has been a classic diagnostic to study the mass distribution of galaxies.

Anisotropy profiles can only be obtained from observations through a dynamical model, since only line-of-sight velocities can be measured for external galaxies. For a number of reasons the construction of such models is more difficult for elliptical galaxies than for spirals. Elliptical galaxies generally have no simple tracer population that could be used to derive the underlying gravitational potential (like HI disks in spirals). These systems are known to support a variety of orbital shapes that largely overlap each other. Furthermore there is a well-known degeneracy between orbital anisotropy and mass. On the observational side, stellar velocity distributions in ellipticals usually have to be derived from absorption lines, which are harder to measure accurately than emission lines.

This situation is improving on various fronts: fully anisotropic dynamical models can now be constructed and matched to observations (see e.g., Rix *et al.* 1997, chapter 2 of this thesis;

Richstone *et al.* 1997; Gerhard *et al.* 1998; Cretton *et al.* 1999, chapter 3 of this thesis). The mass-anisotropy degeneracy can in principle be broken with the use of the full line-of-sight velocity distributions, also called Velocity Profiles (hereafter VPs) (see e.g., Gerhard 1993). Measurements of VPs at several positional angles and extended radii further constrain the dynamical structure of the galaxy (see e.g., Carollo *et al.* 1995; Statler, Smecker-Hane & Cecil 1996).

What is known about the intrinsic dispersion profiles in elliptical galaxies? Keeping in mind the various limitations (in the modeling or in the observations), we can already draw some conclusions from a few recent studies. In chapter 2, we constructed spherical anisotropic models for NGC 2434 and we concluded that half of the mass inside one effective radius (R_{eff}) needs to be dark. Inside this range, the dispersion profiles are flat and obey $\sigma_r/\sigma_{\text{total}} = 0.7$, $\sigma_\theta/\sigma_{\text{total}} = \sigma_\phi/\sigma_{\text{total}} = 0.5$. Matthias & Gerhard (1999) built axisymmetric 3-integral models of NGC 1600. They did not need any DM inside $1 R_{\text{eff}}$ and deduced radial anisotropy in the outer parts ($\sigma_\theta/\sigma_r \simeq \sigma_\phi/\sigma_r \simeq 0.7$), and a more isotropic structure in the center. In their spherical models of NGC 6703, Gerhard *et al.* 1998 again found near isotropy in the center, and a mild radial anisotropy in the outer parts, as in NGC 1399 (Kronawitter *et al.* 1999, in preparation). Dejonghe *et al.* (1996) inferred tangential anisotropy in NGC 4697, as in NGC 1700 (Statler *et al.* 1999) based on axisymmetric 3-integral models built with a quadratic programming technique, but no VPs were used in these two studies. Using fully general axisymmetric models, Gebhardt and collaborators found that NGC 3379, NGC 3377, NGC 4473 and NGC 5845 were radially anisotropic in the range $(0.1 - 1) R_{\text{eff}}$ (Gebhardt *et al.* 1999). Merritt & Oh (1987) derived a slight radial anisotropy $\sigma_r \simeq 1.2\sigma_\phi = 1.2\sigma_\theta$ in the main body of M87.

In this paper, we derive the anisotropy of the giant elliptical NGC 2320, combining constraints from the stellar VPs and from the gas rotation and dispersion curves. The simple kinematics of the gas (mainly circular rotation) allows an independent and straightforward measure of the total mass of the system. Cinzano & van der Marel (1994) constructed dynamical models for NGC 2974 and found that the velocities of the ionized gas disk were consistent with the potential derived from the stellar kinematics.

The paper is organized as follows. In Section 6.1, we present the photometric and kinematic data. We describe the mass and dynamical models in Section 6.2. The range of statistically acceptable models with their associate dynamical structure is derived in Section 6.3. Sec-

TABLE 6.1— Summary of the kinematic observations.

j	name	telescope	slit width	date	radial extent	# of points	S/N
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	MMT-140	MMT	3.5	17 Feb 1996	37	23	25
2	MMT-50	MMT	3.5	17 Feb 1996	23	12	25
3	050	KPNO 4m	2.5	2 Mar 1998	17	8	12
4	100	KPNO 4m	2.5	27 Feb 1998	18	10	10
5	130	KPNO 4m	2.5	27 Feb 1998	27	11	12
6	142	KPNO 4m	2.5	2 Mar 1998	24	12	12
7	175	KPNO 4m	2.5	27 Feb 1998	19	10	10
8	b000	KPNO 4m	2.5	1 Mar 1998	24	9	15
9	b020	KPNO 4m	2.5	1 Mar 1998	27	10	15
10	b100	KPNO 4m	2.5	27 Feb 1998	45	18	10
11	b130	KPNO 4m	2.5	27 Feb 1998	44	19	10
12	b175	KPNO 4m	2.5	27 Feb 1998	43	16	10

Column (1) gives the number of the spectrum; column (2) its name (chosen after its position angle); column (3) the telescope where it was taken; column (4) the slit width (in arcseconds); column (5) the date of observation; column (6) the maximum radial extension (in arcseconds); column (7) the number of data points (after pixel binning) for the stellar data; column (8) the S/N for the stellar data.

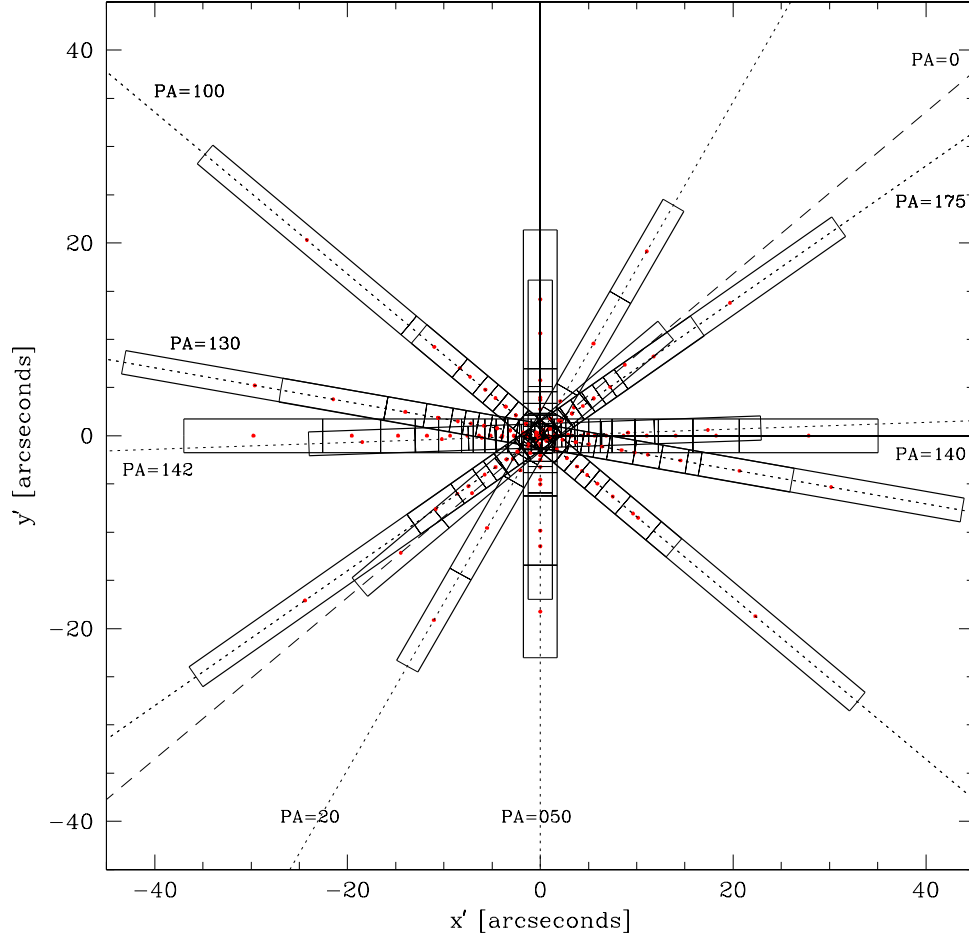


FIGURE 6.1— Position of all the slits used to constrain the dynamical models. The values of the various position angles are indicated (the major axis is $PA=140$ and the minor axis is $PA=50$). The effective radius R_{eff} is at $29.5''$.

tion 6.4 summarizes the results and discusses them in the context of cosmological simulations and mergers.

6.1 The data

6.1.1 Photometric data

NGC 2320 is a luminous elliptical galaxy with a heliocentric velocity of 5725 km/s, implying a distance of $76.3 h_{75}$ Mpc. It has an apparent magnitude $m_V = 11.9$ (NED), which translates to an absolute magnitude of $M_V = -22.5$. The galaxy image was taken at the Steward Observatory 90'' Telescope on March, 1996 on a 2k x 2k CCD (exposure time 600 seconds) in the V-band. The scale was 0.283 arcseconds. The image was calibrated using published aperture photometry available via the Hypercat catalogue (CRAL, Lyon), mostly consisting of data taken at the OHP (Prugniel & Heraudeau 1998). The effective radius R_{eff} (i.e., enclosing half of the light) is $29.5''$. Our photometric data is not good enough to detect a possible central luminosity cusp (the seeing was about $1''$) and no HST archive image exists. Anyway in this paper, we do not con-

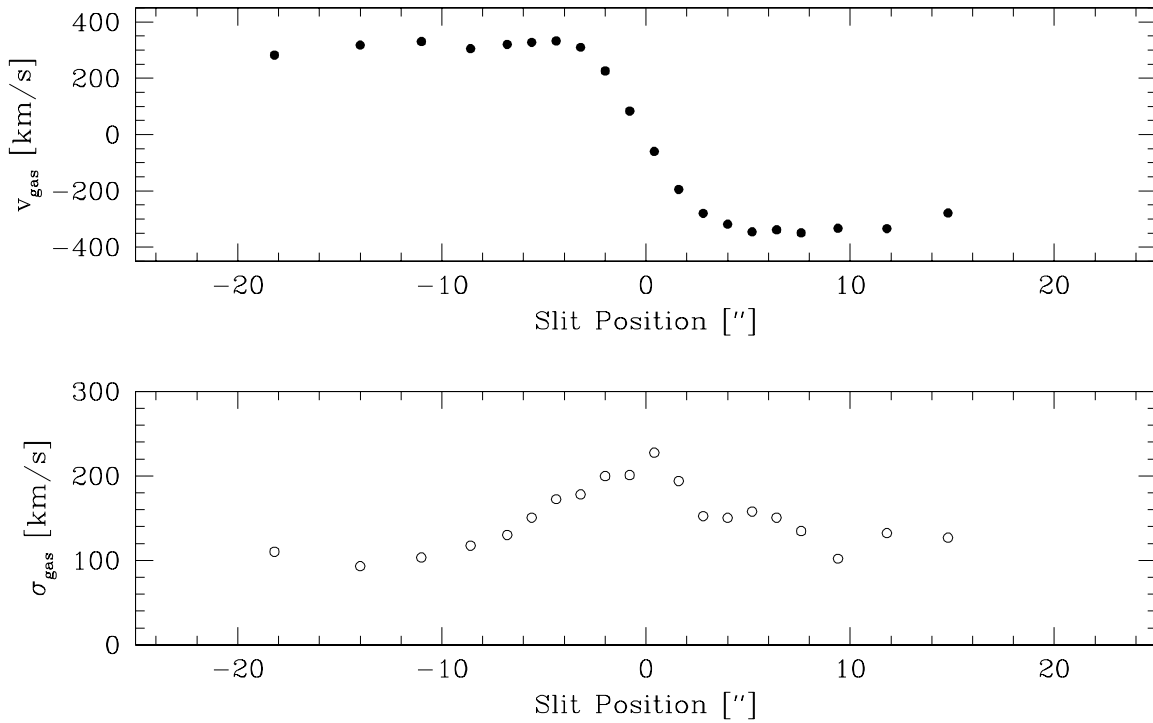


FIGURE 6.2— Mean line-of-sight velocity and dispersion of the gas along the major axis (MMT spectra). The slow rise of the velocity is due to smearing inside the wide slit (3.5").

concentrate on the core properties of NGC 2320, since we have used a wide slit for the spectroscopic observations.

6.1.2 Stellar kinematic data

NGC 2320 was observed during three different runs (February 96, February 98 and March 1998) at the Multi Mirror Telescope (MMT) and the 4 meter telescope at Kitt Peak. Table 6.1 summarizes the observations and Figure 6.1 illustrates all slit positions on the plane of the sky. All these various spectra are statistically independent constraints even if they have the same PAs, because they have a different radial extent (and binning) and, in some cases, have been obtained at a different telescope.

After standard reduction procedure, the VPs were extracted from the absorption lines and quantified using the Gauss-Hermite decomposition (see e.g. Rix & White 1992, van der Marel & Franx 1993). With the same notation as in Cretton & van den Bosch (1999, chapter 5 of this thesis), we parameterize these velocity profiles (VPs) using Gauss-Hermite series with line strength γ , mean radial velocity V , and velocity dispersion σ as free parameters. The anisotropy is reflected in the shape of the VPs (and hence in the values of the GH-moments h_i). For instance VPs more peaked than a Gaussian usually originate from radial anisotropy and translate into greater h_4 values than their isotropic counterparts. Note that the VP shapes not only depend on the dynamical structure, but also on the gravitational potential, which is fortunate since they can then be used to disentangle anisotropy from mass (Dejonghe 1987, Merritt 1993, Gerhard 1993, see also Figure 2 in Gerhard *et al.* 1998). In total we have 632 stellar kinematic constraints (158 data points \times 4 GH-moments $h_{1,\dots,4}$, see Table 6.1).

A careful look at the kinematic data (Figure 6.6) reveals some *systematic* problems, e.g., about

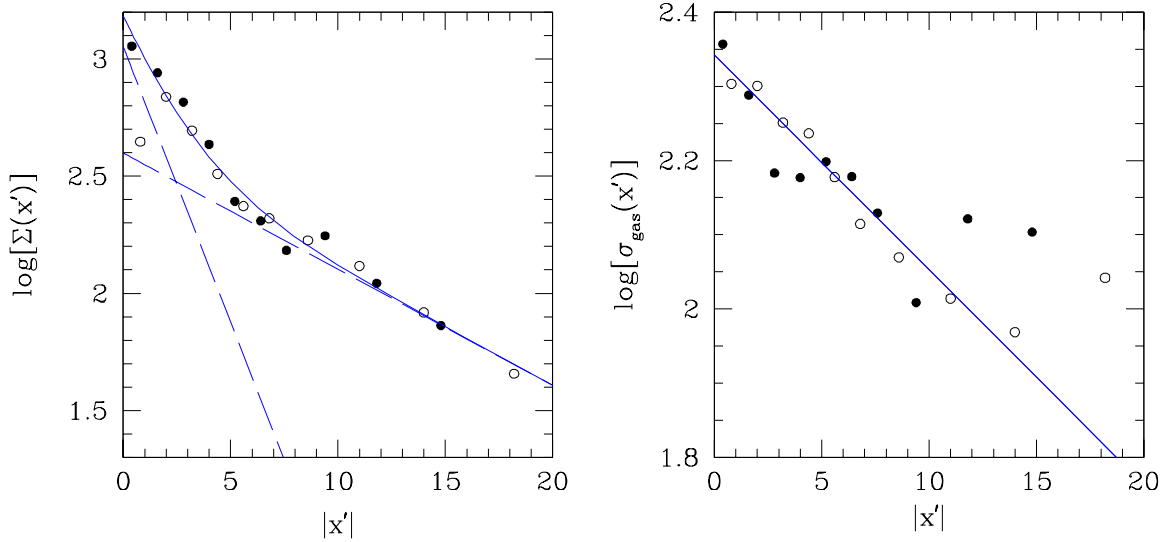


FIGURE 6.3— Surface density profile of the ionized gas disk (left) and dispersion profile (right) along the major axis. The dots are the measured quantities, the full line is the fit. On the left panel, the dashed lines are the individual exponentials. Both sides of the major axis are shown: full symbols correspond to the positive side and open symbols to the negative side of the major axis.

half of the slits show a significantly higher ($\sim 40 \text{ km s}^{-1}$) central dispersion value than for the remaining half. Given their formal errors ($\sim 10 \text{ km s}^{-1}$), such large differences amongst various slits can not be explained statistically.

6.1.3 Kinematics of the ionized gas

After subtraction of the broadened stellar template, emission was found at 5007 \AA and 4965 \AA , corresponding to the [OIII] lines of ionized gas (see e.g., Figure 2 of Rix *et al.* 1995). However, only the MMT data was of sufficient S/N to permit measurement of mean velocity and velocity dispersion.

Usually, gas is assumed to rotate in the equatorial plane of a galaxy on nearly closed orbits. However, this is only true if the velocity dispersion is negligible compared to the mean rotation velocity. In Figure 6.2, we show the mean line-of-sight velocity and dispersion of the gas on the MMT major axis: The velocity reaches 325 km/s (inside $20''$) and the dispersion decreases roughly exponentially from 220 km/s (center) to 100 km/s (at $15''$) and thus can not be neglected.

To derive the true circular velocity $v_c^2 = R \partial \Phi_{\text{grav}} / \partial R$ from the observed v_{gas} and σ_{gas} , we proceed as follows (see e.g., Neistein *et al.* 1999). We first obtain the mean rotation velocity of the gas, $v_\phi(R)$, in the equatorial plane (i.e., the plane of the disk) by de-projecting the observed v_{gas} . Along the major axis, we have $v_\phi(R) = v_{\text{gas}} / \sin i$, where i is the inclination of the galaxy (away from face-on). Similarly, we have $\sigma_\phi(R) = \sigma_{\text{gas}} / \sin i$. Note that in nearly edge-on disks, this is only approximately true, since the line-of-sight integration through the disk will reduce v_{gas} relative to v_ϕ . Neistein *et al.* (1999) estimated that this correction is less than 4% for inclinations $i < 70^\circ$.

We follow Binney & Tremaine (1987, eq. 4–33) to obtain the circular velocity of the cold gas disk:

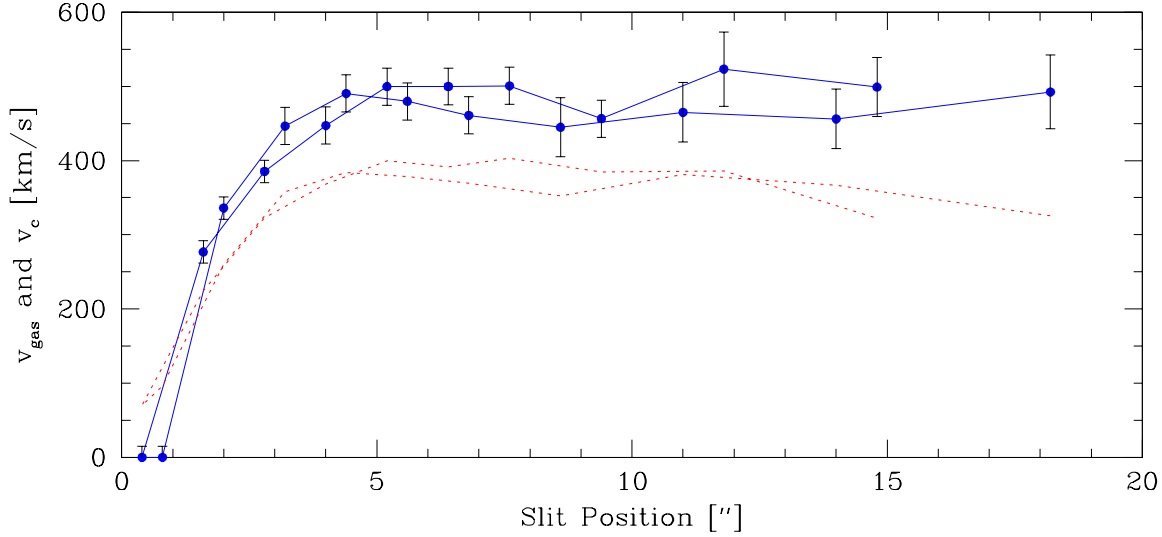


FIGURE 6.4— Circular velocity (full line) when the asymmetric correction has been applied. For comparison, we plotted also the observed gas velocity (dotted line). Both sides of the major axis have been folded to the positive side. Only points at radii larger than $3''$ should be taken seriously, since during the correction we have assumed a flat rotation curve.

$$v_c^2 = \overline{v_\phi^2} + \sigma_\phi^2 - \sigma_R^2 - \frac{R}{\rho} \frac{\partial(\rho \sigma_R^2)}{\partial R} - R \frac{\partial(\overline{v_R v_z})}{\partial z}, \quad (6.1)$$

where we assume that the final term can be neglected (i.e., the gas dynamics is close to isotropy). We fit the luminosity density of the gas as $\rho(R) = \rho_1 \exp(-R/R_1) + \rho_2 \exp(-R/R_2)$ as shown in the left panel of Figure 6.3. For a flat rotation curve, the epicycle approximation gives $\sigma_R^2(R) = 2\sigma_\phi^2(R)$. Since the term σ_R enters a derivative in eq. (6.1) we find it convenient to use an analytic expression instead of the (noisy) data points, so we fit an exponential to $\sigma_{\text{gas}}(R) = \sigma_0 \exp(-R/R_0) = \sigma_\phi(R) \sin i$ (right panel in Figure 6.3). We can now evaluate the circular velocity (Figure 6.4). The error bars for $v_c(R)$ have been estimated from the left/right differences of the rotation curve.

6.2 The models

6.2.1 The mass model

As in chapter 5, we have used the Multi Gaussian Expansion (MGE) of Emsellem *et al.* (1994) to construct a mass model for NGC 2320. Briefly, in this formalism the surface brightness profile, the mass density distribution and the PSF are all described as a sum of Gaussian components. Free parameters include the center of each Gaussian, its position angle, flattening, central intensity, and size (i.e., standard deviation) along the major axis.

We find that the density profile of NGC 2320 can be well fitted with 5 Gaussian components (see Table 6.2), all with the same position angle and center. The mass density is expressed as

$$\rho(R, z) = \Upsilon \sum_{i=1}^5 I_i \exp \left[-\frac{1}{2\sigma_i^2} \left(R^2 + \frac{z^2}{q_i^2} \right) \right], \quad (6.2)$$

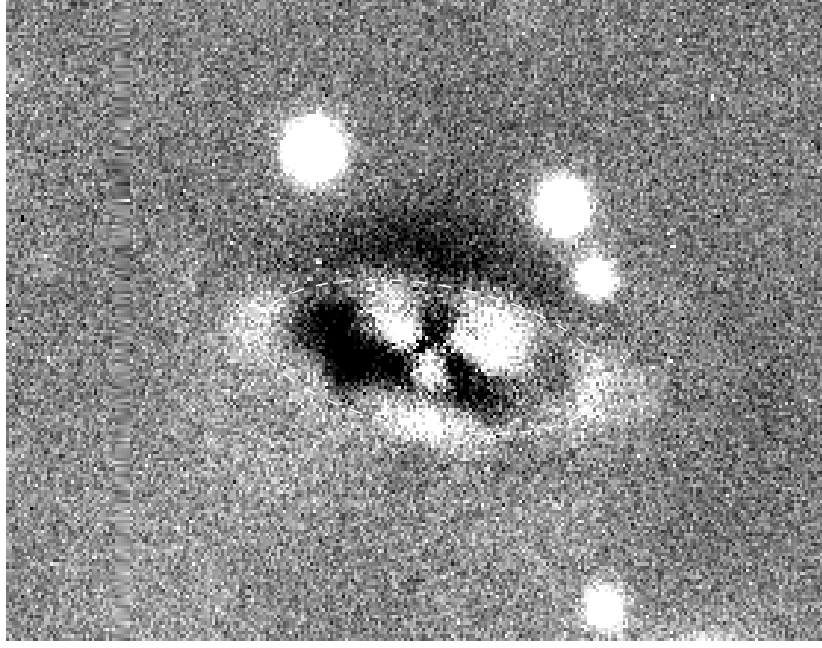


FIGURE 6.5— Ring-like structure once the MGE model has been subtracted from the image. The big white dots outside the ring are foreground stars. We superpose an ellipse with an axis ratio of 0.41 and a semi major axis of $17''$ on top of the ring (white dashed line).

where Y is the mass-to-light ratio, I_i is the central intensity (in $L_\odot/\text{arcsecond}^3$), σ_i the standard deviation (in arcseconds) and q_i the flattening of each Gaussian. The corresponding gravitational potential (and forces) are obtained by solving Poisson equation and can be written as one-dimensional quadratures (see chapter 5 for details). In this case, the mass distribution follows the light.

When we subtract the MGE model from the image, a disk-like structure appears with a radial extent of $\sim 17''$ and an axis ratio of ~ 0.5 (Figure 6.5). Assuming this ring of excess light lies in the equatorial plane, we have a constraint on the inclination angle. From the apparent axis ratio of 0.5, we deduce an inclination $i \approx 60^\circ$, which we adopt for the remainder of this paper. The apparent flattening of the galaxy (0.6–0.7) makes it unlikely to be much more inclined, since galaxies intrinsically as flat as E6 are very rare. Van den Bosch & Emsellem (1998) found a similar stellar ring in the nuclear region of NGC 4570.

We have explored another set of models in which the total gravitational potential and the luminous density are not related by Poisson equation: models with a logarithmic potential

$$\Phi_{\log}(R, z) = \frac{1}{2} v_0^2 \ln \left(R_c^2 + R^2 + \frac{z^2}{q_\Phi^2} \right). \quad (6.3)$$

We still use the MGE density law (equation 6.2) to describe the luminosity profile of NGC 2320. We have chosen the simple logarithmic potential (with flat circular velocity at large radii) to build a sequence of models with dark matter halos. We took $q_\Phi = 0.83$ such that the flattening of the corresponding mass density profile is similar to that of the MGE models, $q_\rho \sim 0.5 - 0.6$. We adjust R_c such as to match the inner parts of the circular velocity curve and we scale the various models with v_0 . The quantity R_c is probably an upper limit of the true core radius, since

we did not fit deconvolved data. Nevertheless it is not likely to have a large influence on the main results of this paper.

In chapter 2, we used dark halo profiles suggested by cosmological N-body simulations (Navarro, Frenk, & White, 1996) and adiabatically contracted them to account for the luminous material. We found that the best fit dark halo profile was nearly indistinguishable from a simple logarithmic model over the radii probed by the kinematics of integrated light.

6.2.2 The dynamical models

The method used here to construct dynamical models is based on the orbit superposition technique of Schwarzschild and was described extensively in chapter 3. We limit ourselves here to the specific parameters we used for NGC 2320.

We sample our orbits using a grid in integral space, energy E , vertical component of the angular momentum L_z and third integral I_3 . We use 20 values of $E = 1/2 R_c \partial\Phi/\partial R + \Phi(R_c, 0)$, sampled through the radius of the circular orbit R_c . We take a logarithmic sampling of R_c in $[0.1'', 300.0'']$, since less than one percent of the total mass of our MGE model lies outside this range. 14 values of L_z per E in $[-L_{z,\max}, +L_{z,\max}]$ and 7 values of I_3 per (E, L_z) were adopted as in chapter 5. We do not imply that I_3 is defined for every value of (E, L_z) , i.e., every orbit is regular, but instead we sample I_3 by choosing starting points on the zero-velocity curve (see chapter 3): when the orbit is indeed regular, this starting point can truly be interpreted as a third integral.

The orbit library is constructed by numerical integration of each trajectory for a fixed amount of time (200 periods of the circular orbit at that E) using a Runge–Kutta scheme. During integration, we store the fractional time spent by each orbit in a cartesian “data-cube” (x', y', v_{los}) , which is further convolved with the PSF and binned into the various slits, where (x', y') are the projected coordinates on the sky and v_{los} , the line-of-sight velocity. Furthermore, we adopt logarithmic polar grids in the meridional plane (R, z) and in the (x', y') plane with the same R_c -radial range and sampling. Orbital occupation times are stored in these grids to make sure the final orbit model reproduces (to a few percent accuracy) the MGE mass model. The lowest order velocity moments of each orbit are also stored on such a grid for analysis of the intrinsic dynamical structure of the model (see chapter 3 for details).

The mass on each orbit is computed with the NNLS algorithm (Lawson & Hanson 1974), such that the non-negative superposition of all orbits aims at reproducing the kinematic data (within the errors) and the projected and intrinsic MGE mass profile. Smoothness in integral space is enforced through a regularization technique. It allows the derivation of smooth anisotropy profiles of a few selected models.

TABLE 6.2— Parameters of MGE model for the luminous density profile.

j	I_j	a_j	q_j
(1)	(2)	(3)	(4)
1	287764585.49	0.820	0.544
2	18868477.1	3.000	0.443
3	1826940.9	8.585	0.419
4	219268.2	19.061	0.505
5	9359.8	56.370	0.676

Column (1) gives the index number of each Gaussian. Column (2) gives its central luminosity density (in L_\odot/arcsec^3); column (3) its standard deviation (which expresses the size of the Gaussian along the major axis) and column (4) its flattening. All Gaussians have the same position angle and the same center.

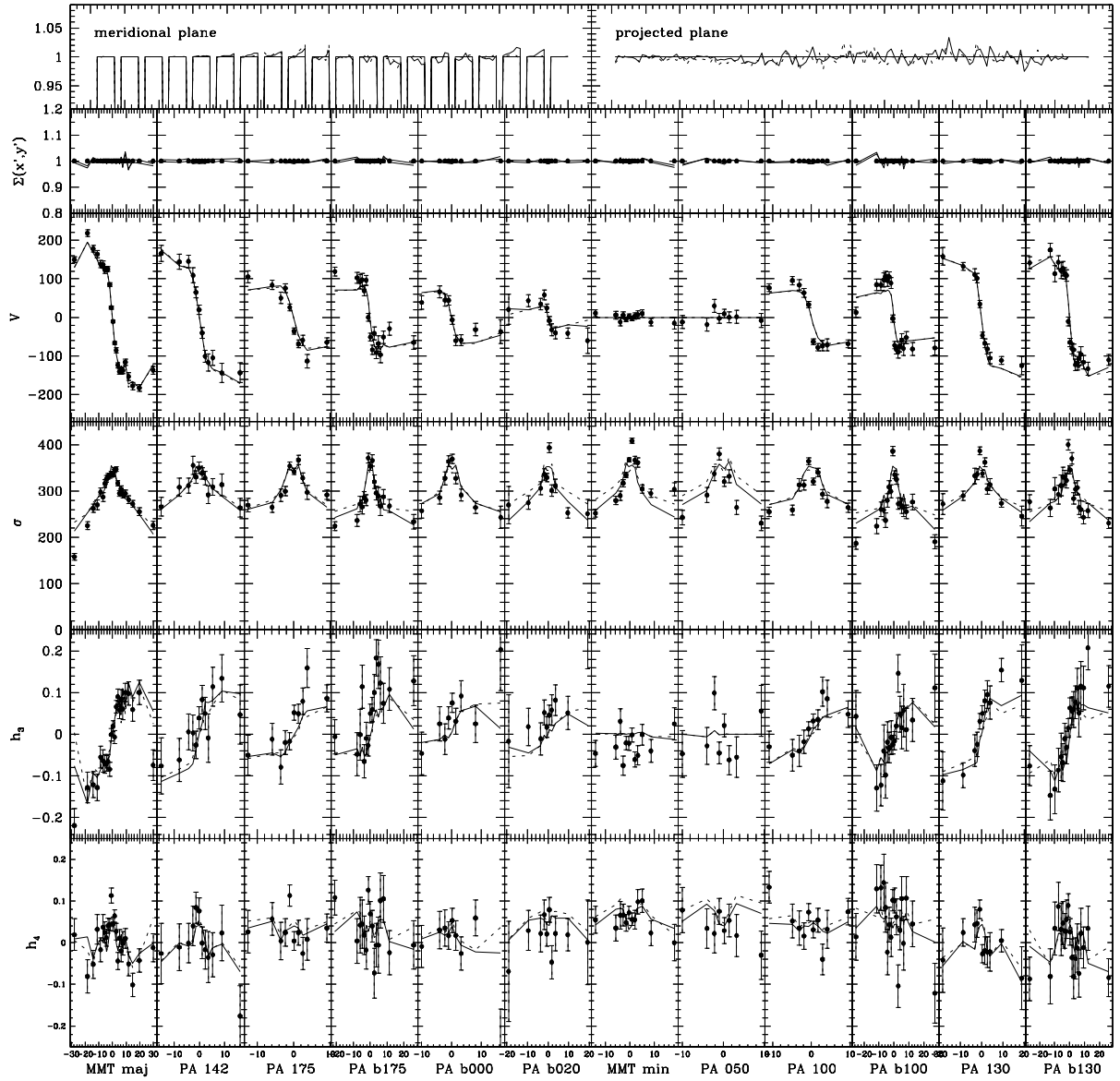


FIGURE 6.6— Kinematic data for all the slits and fits of the two best fit models: the full line is the model where the mass follows the light and the dotted line is the model with the logarithmic potential. The columns have been organized in increasing Position Angle order, starting from the MMT major axis. The top line shows the fit of the (normalized) mass constraints in the meridional plane and in the plane of the sky. In the fit of the meridional plane constraints, we exclude the cells closest to the symmetry axis. This explains the jagged appearance of the fit for these constraints.

6.3 Results

6.3.1 Fits to the data

In Figure 6.6, we show the two best fit models to all the data, in the case where mass follows light (full line) and in the logarithmic potential case (dotted line). The first row displays the fit to the mass constraints (intrinsic and projected), normalized to unity. The second row shows the fit to the integrated surface density in the slit bins (also normalized to unity). The subsequent rows show the kinematic data and the fits of both models. Each column corresponds to a different

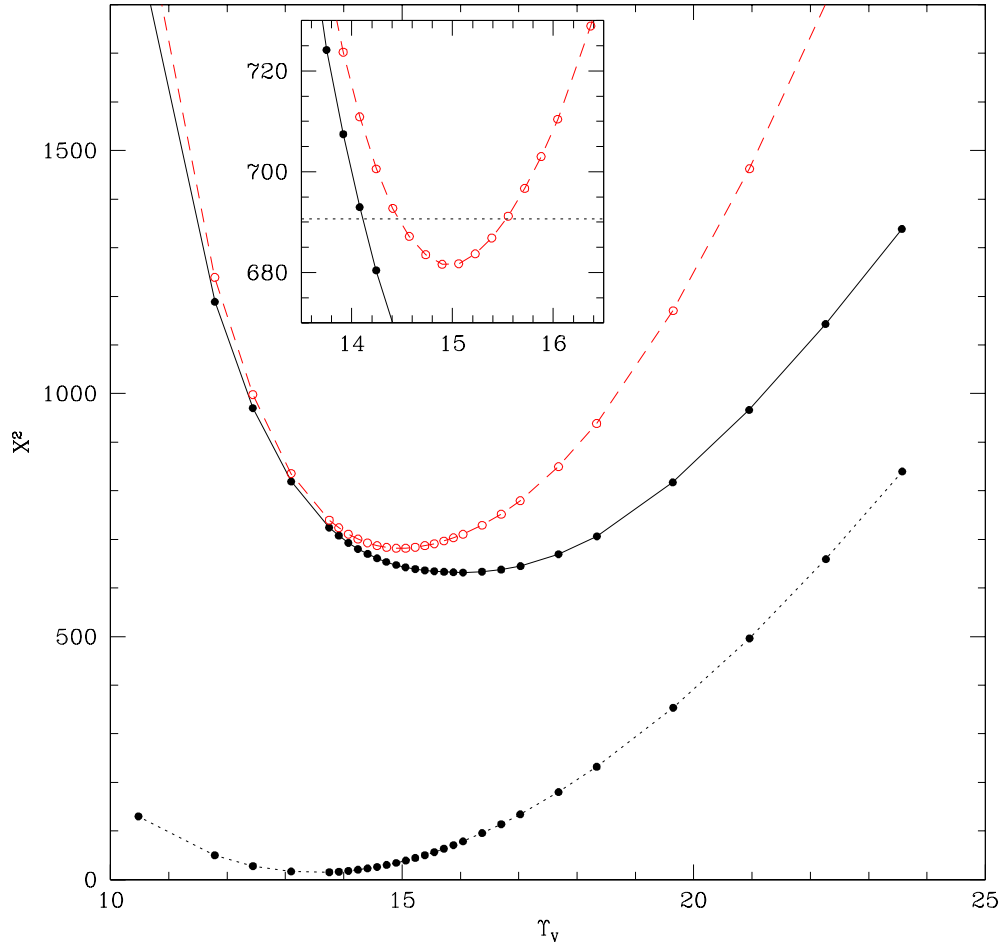


FIGURE 6.7— χ^2 distributions for the models with spatially constant Υ_V . The dotted line is χ^2_{gas} , the full line is χ^2_{stars} and the dashed line (with open dots) corresponds to the combination of both type of constraints. The dots represent the actual model calculations. The inset shows in more details the region near the (combined) χ^2 minimum. The horizontal dotted line has $\chi^2 = \chi^2_{\text{min}} + 9$.

slit (see Table 6.1).

Both models (i.e., with and without DM) fit the data comparatively well. The differences appear at large radii (e.g., in the velocity dispersions), where the circular velocities of the two models (i.e., the enclosed masses) start to diverge. The two models have problems fitting about half of the central dispersion points. This may be due to the absence of a density cusp (and/or central black hole) in our MGE model, but as mentioned earlier (see Section 6.1.2), there are systematic problems with the data in the center. Moreover, for some position angles, the observed dispersion seems to drop too fast compared to the models, whereas for other position angles, it could be well fitted.

6.3.2 Goodness of fit

We wish to assess statistically which value of Υ_V provides the best fit to the data (in terms of χ^2). We obtain a set of different models (with different Υ_V) by rescaling one orbit library. We

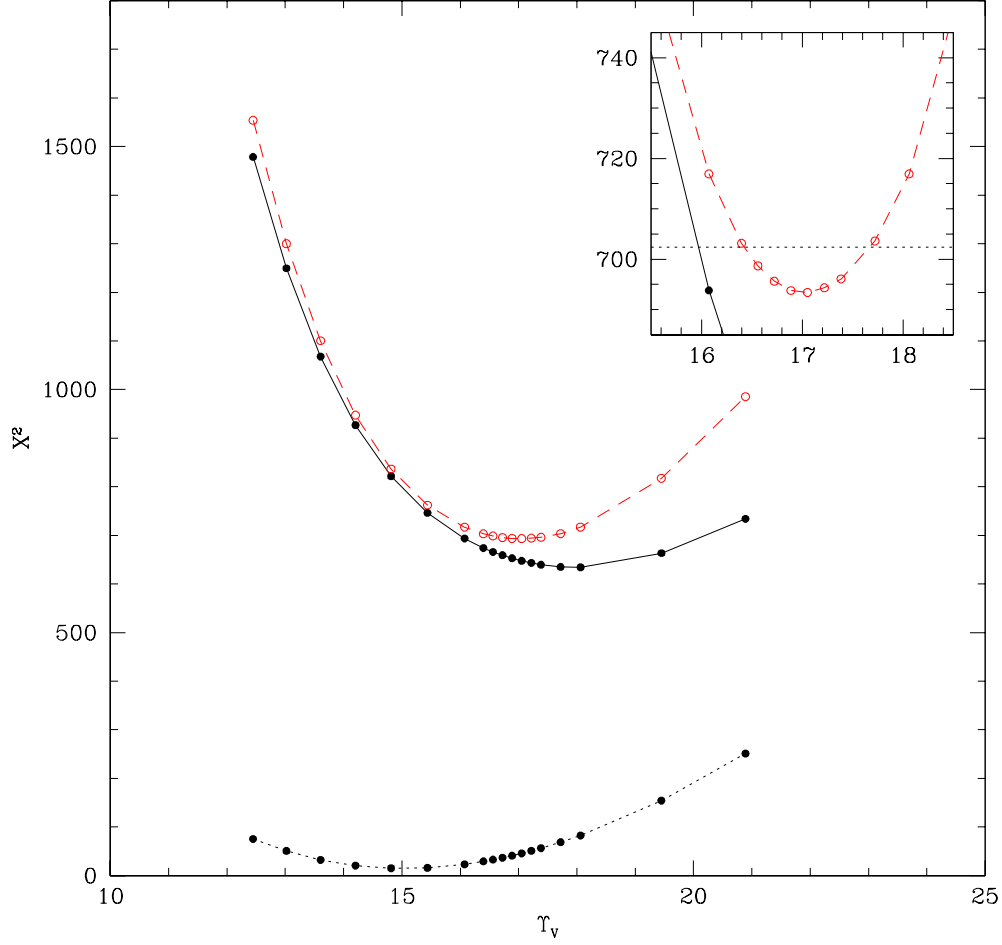


FIGURE 6.8— Same as Figure 6.7, but for the model with the logarithmic potential.

then perform a NNLS fit for each new model and compute the χ^2 difference with the data. In this way, we can study the stellar χ^2 distribution, χ_{stars}^2 , as a function of Υ_V . A similar χ^2 comparison is done with the gas data: After rescaling a model by Υ_V , we ask by how much the new circular velocity (rescaled by $\sqrt{\Upsilon_V}$) differ from the one derived from the gas measurements (see Section 6.1.3), yielding a distribution χ_{gas}^2 . These two distributions can be combined to get better constraints on Υ_V . The combined probability distribution is $\mathcal{P}(\Upsilon_V) \sim \exp[-(\chi_{\text{stars}}^2 + \chi_{\text{gas}}^2)]$ (Press *et al.* 1992).

For a good fit, one expects χ^2 to be roughly equal to the number of data points, which was not the case in the stellar fits, probably due to systematic discrepancies in the data from different runs (Section 6.1.2). Following Kochanek (1994), we rescale the χ_{stars}^2 distribution such that $\chi_{\text{stars, min}}^2 = 632$. Similarly, we consider the 15 gas data points with $|x'| > 3''$ for which the gas rotation curve is roughly flat and rescale the χ_{gas}^2 distribution, so that $\chi_{\text{gas, min}}^2 = 15$. In Figure 6.7, we plot χ_{gas}^2 , χ_{stars}^2 , and the combined χ_{comb}^2 distribution as function of Υ_V for the models in which the mass follows the light. The minimum $\chi_{\text{min, comb}}^2$ is attained for $\Upsilon_V = 15.0 \pm 0.6$, where the error bar corresponds to the formal 99.73 % confidence level or 3σ interval

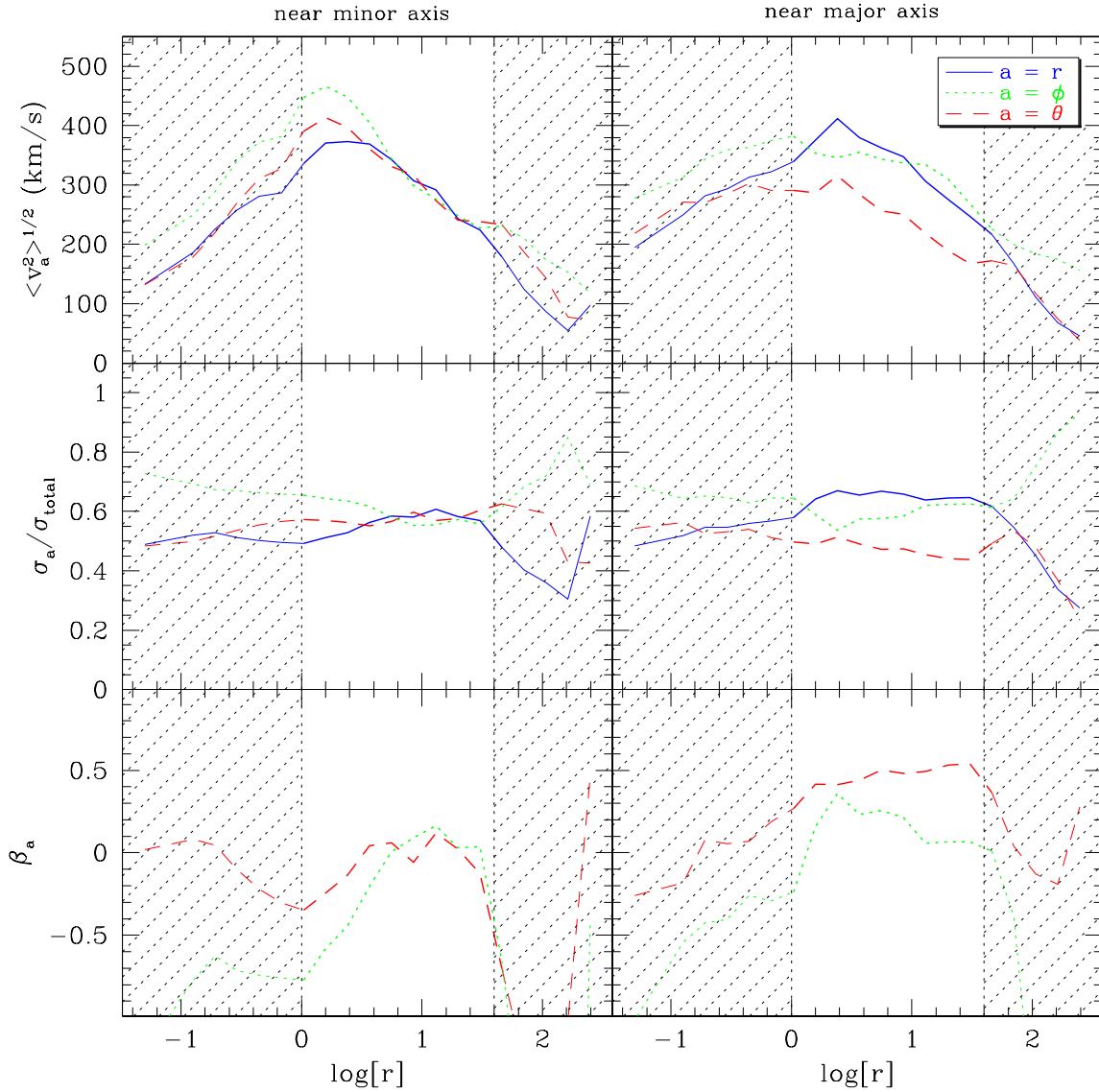


FIGURE 6.9— Anisotropy profiles of the model with $\Upsilon_V = 15$. The first column is an average around the symmetry axis, and the second one around the equatorial plane. The first row shows the second moments in km/s, the second one the ratio of the dispersions to the total dispersion, and the last one shows the anisotropy parameter defined as $\beta_a = 1 - \sigma_a^2/\sigma_r^2$. The inset in the upper right panel explains the line convention.

(i.e., the range of Υ_V for which $\chi_{\text{comb}}^2 = \chi_{\text{min}}^2 + 9$). We perform the same exercise for the model with the logarithmic potential. In that case, Υ_V is defined as the mass enclosed within $15''$ divided by the total amount of light in the same volume. We find $\Upsilon_V = 17.05 \pm 0.7$ at the same confidence level (Figure 6.8).

In order to compare with the sample of 37 bright elliptical studied by van der Marel (1991), we first calculate the total absolute magnitude of NGC 2320 in B using his choice for H_0 , and find $M_B = -23 h_{50}$. We then translate our value of Υ_V (15) in the V-band into the R-band to place NGC 2320 on van der Marel's Figure 6 (upper left panel). Using V-R of 0.77 (typical for K0 giants) we find $\Upsilon_R = 8.0$. This value is 1.8 times higher than the prediction of his least-square

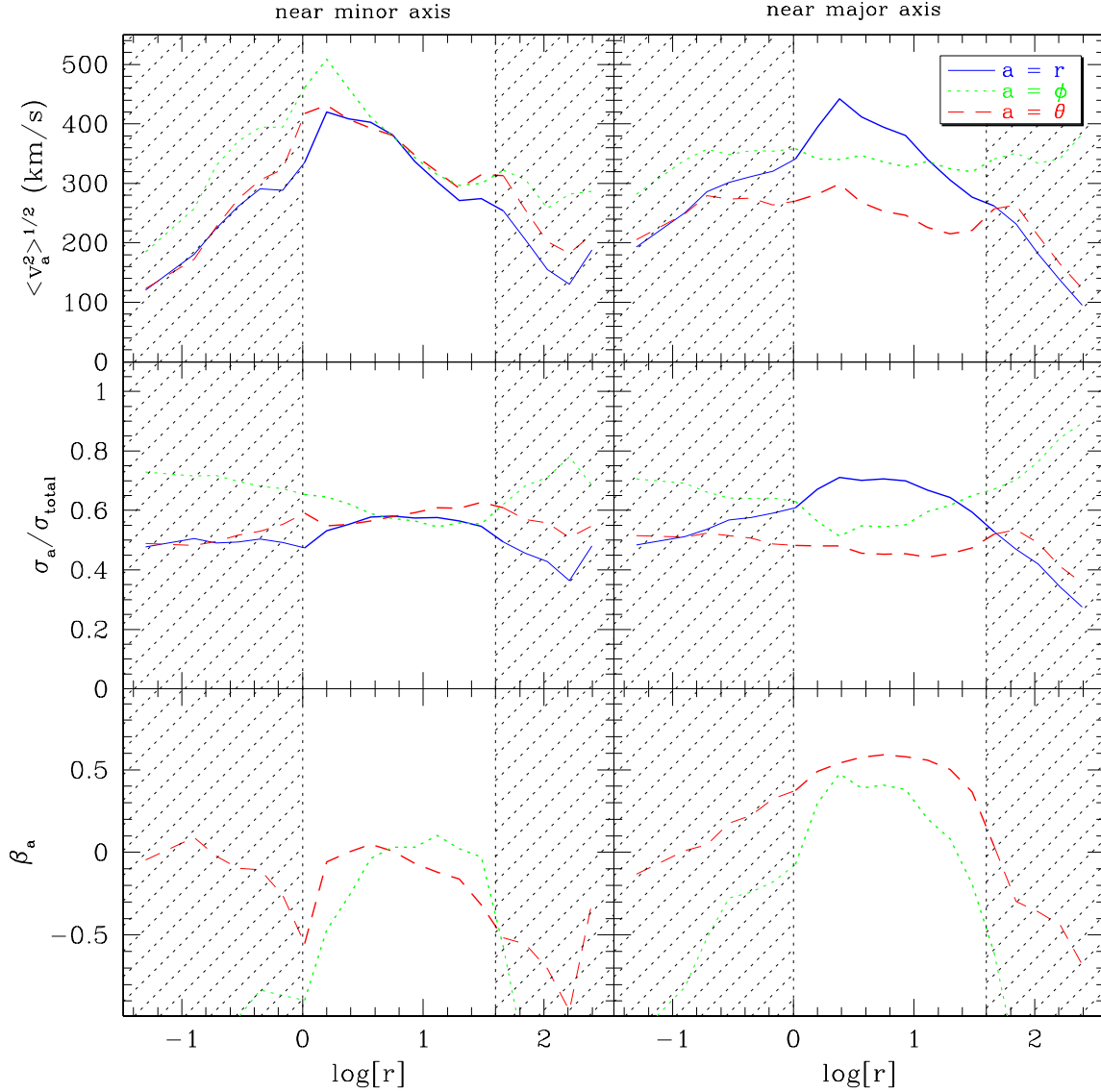


FIGURE 6.10— Same as Figure 6.9, but for the model with the logarithmic potential.

fit relation at that absolute magnitude, making NGC 2320 a slight outlier.

6.3.3 Intrinsic velocity dispersions

Using solutions with a moderate amount of regularization (as defined in Section 5.4.5 of chapter 5), we compute second moments, ratio of dispersion profiles and the usual anisotropy parameter β (see e.g., Binney & Tremaine, 1987) along major and minor axis for the best fit models. In Figure 6.9 we show such quantities for the best fit model with a radially constant Υ_V of 15. The results are averaged over cells with the same radius but with a different angle θ , which is the usual spherical coordinate: In the meridional plane, the first three angular sectors closest to the minor axis are averaged together (left column of Figure 6.9), as well as the remaining four closest to the major axis (right column). The first row displays the second moments as a function of radius. The ratios of the dispersions to the total dispersion are shown in the second

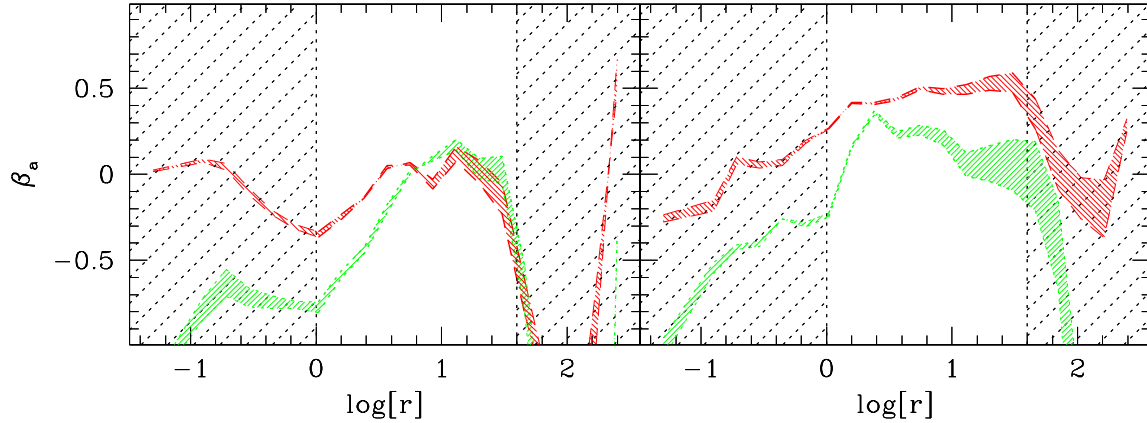


FIGURE 6.11— The anisotropy parameter as in Figure 6.9: the densely shaded region indicates the values of β for models which lie in the formal 99.73 % confidence level region around the best fitting model (see Figure 6.7).

row. In the last row, we plot the anisotropy parameter $\beta_a = 1 - \sigma_a^2 / \sigma_r^2$ for $a = \phi$ (dotted line) or $a = \theta$ (dashed line). Positive β corresponds to radial anisotropy. Figure 6.10 shows the same, but for the model with the logarithmic potential. Both types of models (with and without DM) have a similar general behavior: they are dominated by radial anisotropy for most radii meaningfully constrained by the data. The radii outside the range of the kinematic data are indicated with shades in Figures 6.9 and 6.10. On the true major axis, $\beta_\theta \simeq 0.4 - 0.5$ and $\beta_\phi \simeq 0.0 - 0.3$, whereas $\beta_\theta \sim \beta_\phi \sim 0.0$ on the symmetry axis (except in the central $3''$).

In Figure 6.11 and 6.12, we concentrate on the β parameters. We consider two additional models around the best fit to estimate the difference (in terms of anisotropy) between models that are statistically acceptable. As mentioned in Section 6.3.2, the two limiting models have $\chi^2 = \chi_{\min}^2 + 9$, corresponding to the 99.73 % confidence level. We compute the β -profiles of these two models and shade the region in between them (Figure 6.11 and Figure 6.12). In this way, we can estimate an error bar for the quantity β . All models in the 99.73 % confidence interval follow the same trend: they are radially anisotropic on the major axis and more isotropic on the minor axis.

6.4 Conclusions and Discussion

We have constrained the anisotropy of NGC 2320 using a combination of kinematic data: stellar VPs along multiple position angles (see Table 6.1), and mean velocity and velocity dispersion of the emission line gas along the major axis. The dense coverage of the galaxy by the various long slits is crucial to constrain the dynamical model. The gaseous data provides an additional constraint on the normalization of the potential (i.e., on the total mass).

The dynamical models are constructed following our extension of Schwarzschild's orbit superposition technique described in chapter 3. The best fitting model is dominated by radial anisotropy near the equatorial plane and is more isotropic on the symmetry axis. This conclusion is valid for models in which the mass is proportional to the light and also for models with a flat rotation curve at large radii (logarithmic potential). We find a best fit Υ_V of 15.05 ± 0.6 for the radially constant Υ_V model and 17.0 ± 0.7 for the logarithmic one. We study the dispersion profiles of models for these intervals of Υ_V in order to determine the range of anisotropy cov-

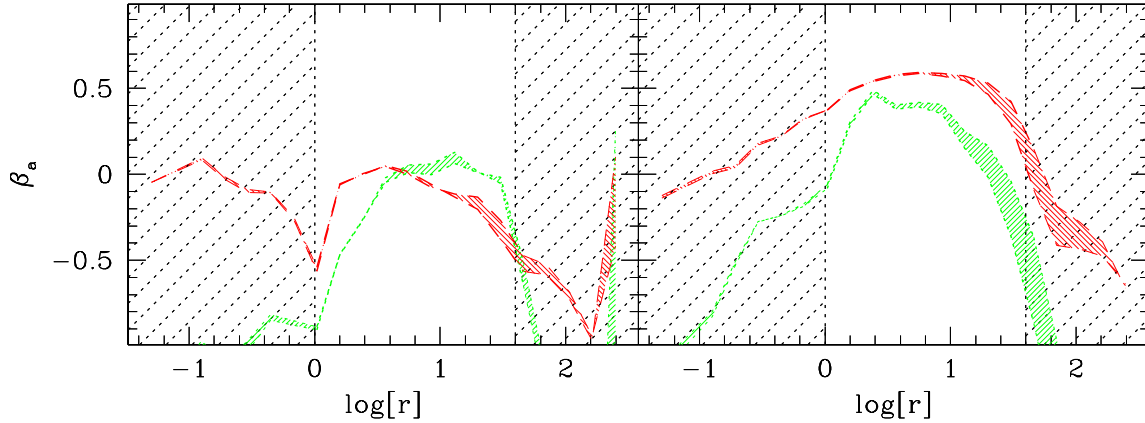


FIGURE 6.12— Same as Figure 6.11, but for the logarithmic potential mass model.

ered by statistically acceptable models. This range is small in general and most models in the Y_V interval follow the same trend (i.e., radial anisotropy in the equatorial plane, isotropy near the symmetry axis).

Radial anisotropy has also been found in a few other objects (NGC 2434, NGC 1600, NGC 6703, NGC 1399, NGC 3377, NGC 3379, NGC 4473, NGC 5845 and M87). It seems worth comparing to the dynamical structure of merger remnants produced by N-body simulations. As pointed out by Dubinski (1998), merger simulations leading to the formation of an elliptical galaxy have concentrated on mergers of a pair of disk galaxies (Barnes 1992; Barnes & Hernquist 1996), of several disk galaxies (Barnes 1989; Weil & Hernquist 1996) or of small virialized clusters of spherical galaxies (Funato, Makino & Ebisuzaki 1993; Garijo, Athanassoula & Garcia-Gomez 1997). For these reasons, Dubinski carried out more realistic cosmological simulations. He found that the anisotropy of the brightest cluster galaxy grows gently from $\beta = 0.0$ near the center to $\beta = 0.5$ at $3 R_{\text{eff}}$. Inside one R_{eff} , his N-body model is mildly radially anisotropic ($\beta < 0.3$). This is qualitatively consistent with our results for NGC 2320.

A comparison of NGC 2320 with Barnes' results is more complicated: In his collisionless simulations, Barnes (1992) typically finds *triaxial* merger remnants with a large fraction of box orbits (see e.g., his figure 20). These box orbits do not exist in the axisymmetric case, only the tubes are present. But if NGC 2320 is indeed triaxial, the box orbits may be responsible for the radial anisotropy. Since our models are axisymmetric, they have no other choice than to put a lot of weight on low L_z tubes to produce radial anisotropy. This is not the case for Barnes' models: the box orbits (with $L_z = 0$) are responsible for the radial anisotropy whereas his population of tube orbits seems fairly uniform in L_z (see his Figures 20b, 21b and 22b).

It is tempting to envision (axisymmetric) radially anisotropic models as previously triaxial models that evolved towards axisymmetry: in this picture, the box orbits are the progenitors of the radial tube orbits (responsible for the radial anisotropy in axisymmetric systems). Black holes, steep central density cusps and central gas concentrations can efficiently scatter stars on box orbits and destroy the triaxiality (see e.g., Gerhard & Binney 1985; Barnes & Hernquist 1996; Merritt & Quinlan 1998). If the amount of gas accumulated at the center is large, one can imagine that box orbits are so efficiently destroyed that very little radial anisotropy is left in the distribution of tube orbits. Therefore, in this scenario, the prevalence of radial orbits in NGC 2320 argues against a merger in which large amounts of gas flowed to the center.

Acknowledgments

We thank Eric Emsellem for discussions and help with the MGE software and Roeland van der Marel for a careful reading of the manuscript. N.C. acknowledges the hospitality of the Max-Planck Institut für Astronomie, where most of this work was done.

References

- Barnes, J. E. 1992, *ApJ*, 393, 484
- Barnes, J. E., & Hernquist, L. 1996, *ApJ*, 471, 115
- Binney, J. J., & Tremaine S. D. 1987, *Galactic Dynamics* (Princeton: Princeton University Press)
- Carollo, C. M., de Zeeuw, P. T., van der Marel, R. P., Danziger, I. J., & Qian, E. E. 1995, *ApJ*, 441, L25
- Cinzano, P., van der Marel, R. P. 1994, *MNRAS*, 270, 325
- Cretton, N., de Zeeuw, P. T., van der Marel, R. P., & Rix, H.-W. 1999, *ApJ Supplements*, in press (astro-ph/9902034), chapter 3 of this thesis
- Cretton, N. & van den Bosch, F. C. 1999, *ApJ*, 514, 704, chapter 5 of this thesis
- Dejonghe, H. 1987, *MNRAS*, 224, 13
- Dubinski, J. 1998, *ApJ*, 502, 141
- Emsellem, E., Monnet, G., & Bacon, R. 1994, *A&A*, 285, 723 (EMB94)
- Funato, Y., Makino, J., & Ebisuzaki, T. 1993, *PASJ*, 45, 289
- Garijo, A., Athanassoula, E., & Garcia-Gomez, C. 1997, *A&A*, 327, 930
- Gebhardt, K., *et al.* 1999, *AJ*, submitted
- Gerhard, O. E., Binney, J. J. 1985, *MNRAS*, 216, 467
- Gerhard, O. E. 1993, *MNRAS*, 265, 213
- Kochanek, C. S. 1994, *ApJ*, 436, 56
- Lawson, C. L., & Hanson, R. J. 1974, *Solving Least Squares Problems* (Englewood Cliffs, New Jersey: Prentice-Hall)
- Merritt, D. 1993, *ApJ*, 413, 79
- Merritt, D., & Oh, S. P. 1997, *AJ*, 113, 1279
- Merritt, D., & Quinlan, G. D. 1998, *ApJ*, 498, 625
- Navarro, J., Frenk, C., & White, S. D. M. 1996, *ApJ*, 462, 563
- Neistein, E., Maoz, D., Rix, H.-W., Tonry, J. L. 1999, preprint (astro-ph/9903007)
- Prugniel, Ph., Heraudeau, Ph. 1998, *AAS*, 128, 299
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, *Numerical Recipes* (Cambridge: Cambridge University Press)
- Richstone, D. O, et al. 1997, in *The Nature of Elliptical Galaxies*, Proceedings of the Second Stromlo Symposium, eds. Arnaboldi, M., da Costa, G., & Saha, P., p. 123
- Rix, H.-W., & White, S. D. M. 1992, *MNRAS*, 254, 389
- Rix, H.-W., Kennicutt, R. C., Braun, R., Walterbos, R. A. M. 1995, *ApJ*, 438, 155
- Rix, H.-W., de Zeeuw, P. T., Cretton, N., van der Marel, R. P., & Carollo, C. M. 1997, *ApJ*, 488, 702, chapter 2 of this thesis
- van den Bosch, F. C., & Emsellem E. 1998, *MNRAS*, 298, 267
- van der Marel, R. P. 1991, *MNRAS*, 253, 710
- van der Marel, R. P., & Franx, M. 1993, *ApJ*, 407, 525

Nederlandse samenvatting

DIT proefschrift bevat een studie naar de eigenschappen van vroeg-type (elliptische en lensvormige) sterrenstelsels met behulp van gedetailleerde dynamische modellen. Deze modellen zijn gebaseerd op de veronderstelling dat een sterrenstelsel eenvoudiger te beschrijven is in de vorm van een verzameling van banen dan als een verzameling van sterren. Dit idee wordt in dit proefschrift gebruikt om dynamische modellen voor enkele vroeg-type sterrenstelsels te bouwen. Daartoe wordt een grote verzameling sterbanen berekend, en wordt het aantal sterren in elke baan zodanig aangepast dat waargenomen eigenschappen (de licht- en snelheidsverdeling) worden gereproduceerd. Het uiteindelijke 'baanmodel' is bedoeld om een zo getrouw mogelijke representatie van het waargenomen sterrenstelsel te zijn. Het voordeel van deze aanpak (en zijn voornaamste drijfveer) is dat er van te voren geen vereenvoudigende aannamen worden gemaakt wat betreft de dynamische structuur van het sterrenstelsel. Dit is essentieel bij het zoeken naar zware zwarte gaten in de kernen van sterrenstelsels, omdat een vereenvoudigd model onjuiste resultaten kan geven.

Hoge kwaliteit gegevens, zowel verkregen vanaf Aarde als vanuit de ruimte met de Hubble Space Telescope, worden in dit proefschrift geïnterpreteerd met behulp van zulke baanmodellen (hoofdstuk 3). Deze gegevens, in combinatie met de gepresenteerde modellen, leveren overtuigend bewijs voor de aanwezigheid van *donkere materie* (zie onder). In het bijzonder worden de centrale concentraties van donkere materie in M32 en NGC 4342 bestudeerd. Zwarte gaten vormen hiervoor de meest waarschijnlijke verklaring (hoofdstuk 4 en 5). De waargenomen snelheden in NGC 2434 laten zien dat dit sterrenstelsel omgeven moet zijn door een zware donkere halo die een aanzienlijk deel van de totale massa bevat (hoofdstuk 2). Tenslotte wordt de interne dynamica van NGC 2320 behandeld. Hieruit blijkt dat de sterren voornamelijk in radiële banen bewegen (hoofdstuk 6).

Sterrenstelsels

Sterrenstelsels zijn reusachtige verzamelingen van sterren, gas, stof en planeten. Grote exemplaren bevatten ongeveer honderdmiljard sterren. De hoeveelheid gas en stof varieert van een paar procent van de totale sterrenmassa tot tien procent voor de meest gasrijke objecten. Hoewel de totale massa van de planeten volkomen onbeduidend is vergeleken met de stellaire en gasmassa's, beschouwt de mensheid deze objecten als zeer belangrijk!

Sterrenstelsels kunnen aan de hand van hun schijnbare vorm grofweg in drie categorieën worden onderverdeeld: schijfvormige, elliptische en onregelmatige (zie figuur 6.13). Schijfvormige stelsels zijn erg platte objecten, soms met prachtige spiraalarmen en een bolvormig centraal gedeelte, ook wel 'lens' genoemd. De spiraalarmen kunnen in meer of mindere mate rond de lens gewikkeld zijn, die op haar beurt verschillende groottes kan hebben. Deze kenmerken laten een verdere klassificatie van de verschillende typen spiraalstelsels toe. Omdat de Zon

zich *in de schijf* van ons sterrenstelsel bevindt, zijn er langs een gezichtslijn in het vlak van de schijf meer sterren zichtbaar dan loodrecht op het vlak. Dit verklaart de verschijningsvorm van ons sterrenstelsel aan de nachthemel en zijn naam, de 'Melkweg'. Lensvormige sterrenstelsels bestaan uit schijven zonder spiraalarmen en met een grote centrale lens. Elliptische sterrenstelsels, ook wel vroeg-type sterrenstelsels genaamd, zien er rond uit (grofweg als een ei) en bevatten veel minder gas dan spiraalstelsels. Hun voorkomen is gladder, wat ertoe leidde dat sterrenkundigen ervan uit gingen dat deze objecten een eenvoudige dynamische structuur hebben. Tenslotte omvat de laatste categorie (de onregelmatige sterrenstelsels) alle sterrenstelsels die niet in de andere twee categorieën konden worden opgenomen vanwege hun verstoorde vorm.

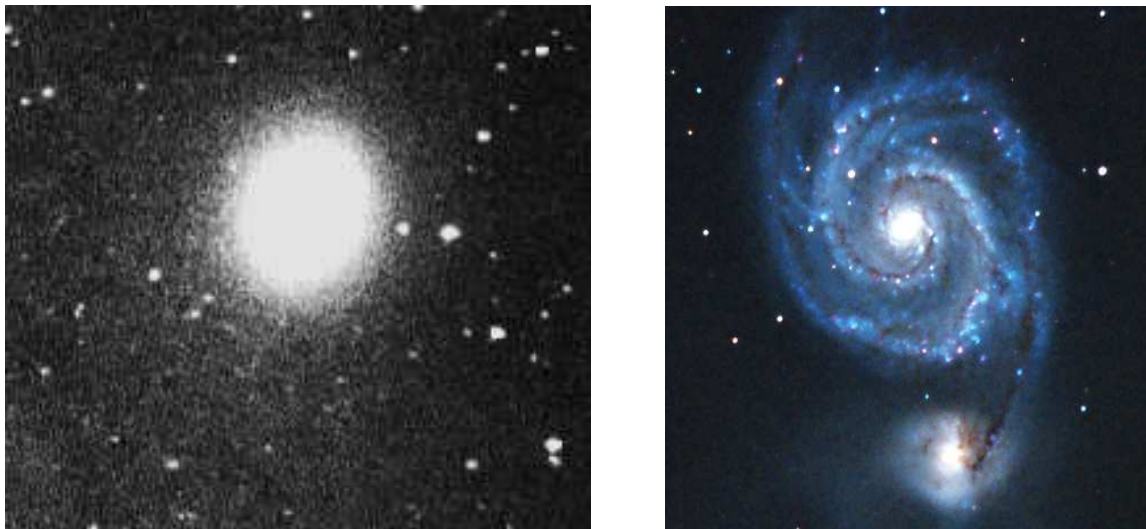
De overheersende gedachte is dat sterrenstelsels vroeg in de geschiedenis van het Heelal ontstaan zijn, ongeveer 15 miljard jaar geleden. Sterrenkundigen proberen te begrijpen hoe zij gevormd zijn door *vandaag* naar ze te kijken, dat wil zeggen na 15 miljard jaar evolutie. Met deze evolutie moet rekening worden gehouden in scenario's van de vorming van sterrenstelsels. De eindprodukten van deze scenario's moeten er uit zien als hedendaagse sterrenstelsels. Dit is te vergelijken met het reconstrueren van de verhaallijn van een drama waarvan alleen de afloop bekend is! De eindige lichtsnelheid biedt een andere mogelijkheid om de geschiedenis van sterrenstelsels te ontrafelen. Het licht heeft een seconde nodig om van de Maan naar de Aarde en acht minuten om van de Zon naar de Aarde te reizen. Met andere woorden: het waargenomen beeld van de Zon is dat van acht minuten geleden. Voor sterrenstelsels die veel verder weg staan kan deze vertraging oplopen tot enkele miljarden jaren. Ver weggelegen sterrenstelsels worden dus nu waargenomen zoals ze er miljarden jaren geleden uitzagen, vlak na hun ontstaan.

Sterrenkunde is een door waarnemingen gedreven wetenschap: theorieën en modellen worden gebruikt om vragen te beantwoorden die door waarnemingen worden opgeroepen. Op deze wijze wordt getracht om met behulp van alle waarnemingen één coherent beeld te vormen. In dit proefschrift wordt zo'n beeld van de ruimtelijke en snelheidsverdeling van sterren in sterrenstelsels beschreven. Hierbij worden de waargenomen lichtverdeling van sterrenstelsels aan de hemel (zie bijvoorbeeld figuur 6.13) en de snelheden van sterren langs de gezichtslijn gebruikt. Het is in het algemeen niet mogelijk individuele sterren in sterrenstelsels te zien. Een waarneming geeft slechts de gemiddelde eigenschappen van alle sterren langs de gezichtslijn. Het is dus niet mogelijk individuele snelheden van sterren te meten. Wel kan de snelheidsverdeling langs de gezichtslijn, het zogenaamde snelheidsprofiel, worden bepalen. Het snelheidsprofiel geeft in principe weer hoeveel sterren er met welke snelheid bewegen.

Enkele van de grote vragen over sterrenstelsels zijn: hoe zijn ze ontstaan, hoe zitten ze in elkaar en hoe ontwikkelen ze zich in de toekomst? Een eerste stap in de zoektocht naar antwoorden op deze vragen is het construeren van dynamische modellen van deze fascinerende objecten. Dit geeft een beschrijving van de posities en de bewegingen van de sterren in sterrenstelsels. Om het model te vinden dat het best bij de waarnemingen past worden verschillende modellen hiermee vergeleken.

Dynamische modellen

Sterrenstelsels zijn geen statische objecten: ze zijn opgebouwd uit miljarden sterren die allemaal onder invloed van hun onderlinge zwaartekracht bewegen. Aangezien sterrenstelsels zeer oud zijn, hebben de sterren voldoende tijd gehad om een stabiele evenwichtssituatie te bereiken. Dit evenwicht wordt soms verstoord door een wisselwerking met een ander sterrenstelsel. Afhankelijk van de sterkte van deze wisselwerking kunnen de sterrenstelsels in meer of mindere mate



FIGUUR 6.13— In het linker paneel is het elliptische stelsel M32 afgebeeld. M32 is de begeleider van het grote spiraalstelsel Andromeda (zie omslag) en wordt bestudeerd in hoofdstuk 4 van dit proefschrift. In het rechter paneel zijn het spiraalstelsel M51 (ook wel de ‘draaikolknevel’ genaamd) en zijn begeleider NGC 5194 afgebeeld (onder). De spiraalarmen van M51 zijn duidelijk zichtbaar.

van vorm veranderen. Soms kan de botsing van twee spiraalstelsels leiden tot de vorming van een elliptisch sterrenstelsel. Kleine sterrenstelsels kunnen zelfs geheel opgeslokt worden door grote sterrenstelsels. In dit proefschrift worden de effecten van wisselwerkende sterrenstelsels echter genegeerd, en wordt een gedetailleerde beschrijving van enkele vroeg-type objecten gegeven. Hierbij worden *dynamische modellen* gebruikt met als doel de bewegingen van sterren in deze sterrenstelsels te verklaren. Een dynamisch model is een wiskundige beschrijving van een sterrenstelsel wat door de astrofysicus als een gereedschap gebruikt wordt. Zulke modellen worden gebruikt omdat de praktische implementatie van Newton’s zwaartekrachtswet voor honderden miljarden wisselwerkende sterren onmogelijk is. Gebaseerd op een paar individuele objecten wordt geprobeerd de fundamentele wetten te vinden die het leven van sterrenstelsels bepalen. In dit proefschrift wordt alleen aandacht besteed aan de bewegingen van sterren in vroeg-type sterrenstelsels, en worden hun andere eigenschappen genegeerd: sterren worden beschouwd als puntmassa’s en bewegen in banen die bepaald worden door de zwaartekracht van alle aanwezige massa in het stelsel (voornamelijk de sterren en het gas).

Het construeren van een dynamisch model is gecompliceerder voor een elliptisch sterrenstelsel dan voor een spiraalstelsel. In de schijven van spiraalstelsels bewegen sterren voornamelijk in (bijna) cirkelvormige banen. Daarentegen zijn de bewegingen van sterren in elliptische stelsels niet zo eenvoudig, en kunnen de banen veel verschillende vormen hebben. Sommige banen zijn zelfs chaotisch. Zonder voorkennis over de precieze vorm van sterbanen in elliptische stelsels wordt een grote verzameling banen berekend. Hierna wordt bepaald hoeveel sterren in elke baan bewegen, onder de voorwaarde dat het uiteindelijke model een goede weergave geeft van het waargenomen stelsel, zowel wat betreft de vorm als de snelheidsprofielen. Het uiteindelijke model stelt ons in staat de volledig drie-dimensionale snelheids- en positieverdeling te bestuderen. In dit proefschrift worden zulke baanmodellen gemaakt voor enkele vroeg-type sterrenstelsels: NGC 2434 (hoofdstuk 2), M32 (hoofdstuk 4), NGC 4342 (hoofdstuk 5) en NGC 2320 (hoofdstuk 6).

Donkere materie

De snelheid van een ster onder invloed van de zwaartekracht wordt bepaald door de totale hoeveelheid aanwezige massa. Massa en snelheid zijn dus nauw verbonden: hoe groter de massa, hoe groter de snelheid. De massa van een sterrenstelsel kan dus bepaald worden door snelheden van sterren te meten. Een andere manier om de massa te bepalen is de waargenomen hoeveelheid sterlicht om te zetten in massa (de omzettingfactor van sterlicht naar sterrenmassa is immers bekend).

Voor spiraalstelsels geven deze twee methoden significant verschillende resultaten. De waargenomen snelheden van sterren in de buitendelen van spiraalstelsels zijn te groot om enkel veroorzaakt te kunnen worden door de zichtbare massa. Deze is slechts tien procent van de massa, die vereist is om de bewegingen te verklaren. Een substantieel deel van de totale massa van spiraalstelsels is dus *onzichtbaar*. Deze niet-zichtbare massa wordt *donkere materie* genoemd.

In dit proefschrift worden twee vormen van donkere materie in vroeg-type sterrenstelsels bestudeerd, namelijk zwarte gaten in de centrale delen en donkere halo's in de buitendelen.

Zwarte gaten

Quasars zijn extreem heldere, ver weggelegen en dus jonge sterrenstelsels. Zij stralen veel meer energie uit dan enkel op grond van hun sterren verwacht zou kunnen worden. De meest aannemelijke verklaring voor dit overschot aan uitgezonden energie is de aanwezigheid van een zwart gat in hun kern. Zulke zwarte gaten kunnen zo zwaar zijn als enkele honderden miljoenen zonsmassa's. Zwarte gaten zijn objecten waarvan de zwaartekracht zo sterk is dat zelfs licht niet kan ontsnappen. Invallende sterren en gas- en stofwolken worden opgeslokt door het zwarte gat. Bij dit proces komt een enorme hoeveelheid energie vrij, wat een aannemelijke verklaring vormt voor de waargenomen eigenschappen van quasars.

Het feit dat jonge sterrenstelsels zware zwarte gaten in hun kernen hebben, doet vermoeden dat ook in de hedendaagse, oude sterrenstelsels zwarte gaten aanwezig zijn, ook al stralen deze stelsels bij lange na niet dezelfde hoeveelheid energie uit als quasars. Zoals hierboven al besproken is, zijn de bewegingen van sterren gekoppeld aan massa. Zodoende voorspelt de aanwezigheid van zware zwarte gaten in de kernen van sterrenstelsels hoge snelheden van sterren in de nabijheid van het zwarte gat. Zulke hoge snelheden zijn inderdaad waargenomen in de kernen van vele elliptische sterrenstelsels, wat ook hier de aanwezigheid van een zwarte gat aannemelijk maakt.

In dit proefschrift worden de centrale delen van twee vroeg-type sterrenstelsels, M32 en NGC 4342 (hoofdstuk 4 en 5), nauwkeurig bestudeerd met behulp van dynamische baanmodellen. Modellen zonder een zwart gat zijn niet in staat om de waargenomen snelheden te reproduceren: sterren in zulke modellen bewegen niet zo snel als geobserveerd. De aanwezigheid van een zwart gat in de kern van beide stelsels is de meest aannemelijke verklaring.

Donkere halo's

Uit de beweging van sterren en gas in de buitendelen van spiraalstelsels kan worden afgeleid dat er donkere materie in de vorm van een zware donkere halo aanwezig is (zie boven). In tegenstelling tot spiraalstelsels is er bijna niets bekend over donkere halo's van elliptische stelsels. Dit wordt gedeeltelijk veroorzaakt door een gebrek aan bruikbare objecten om snelheden in de buitendelen van elliptische stelsels te meten: er zijn te weinig sterren en er is te weinig gas. Bovendien kan de snelheidsstructuur die wordt veroorzaakt door een donkere halo evengoed worden gegenereerd door de ingewikkelde baanstructuur van deze stelsels, zodat het moeilijk is onderscheid te maken tussen deze twee oorzaken.

Dynamische baanmodellen gecombineerd met nauwkeurig geanalyseerde kinematische waarnemingen van het ronde elliptische sterrenstelsel NGC 2434 laten zien dat de zichtbare massa onvoldoende is om de bewegingen van de sterren te verklaren (hoofdstuk 2). Het toevoegen van een donkere halo in de baanmodellen leidt tot resultaten die wel overeenkomen met de waarnemingen. De donkere halo rond NGC 2434 is zwaarder dan de typische donkere halo van een spiraalstelsel. Voor het afgeplatte elliptische stelsel NGC 2320 wordt een soortgelijk analyse uitgevoerd, waarbij naast de bewegingen van de sterren ook de snelheden van het aanwezige gas worden gebruikt (hoofdstuk 6). Ook dit stelsel heeft een donkere halo; de beweging van de sterren is voornamelijk in de radiële richting.

English Summary

THIS thesis concerns the study of early-type (elliptical and lenticular) galaxies through detailed dynamical modeling. The models are based on the idea that a galaxy is a dynamical object that can be better understood as a collection of *orbits* rather than a collection of *stars*. In this thesis, we exploit this idea and construct dynamical models for a few early-type galaxies: we compute numerically a large set of orbits and adjust the number of stars on each of them so as to match a set of observed properties (light distribution and velocities). The final 'orbit-model' is aimed to be as close as possible to the observed galaxy. The advantage of this approach (and its primary motivation) is that no a priori simplifying assumptions are made concerning the dynamical structure of the galaxy. This is crucial in the search for massive black holes (BHs) in the center of galaxies, because a simplified model can give a wrong answer concerning the existence of the BH.

In this thesis, such orbit-models (chapter 3) are used to interpret high quality ground-based and Hubble Space Telescope observations. This data in combination with the models provide compelling evidence for *dark matter* (see below). In particular, we study central dark mass concentrations in M32 and NGC 4342, for which BHs are the most likely explanations (chapters 4 and 5). In order to explain the observed velocities of NGC 2434, we show that this galaxy has to be embedded in a massive dark halo that accounts for a substantial fraction of the total mass (chapter 2). Finally, we focus on the internal dynamics of NGC 2320 and show that the motions of the stars are predominantly radial (chapter 6).

Galaxies

Galaxies are huge collections of stars, gas, dust and planets. Large galaxies contain approximately one hundred billion (10^{11}) stars. The quantity of gas and dust varies from a few percent of the total stellar mass (for the lenticular galaxies), to ten percent for the most gas-rich objects (for the spiral galaxies). Even if the total mass of planets is completely insignificant compared to the stellar and gas mass, we humans consider these objects as quite important!

Galaxies can be classified in roughly three categories according to their apparent shape: disk, elliptical (see Figure 6.14) and irregular. Each class is further subdivided according to the appearance of the galaxy. Disk galaxies are very flat objects, sometimes with beautiful spiral arms and a rounder central part, called the 'bulge'. The spiral arms can be more or less tightly wound around the bulge, which in turn can have different sizes: these features allow a classification of the various types of spiral disk galaxies. Our own galaxy is a disk galaxy and it can be recognized as the white band that crosses the sky. Since our Sun is located *in the disk* of our galaxy, when we look through it, we see a lot more stars than looking perpendicular from it: this explains the appearance of our galaxy at night and its name, the 'Milky Way'. The lenticular galaxies are disks without spiral arms and with a large central bulge. Elliptical galaxies,

also called 'early-type' galaxies, have a rounder overall shape (roughly like an egg) and have much less gas than the spirals. Their appearance is smoother, which led astronomers to think that these objects are simpler, dynamically speaking. Finally, the last class (irregular) includes all the other galaxies that could not be classified in the first two categories because of their disturbed appearance.

Galaxies are believed to form early on in the history of the universe, approximately 15 billion years ago. Astronomers try to understand how they have formed by looking at them *today*, i.e., after ~ 15 billions years of evolution. All this evolution has to be taken into account in galaxy formation scenarios and the end-products of such scenarios should look like present day galaxies. It is like trying to reconstruct the premises of the drama, knowing already the end! There is another option to reveal the history of galaxies, i.e., to look at them in the past. Because the speed of light is finite, when we look at objects we do not see them as they are now, but as they were at the time when the light was emitted (or reflected) from the object. If the distance between the object and the observer is large, the travel time of the light can be long. For instance, it takes about one second for the light to travel the distance Moon–Earth, and eight minutes for the distance Sun–Earth. In other words, we see the Sun as it was eight minutes ago. If we look at very distant galaxies, we see them as they were a long time ago, in some cases, billions of years ago.

Astronomy is an observational science: observations provide facts and trigger questions, whereas theories and models try to answer these questions and put the facts into a coherent picture. We can measure the intensity of stellar light of a galaxy, e.g., as a function of the distance from the center, and the velocity of stars along the line-of-sight. The other components of the velocity can not be measured, except for very nearby objects. Since individual stars can not be resolved in external galaxies, one can not measure individual stellar velocities. In projection, the light of many stars contributes to the same position on the galaxy image. Therefore, we do not measure a single line-of-sight velocity v_{los} , but a *distribution* of velocities. In other words, from the observations, we can reconstruct the number of stars that have a velocity v_{los} , and hence build the whole distribution as a function of v_{los} . We call such a distribution 'Velocity Profile' (hereafter VP).

We want to understand what these fascinating objects are, how they function, how they were formed and into what they are evolving. Constructing dynamical models for galaxies is a first step to understand how galaxies "work": it gives a description of the position and motion of the stars inside the galaxy. Different models can be compared with observations to select which ones provide the best match to the data.

Dynamical models

Galaxies are not static objects: they are composed of billions of stars that move under their mutual gravity. Since galaxies are very old, the stars in them had time to settle down and reach an equilibrium situation. Sometimes this internal equilibrium can be disturbed by the interaction with another galaxy. Depending on the strength of this interaction, the galaxy can modify its shape more or less drastically. In extreme cases, a collision between two spirals can lead to the formation of an elliptical. Small galaxies can even be cannibalized by bigger ones. However in this work, we ignore the effect of interactions between galaxies and concentrate on the detailed description of a few early-type objects. We construct *dynamical models* to explain the internal motions of the stars in isolated galaxies. We rely on models, because a physical theory describing the motion of hundred billion stars is out of reach: even though Newton's law of gravity is simple, a frightening complexity arises when many stars interact together through

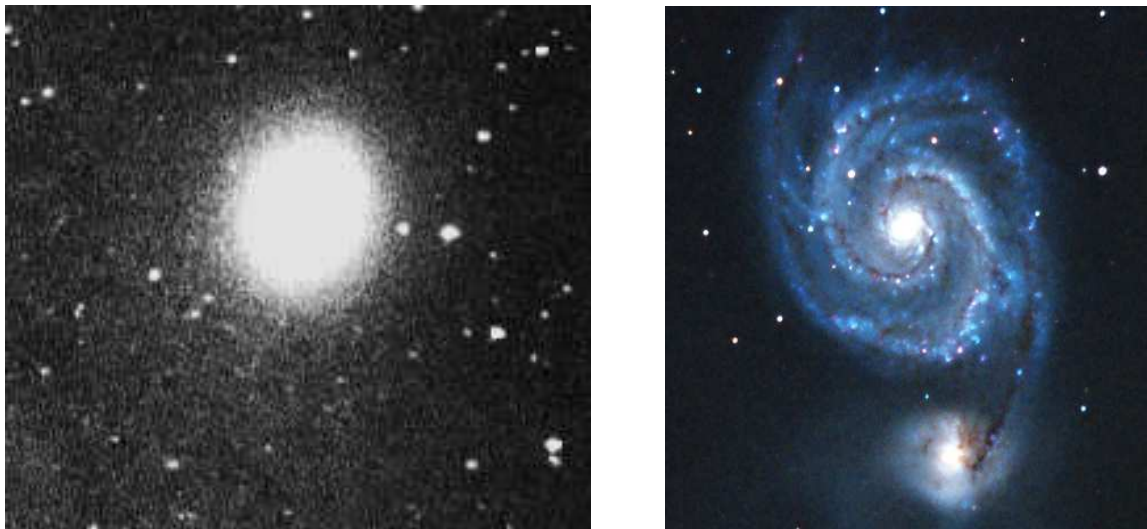


FIGURE 6.14— In the left panel, the elliptical galaxy M32 is shown. It is the companion of the large spiral galaxy Andromeda (see the cover of this thesis) and the subject of investigation of chapter 4 of this thesis. On the right panel, we see the spiral galaxy M51 (also called ‘The whirlpool’) viewed from above and its companion, NGC 5194, at the bottom. The spiral arms of M51 are clearly visible.

gravitational forces. A dynamical model is a mathematical description of a galaxy and as such, provides a working tool for the astrophysicist.

From the study of individual objects, we wish to extract the fundamental laws that govern the life of galaxies. Here we focus on the motion of stars inside early-type galaxies and neglect their other properties: stars are considered as “point-masses” that follow orbits in the combined gravitational field of all the mass of the galaxy (mostly the other stars and the gas).

Building a dynamical model for an elliptical galaxy is more complicated than for a spiral. In spiral disks, stars travel mostly on circular (or nearly circular) orbits. By contrast, motion in elliptical galaxies is not so simple, and orbits can have many different shapes (some orbits can even be chaotic). Without the a priori knowledge of what the stellar orbits are in elliptical galaxies, we compute a large variety of orbits (hopefully covering all types) and then solve for their occupation, i.e., the number of stars traveling on each of them. This is done under the condition that the model made out of orbits resembles the real galaxy, not only in shape, but also in velocities. As kinematic constraints, we use the observed VPs at different points of the galaxy. Once the model is complete, we can peer into it and extract the full velocity and position of any star on any orbit. In this thesis, we constructed such orbit models for several early-type galaxies: NGC 2434 (chapter 2), M32 (chapter 4), NGC 4342 (chapter 5) and NGC 2320 (chapter 6).

Dark Matter

The velocity of a star is due to the gravity field: mass produces gravitational forces that can “pull” the star and change its velocity (a falling apple on Earth experiences the same phenomenon). Hence mass and velocity are intimately related: the larger the mass, the higher the velocity. We can measure velocities of stars and deduce the gravitational field (i.e., the amount of mass) through the equations of motion. Alternatively, one can deduce the amount of mass by measuring the stellar light and converting it into mass (since we know what is the conversion factor between mass and luminosity for a typical stellar population).

These two determinations of mass do not agree when applied, e.g., to the total mass of spiral galaxies. The observed velocities in the outer parts of the galaxy are too high to be explained only by the luminous mass. The discrepancy is very significant, since the “luminous” mass can account for only ten percent of the “dynamical” mass. Therefore, in spiral galaxies, there is a substantial fraction of the total mass that is *not* luminous, called *dark matter*.

Black Holes (BHs)

Quasars are extremely bright and very distant galaxies. As we explained previously, we see these objects as they were in their “infancy”. They emit much more energy than what could be expected from their stars only. The amount of energy released is so large that an “extreme” explanation was needed to describe the early stages of the life of these so-called active galaxies.

BHs are objects that have such a strong gravitational field that nothing can escape from it, not even a gleam of light. BHs at the center of galaxies can be very massive, e.g., as massive as a hundred million Suns! Gas, stars and dust can be swallowed by the hole, releasing vast amount of energy during the process: this is the only mechanism that can explain such large energy generation. So if BHs are suspected in the center of active galaxies, what is the situation for normal galaxies (i.e., the case for most galaxies today)? BHs can also reveal their presence through the strong gravitational field they exert on the surrounding stars. As we have seen, stars orbiting very near a super-massive BH will have a very high velocity. Such high velocities are indeed observed in the center of many elliptical galaxies, so massive BHs were proposed as plausible explanations.

In this thesis, we examine in detail the central regions of two early-type galaxies, M32 and NGC 4342. We have constructed general dynamical models to interpret the observed velocities of the stars. Models that do *not* include a massive BH are unable to explain the observations: stars in such models do not travel as fast as observed. We conclude that for both galaxies, a massive BH in the center is the most likely explanation.

Dark Halos (DHs)

We mentioned that the dynamical mass measurements do not agree with the luminous mass estimation for spiral galaxies. In a dynamical model, one can increase the stellar velocities in the outer regions by embedding the model galaxy into a halo of dark material. Since models based on the light only (i.e., without dark matter) tend to underestimate the velocities at large radii, dark halos (DHs) were postulated to exist around spiral galaxies to explain the large observed velocities.

On the contrary, much less is known concerning the DHs of elliptical galaxies. This is partly due to the more complex orbit families that populate these objects, and hence the greater difficulty to build a dynamical model. Another reason is that velocity measurements are hard to perform in the (very faint) outer parts of elliptical galaxies.

Using extended kinematic measurements and general dynamical models, we studied the round elliptical galaxy, NGC 2434 (chapter 2). The models based on the light only could not explain simultaneously all the kinematic observations, indicating that some dark matter was required. We built a grid of models to study the properties of the DH around that galaxy. We find that the best fit is obtained with a very massive halo (heavier than the typical DH around a *spiral* that would have the same total luminosity than NGC 2434).

In chapter 6, we carried out a similar study for the flattened galaxy NGC 2320 with the axisymmetric code of chapter 3, and using not only the stellar kinematics but also that of the gas.

Curriculum Vitae

I was born on November 18, 1968 in Sion (Switzerland). In 1988, I obtained my high school degree (certificat de maturité) at the “Collège des Creusets” in Sion. I studied physics, and later on astronomy at the university of Geneva until the end of 1992. Under the supervision of Prof. Dr. Martinet and Dr. Pfenniger, I did my diploma thesis at the “Observatoire de Genève”.

After completion of this work in September 1993, I accepted an invitation of Prof. Dr. P. T. de Zeeuw and came to Leiden. I had funding for one year from the Swiss government, but the project I embarked on with Professor de Zeeuw turned out to be much longer and much richer than it originally seemed (at least to me): it was the seed of what eventually became this thesis.

During the course of this work, I had the chance to collaborate with many people and in particular with Prof. Dr. H.-W. Rix, who became my second thesis advisor. I visited him in MPIA (Garching), Steward Observatory (Tucson, Arizona) and MPIA (Heidelberg). During my research in Leiden, I had the opportunity to attend several international conferences and workshops to present my work: Paris, Canberra, Rutgers and Venice. I participated in two NOVA fall schools for astronomy in Dwingeloo and also attended the Canary Island Winter School.

After my promotion, I will continue my research as a post-doc in MPIA, Heidelberg.

Acknowledgments

A thesis is very rarely a “solo voyage” and this one was no exception. In the course of this research, I have benefited from the support of many people: colleagues, friends and family.

At the Leidse Sterrewacht, Richard Arnold had a decisive influence on the progress of this work. He was always ready to help and discuss intricate bugs. Perhaps most importantly, he showed me how to attack an astrophysical problem and solve it efficiently. Richard convinced me *always* to be dubious of a result generated by a computer code! For numerous discussions and advice, I am grateful to my (former and present) colleagues at the Sterrewacht, in particular Hong-Sheng Zhao, Eric Emsellem, Frank van den Bosch, Yannick Copin, Amina Helmi and Wouter Vlemmings. The Sterrewacht is a very lively place and gives plenty of opportunities to meet visitors: I feel thankful to the people responsible for this situation, that I find truly favorable to scientific collaboration. For their proficiency and availability, I am indebted to the computer managers of the Sterrewacht, Erik Deul, David Jansen and Aart Vos. My gratitude goes to Jos de Bruijne, Ronnie Hoogerwerf and Ellen Verolme for their help with the Dutch summary.

I had a rather jagged funding history during this thesis and I would like to mention here the various contributors. I first came to Leiden with a one year grant from the canton Valais (“bourse de l’Etat du Valais”). During the next year, I was supported by the Leidse Sterrewacht. For the subsequent nine months, NUFFIC (Netherland Organization for Higher Education) awarded me a grant, which was supplemented by the Sterrewacht. From October 1996 until January 1998, I was a so-called Beurs Promovendus, i.e., I received the standard PhD grant of Leiden University. The publication of this thesis was supported by the Sterrewacht and the Leids Kerkhoven-Bosscha Fonds (LKBF).

I gratefully acknowledge the hospitality of MPIA (Garching), Steward Observatory (Tucson, Arizona) and MPIA (Heidelberg), where I had the chance to spend fruitful working visits (varying from one week to three months). During those visits, partial financial support was offered by these institutions. LKBF, NWO and the Sterrewacht provided generous financial support for travels to various conferences and astronomical institutes.

Finally, I would like to thank warmly my close family (Mike and Anne) for their continuous support in uncertain times. During my Leiden years, Barbara has been my daily and most faithful supporter. She managed to alleviate most of the stress generated by this thesis. At last, I am most grateful to her for teaching me (or at least trying to) her invaluable diplomatic skills.