Numerical Simulations of Galaxy Interaction

A thesis submitted for the degree of Doctor of Philosophy at the University of Leicester.

by

Paul McMillan

Theoretical Astrophysics Group
Department of Physics and Astronomy
University of Leicester

September 2006
To anyone who deserves a thesis dedicated to them
Acknowledgements

I would like to thank my officemates, past and present, for their help, patience, and occasional paper airplanes. Sam, Yohann, Skippy, Matthew, Patrick, Jalpesh and Emma were all a joy to share an office with; Dave, Owen and Rosie were that too and provided many of the templates and style files without which this thesis would have been many, many times uglier and harder to put together than it was.

I’d like to thank my supervisor, Walter Dehnen for his insight and advice.

I’d also like to thank the other members of the Theory group, Klaus, Dan, Martin, James, Ralf, Jim, Lynette, Andrew, Sergei and Dean, without whom my working days would have been much duller. I also want to thank Graham for his guidance when I needed it; Lisa for making the group run that much more smoothly; and Stuart, Chris and Richard for keeping the computers running so I could actually do any of this work.

My sanity has been preserved by my friends outside the office. Matt, Griff, Johnny, Harriet et al, who are always happy to join me for a drink in the Dorset whenever I return home. Robin, Yusef, Bober, Canavan, Beccy and others who cast me in and/or acted with me in various plays. Most of all my housemates Matt, Mark, Emma and Dan; and Harmony, who defies easy categorisation.

My time in Marseille was of great value for the guidance I received from Lia Athanas-soula, but it would have been a much duller experience but for the welcome and friendship of Lia, Albert, Jorge and Merce. I’d also like to thank Jean-Charles for all his computing support.

Finally, I’d like to thank my family, for their support, wit, warmth, and for absolutely everything else.
Abstract

Cosmological theories that include a non-baryonic dynamically cold dark matter (CDM) have been stunningly successful at explaining observations of the universe on large scales. On the scale of individual galaxies, however, observations have been made which call into question the CDM paradigm.

In particular, simulations of structure formation show CDM haloes with a density “cusp”, such that in the centre of the halo \( \frac{d \ln \rho}{d \ln r} \sim 1 - 1.5 \). In contrast, observational studies suggest that CDM haloes have constant density cores.

In this thesis I use gravitational \( N \)-body simulations to investigate the claim that the dark matter halo cusp can be removed by angular momentum transport from a rotating bar in a disc galaxy. I find that the simulations which were used to support this claim were seriously flawed, and similar simulations designed to mitigate these flaws suggest that this is unlikely to be a mechanism for turning a cusp into a core.

In the interests of further work on dark matter haloes, and on other problems in astrophysics, I design and implement a new method for constructing model galaxies with halo, bulge, and disc components. This method avoids the use of an approximation to a Maxwellian velocity distribution. I show that this creates stable galaxy models, well suited to many applications.

As an example of these applications, I conduct a thorough investigation of the structural and kinematic properties of the haloes of the remnants of 1:1 mass ratio mergers. I determine that the merger has virtually no effect on the halo cusp strength, but a substantial effect on the halo velocity distribution. The remnant haloes are significantly less spherical that those described in studies of mergers which consider gas cooling. Other properties of the remnants are noted and discussed.
1 Introduction

1.1 Galaxies ................................................................. 2
1.2 Dark matter ............................................................ 3
1.3 Cosmological simulations ........................................... 5
1.4 Collisionless dynamics: Space is big ................................ 6
  1.4.1 Collisionless Boltzmann equation .............................. 8
1.5 Modelling ............................................................... 8
  1.5.1 Eddington inversion .............................................. 8
  1.5.2 Osipkov-Merritt models ........................................ 9
  1.5.3 The epicycle approximation .................................. 10
1.6 Dynamical Friction .................................................. 12

2 Numerical methods .................................................... 13

2.1 N-body simulations ................................................ 14
2.2 Initial conditions ................................................... 14
2.3 Particle smoothing ................................................ 15
2.4 Force calculation ................................................... 16
  2.4.1 Tree-codes ....................................................... 18
  2.4.2 falcON ............................................................. 18
2.5 Integrator ............................................................. 21
3 Halo evolution in the presence of a disc bar

3.1 Introduction ........................................................................................................ 25
  3.1.1 Possible explanations for observed cored haloes .................................. 25
  3.1.2 Dynamical friction ......................................................................................... 26
  3.1.3 Previous work ................................................................................................ 27

3.2 Modelling .............................................................................................................. 28
  3.2.1 Technical details ............................................................................................ 30

3.3 Results .................................................................................................................. 30
  3.3.1 Unstable evolution ......................................................................................... 33
  3.3.2 NFW halo ....................................................................................................... 33
  3.3.3 Preventing instability ..................................................................................... 36
  3.3.4 Angular momentum conservation ................................................................. 39

3.4 Conclusions ........................................................................................................... 39

4 Initial conditions for simulations of multi-component galaxies

4.1 Introduction ............................................................................................................ 43

4.2 Existing methods .................................................................................................. 44

4.3 The new method .................................................................................................. 46
  4.3.1 Creating the initial \(N\)-body halo/bulge ...................................................... 46
  4.3.2 Evolving the halo and bulge to adapt to the disc ...................................... 48
  4.3.3 Populating the disc ....................................................................................... 49

4.4 Testing .................................................................................................................. 52
  4.4.1 Testing the disc growth ................................................................................. 54
  4.4.2 Testing the disc model .................................................................................. 58
  4.4.3 Testing the full model ................................................................................... 59

4.5 Conclusions ........................................................................................................... 69
## Contents

5  The haloes of merger remnants  

5.1  Introduction  

5.1.1  Dynamical processes of importance  

5.2  Galaxy models and simulations  

5.2.1  Initial Conditions  

5.2.2  Suites of simulations  

5.3  Results  

5.3.1  Halo density profiles  

5.3.2  Halo velocity distributions  

5.3.3  Halo Shape  

5.4  Discussion  

5.5  Conclusions
Chapter 1

Introduction

“In the beginning the Universe was created. This made a lot of people very angry and has been widely regarded as a bad move.”

Douglas Adams
1.1 Galaxies

Galaxies are some of the most beautiful objects in the night sky (Figure 1.1, title page). Our own galaxy (Figure 1.1 top-left) is visible to the naked eye as a continuous white stream across the sky. It was known as “γαλαξιας – galaxias” (the milky circle) to the ancient Greeks, and it is from this that we derive the word galaxy. The Romans referred to it as *Via Lactea – the Milky Way*. Until Galileo Galilei (1610) looked at it through a telescope, no one had realised that the Milky Way was made up of millions upon millions of stars, too faint and close together for the naked eye to resolve.

Also visible to the naked eye is the Andromeda galaxy (also known as M31, figure 1.1 bottom left), and the Magellanic clouds (visible from the southern hemisphere). The first recorded observations of these galaxies is from Persian astronomer ’Abd Al-Rahman Al Sufi (964), who referred to them as “little clouds”. European astronomers were not aware of the Magellanic Clouds until the crew which had been lead by Ferdinand Magellan – until his death at sea in 1521 – returned to Spain after circumnavigating the globe in 1522. The Andromeda galaxy (or “nebula”, as it was thought to be) was only “discovered” in Europe when Simon Marius observed it in 1612.

It was Immanuel Kant (1755) who first suggested that Andromeda, the Magellanic clouds and the other nebulae that had, by then, been observed might be “island universes”, like the Milky Way, viewed at great distances. This was not immediately accepted, and the observational evidence remained unclear enough, for long enough, that as late as 1920 it was the subject of what became known as the “Great Debate”. This was a lecture each from Harlow Shapely, who argued that the nebulae were part of the Milky Way, and Heber Curtis who argued that they were “island universes”. Neither was thought to have won the argument. It wasn’t until Edwin Hubble (1922) observed Cepheid variable stars in M31, and could compare their apparent brightness with estimates of their luminosity, that there was compelling evidence that M31 was external to the Milky Way.

The Milky Way contains \( \sim 10^{11} \) stars and \( \sim 10^{10} \text{M}_\odot \) of gas, with a total mass \( \sim 6 \times 10^{10} \text{M}_\odot \), mostly in a thin disc of radius \( \sim 15 \text{kpc} \). The largest galaxies contain \( \sim 10^{12} \) stars. This, however, represents a small fraction of the total mass and scale of a galaxy. The rest of the galaxy is made up of so-called “dark matter”.
1.2 Dark matter

The idea of “dark matter”, matter that is only observed through its gravitational effect on other (luminous) matter, has existed in astrophysics since the 1930’s. Fritz Zwicky (1933) estimated a mass-to-light ratios of the nearby Coma cluster of galaxies from the velocities of the galaxies with respect to one another. He compared this to the mass-to-light ratios of nearby galaxies, and noted that the estimated cluster mass-to-light ratio was higher by a factor of $\sim 400$. This estimate was crude, and based on some faulty assumptions, for instance of the distance to the Coma cluster, but it was prescient. In the 1960’s and 70’s it became possible to make high quality measurements which showed
that the rotation curves of spiral galaxies did not start falling at large radii. It was Freeman (1970) who first stated that galaxies (in this case NGC300 and M33) must, therefore, contain “additional matter which is undetected” in the form of a massive dark halo, more extensive than the visible galaxy. Even today, while the existence of dark matter is generally accepted, some still argue that the evidence from rotation curves (at least) can be explained by a modification in Newton’s laws of gravity in the small acceleration regime (Modified Newtonian Dynamics or MOND: e.g. Bekenstein 2004).

The precise nature of dark matter remains a mystery. For some time it was believed that the haloes could be made up of massive compact halo objects (MACHOs) such as white dwarfs, neutron stars, black holes and brown dwarfs. However, there is now a great deal of evidence that these cannot make up most of the halo. This evidence comes from searches for the microlensing expected due to MACHOs (e.g. Alcock et al. 2000), and from measurements of deuterium abundance which indicate the amount of primordial nucleosynthesis, and thus the baryon density of the universe (e.g. Burles, Nollett & Turner 2001).

It is now generally assumed that the majority of dark matter is non-baryonic, and is made up of some dynamically cold, unknown elementary particle which only interacts with other matter through gravity and the weak force. Popular candidates range on the mass scale from the axion \( (10^{-6} \text{eV}/c^2 \lesssim m_{\text{axion}} \lesssim 10^{-3} \text{eV}/c^2) \) to the neutralino (a supersymmetric or weakly interacting massive particle [WIMP]; \( 46 \text{GeV}/c^2 \lesssim m_{\text{neutralino}} \lesssim 2000 \text{GeV}/c^2 \)). This is known as the cold dark matter, or CDM, model.

On cosmological scales the CDM model, or more accurately its extension the lambda cold dark matter (ΛCDM) model has been exceptionally successful. The ΛCDM model includes both non-baryonic CDM and so-called “dark energy”, the precise nature of which is also unknown. Observations of high redshift supernovae (e.g. Riess et al. 1998) indicate that the expansion of the universe is accelerating. Dark energy, it is thought, has a negative pressure, which causes this acceleration on large scales. The most common theories (e.g. Carroll 2003) are that dark energy is a property of the vacuum like Einstein’s cosmological constant, or is “quintessence” which is not constant, and is predicted by some theories in particle physics.

The most spectacular success of the ΛCDM model came in the observations of anisotropy across the whole sky in the temperature of the Cosmic Microwave Background (CMB) observed by the Wilkinson Microwave Anisotropy Probe (WMAP: Bennett et al. 2003). Other evidence comes from CMB data on smaller scales (e.g. Kuo et al. 2004), light element abundances (e.g. Kirkman et al. 2003), and galaxy surveys which study the large scale structure of the universe (e.g. Abazajian et al. 2003).
The current best fit to the WMAP data, which also fits the other evidence (Spergel et al. 2006), has $\Omega_m h^2 = 0.127^{+0.007}_{-0.013}$, $\Omega_b h^2 = 0.0223^{+0.0007}_{-0.0009}$, where the Hubble constant $H_0 = 100h \text{Mpc}^{-1}\text{km s}^{-1}$, $\Omega_m$ is the total matter density, $\Omega_b$ is the baryonic matter density where the values of $\Omega_m$ and $\Omega_b$ are given relative to the “critical density”, the density of a universe – without dark energy – with an overall spatial curvature of zero (also thought to be the energy density of this universe when dark energy is included). The best fit to the data gives $h = 0.73 \pm 0.03$. This means that there is $\sim 5$ times more non-baryonic matter than there is baryonic.

The $\Lambda$CDM model naturally results in hierarchical structure formation (e.g. Blumenthal et al. 1984). Gravitation collapse creates haloes of dark matter, with small structures forming first and then merging to create large ones. This provides the gravitational potential wells in which it is thought that the observable components of galaxies form (e.g. White & Rees 1978). Evidence for this model comes from comparison of simulations (e.g. Springel et al. 2005) to large scale galaxy surveys, such as the Sloan Digital Sky Survey (Abazajian et al. 2003).

### 1.3 Cosmological simulations

Cosmological simulations of structure formation typically follow the evolution of the dark matter distribution starting with small initial perturbations from homogeneity. The exact nature of the perturbations is dependent on the cosmological model chosen for study. The evolution under gravity is followed using an $N$-body code (chapter 2), with, in some cases, semi-analytic prescriptions to follow gas, star and black hole processes. On large scales (>few 100kpc, the scale of clusters or large numbers of clusters) this model has been immensely successful (e.g. Springel, Frenk & White 2006); it is on the scale of individual galaxies that there are conflicts between the predictions of CDM theory and observational data. The two most frequently cited conflicts are the substructure (or “missing satellite”) problem and the “cusp controversy”.

In the hierarchical structure formation model, the small structures that merge to form larger ones survive, as substructure in the halo (Navarro, Frenk & White 1996). Substructure is observed in the Milky Way’s halo, in the form of dwarf galaxies. However it is well known that the number of substructure haloes around the Milky Way, as predicted by $\Lambda$CDM simulations, is $10 - 100$ times greater than the number of dwarf galaxies observed. This discrepancy can be explained if only a subset of substructure haloes contain dwarf galaxies. Typically, though not always, these are thought to be the most massive...
Introduction: 1.4 Collisionless dynamics: Space is big

haloes (e.g. Stoehr et al. 2002, Kravtsov, Gnedin & Klypin 2004). Alternatively, Avila-Reese et al. (2001) and others have showed that the problem can be resolved by damping out the small scale density fluctuations, for instance by giving the dark matter a thermal velocity distribution (so-called “warm dark matter”).

The cusp controversy is another conflict between models and observations. CDM cosmological simulations consistently predict that the spherically-averaged density profile of dark matter haloes has a sharp peak at the centre, creating a ‘cusp’, where the density increases like a inverse power-law. Navarro, Frenk & White (1997) proposed

\[ \rho(r) = \frac{\rho_c}{(r/r_s)(1 + r/r_s)^2}, \]

as a universal density profile (hereafter called ‘NFW’) for CDM haloes of all masses. Here, \( r_s \) is a scale radius and \( \rho_c \) is a characteristic density. This density profile has a central slope of \( \frac{d \ln \rho}{d \ln r} = -1 \). Later work has questioned this inner slope and the universality over all halo masses, but the existence of a cusp with \( \frac{d \ln \rho}{d \ln r} \) in the range of approximately \(-1\) to \(-1.5\) at the innermost resolved point is a robust prediction of such simulations (e.g. Moore et al. 1998, Power et al. 2003).

Observations of the rotation curves of disc galaxies are the best probes of dark matter on galactic scales, and while some studies cannot rule out consistency with a cuspy NFW-like profile (Swaters et al. 2003, van den Bosch & Swaters 2001), the majority of the observations are claimed to be inconsistent with the density profile model (Côté, Carignan & Freeman 2000, de Blok, McGaugh & Rubin 2001, de Blok et al. 2001, Blais-Ouellette, Amram & Carignan 2001, Salucci, Walter & Borriello 2003, Gentile et al. 2004). Investigations combining rotation curves with gravitational lensing observations came to the same conclusion (Trott & Webster 2002). There is also evidence from observations of the Milky Way itself (Binney & Evans 2001). It has been suggested that the discrepancy may be a result of direct comparison with the NFW fitting formula, rather than with more realistic triaxial haloes (Hayashi et al. 2004), though Gentile et al. (2005) examined this in the case of dwarf galaxy DDO 47 and found that a cusped triaxial halo could not explain their observations. This “cusp controversy” is considered one of the greatest current challenges to the CDM paradigm.

1.4 Collisionless dynamics: Space is big

As Douglas Adams (1979) put it, “Space is big... I mean, you may think it’s a long way down the road to the chemist’s, but that’s just peanuts to space”. Outside of binaries or
higher multiple stars, the distances between stars in a galaxy are vastly greater than the stars themselves. For instance, the nearest stars to the Sun – the Alpha Centurai system – are more than 1pc away, which is $\sim 10^8$ times greater than the radius of the sun. The gravitational force on the Sun from the Alpha Centurai system is of approximately the same magnitude as that from the most massive object in the asteroid belt, Ceres, and 1000 times weaker than the net force on the Solar system that keeps it orbiting the Galactic centre. Because direct interactions between stars in a galaxy are typically weak, it is reasonable to approximate that the force on any individual star (or particle of dark matter) is entirely due to the bulk distribution of mass in the galaxy, and not dominated by the interaction with nearby objects. This is essentially approximating that the potential of the galaxy is smooth.

It is reasonable to question whether or not the many weak interactions between individual stars has a significant effect on the dynamics of a system, as compared to the approximation that stars move in a smooth potential. A quantitative estimate of the significance comes from the idea of a relaxation time (e.g. Binney & Tremaine 1987). One assumes that a star travelling with velocity $v$ undergoes many weak deflections from other stars, which alter its velocity by a total amount $\Delta v$. It is possible to make a crude estimate of $n_{\text{relax}}$, the number of crossings of a galaxy that a star would require for the deflections to cause a total deflection $\Delta v \sim v$,

$$n_{\text{relax}} \sim \frac{0.1N}{\ln(N)}.$$  \hspace{1cm} (1.2)

For a galaxy of $\sim 10^{-11}$ stars, $n_{\text{relax}} \sim 4 \times 10^8$. The crossing time for a star ($t_{\text{cross}}$) is simply $v/R$, where $R$ is a characteristic radius for the galaxy. This is similar to the dynamical time $t_{\text{dyn}}$, which for a system of mean density $\rho$ is

$$t_{\text{dyn}} = \sqrt{\frac{3\pi}{16G\rho}}.$$  \hspace{1cm} (1.3)

A typical star in a Milky Way-like galaxy of $\sim 10^{-11}$ stars has only had a few hundred crossing times since the galaxy was formed, which is massively smaller than $n_{\text{relax}}$. The relaxation time is simply $n_{\text{relax}} \times t_{\text{cross}}$. Dark matter particles in a galaxy halo are almost undoubtedly many orders of magnitude more numerous than stars in a galaxy, so the approximation is even more valid for the halo.
1.4.1 Collisionless Boltzmann equation

A description of any collisionless system moving in a smooth potential $\Phi(x, t)$ can be written as $f(x, v, t)$, where the mass of stars or dark matter in the small volume $d^3x$ centred at $x$, with velocities in the range $d^3v$ centred at $v$ is $f(x, v, t)d^3x d^3v$. $f(x, v, t)$ can alternatively be used to represent the number density. Since the system is collisionless, and it is approximated that dark matter and stars are neither created nor destroyed, one can obtain the collisionless Boltzmann (or Vlasov) equation (e.g. Binney & Tremaine 1987)

$$\frac{df}{dt} = 0,$$  \hspace{1cm} (1.4)

where this is a full convective derivative in phase space such that

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + v \cdot \frac{\partial f}{\partial x} + a \cdot \frac{\partial f}{\partial v}. \hspace{1cm} (1.5)$$

$f$ is known as the distribution function or phase space density. A distribution function is in equilibrium if the partial derivative $\frac{\partial f}{\partial t} = 0$.

A description of the dynamics of a collisionless system (ignoring relativity) can be determined from the collisionless Boltzmann equation (often referred to as the CBE), Newton’s laws of motion, and the Poisson equation:

$$\nabla^2 \Phi = 4\pi G \rho.$$  \hspace{1cm} (1.6)

1.5 Modelling

1.5.1 Eddington inversion

Under certain limited circumstances it is possible to derive an exact equilibrium distribution function for a defined density profile. The most common example is that of spherically symmetric, isotropic system. In that case Eddington (1916) inversion can be used. Typically in this case calculations are performed in terms of the “relative potential” $\Psi \equiv -\Phi + \Phi_0$ and the “relative energy” $\varepsilon \equiv -E + \Phi_0 = \Psi - \frac{1}{2}v^2$, where $\Phi_0$ is some constant defined such that $f(\varepsilon \leq 0) = 0$, and $v = |v|$.

By symmetry, any spherical system whose distribution function is a function of $\varepsilon$ alone will be isotropic, that is to say that the distribution function will be independent of
the direction of the velocity, so \( f(x, v, t) = f(r, v) \) As such the density
\[
\rho(r) = \int f d^3v = 4\pi \int_0^{\sqrt{2\Psi}} f(\Psi - \frac{1}{2}v^2) v^2 dv.
\] (1.7)
Since \( \Psi \) is a monotonic function of \( r \), this can be rewritten in terms of \( \Psi \) and \( \varepsilon \) as
\[
\frac{1}{\sqrt{8\pi}} \rho(\Psi) = 2 \int_0^\Psi f(\varepsilon) \sqrt{\Psi - \varepsilon} d\varepsilon,
\] (1.8)
and then differentiated,
\[
\frac{1}{\sqrt{8\pi}} \frac{d\rho}{d\Psi} = \int_0^\Psi \frac{f(\varepsilon)}{\sqrt{\Psi - \varepsilon}} d\varepsilon.
\] (1.9)
This is an Abel (1823) integral equation with solution
\[
f(\varepsilon) = \frac{1}{\sqrt{8\pi^2}} \frac{d}{d\varepsilon} \int_0^\varepsilon \frac{d\rho}{d\Psi} \sqrt{\varepsilon - \eta}.
\] (1.10)

1.5.2 Osipkov-Merritt models

It is not straightforward to find the distribution function of an equilibrium spherically symmetric system with an anisotropic velocity dispersion. However, under limited circumstances it is possible. Both Osipkov (1979) and Merritt (1985) independently investigated distribution functions for spherically symmetric systems that depend on energy and angular momentum only through the variable
\[
Q \equiv \varepsilon - \frac{L^2}{2r_a^2},
\] (1.11)
where \( r_a \) is known as the anisotropy radius.

To find \( f(Q) \) for a given \( \rho(r) \) one first writes the components of velocity in real space in terms of the polar coordinates of velocity in velocity space \((v, \eta, \psi)\) oriented such that
\[
v_r = v \cos \eta \quad ; \quad v_\theta = v \sin \eta \cos \psi \quad ; \quad v_\phi = v \sin \eta \sin \psi.
\] (1.12)
One can then write \( Q \) in terms of \( v \) and \( \eta \)
\[
Q = \Psi - \frac{v^2}{2} \left( 1 + \frac{v^2}{r_a^2} \sin^2 \eta \right),
\] (1.13)
the integral over velocity space is an integral over \( v \) and \( \eta \), which can then be written as an integral over \( Q \) and \( \eta \) in much the same way as Eddington inversion converts an
integral over $v$ to an integral over $\varepsilon$

$$
\rho(r) = 2\pi \int_0^\pi \frac{r_a^2 \sin(\eta)}{(r_a^2 + r^2 \sin^2(\eta))^{3/2}} \int_0^\Psi f(Q) \sqrt{2(\Psi - Q)} dQ.
$$

(1.14)

The integral over $\eta$ is trivially soluble, and the integral over $Q$ is the same as that in equation 1.9 and can be solved in the same way, yielding

$$
f(Q) = \frac{1}{\sqrt{8\pi^2}} \frac{d}{dQ} \int_0^Q \frac{d\rho_Q}{d\Psi} \frac{d\Psi}{\sqrt{\Psi - Q}},
$$

(1.15)

where

$$
\rho_Q \equiv \left(1 + \frac{r^2}{r_a^2}\right) \rho.
$$

(1.16)

In the limit $r_a \to \infty$, $f(\varepsilon, L) \to f(\varepsilon)$, so the distribution function tends towards the isotropic case (section 1.5.1). For finite $r_a$ we require a method of quantifying the velocity anisotropy. The most usual measure is $\beta$ (Binney 1980), defined as

$$
\beta \equiv 1 - \frac{\sigma_r^2}{\sigma_t^2},
$$

(1.17)

where $\sigma_r$ is the velocity dispersion in the radial direction, and $\sigma_t$ is the velocity dispersions in a tangential direction. In the case of Osipkov-Merritt models

$$
\beta(r) \equiv \frac{r^2}{r^2 + r_a^2}.
$$

(1.18)

### 1.5.3 The epicycle approximation

The modelling of disc galaxies is a major topic in chapter 4 of this thesis. Here I introduce the epicycle approximation, which is useful both as a starting point for this modelling, and in the analysis of disc dynamics in general.

In a disc galaxy, a large number of stars are on near-planar, near-circular orbits. As such it is convenient to have an approximate solution for the motion in these orbits.

An orbit in an axisymmetric system can be considered to be motion in an effective potential

$$
\Phi_{eff} \equiv \Phi(R, z) + \frac{L_z^2}{2R^2},
$$

(1.19)

with

$$
\ddot{R} = -\frac{\partial \Phi_{eff}}{\partial R}; \ddot{z} = -\frac{\partial \Phi_{eff}}{\partial z}.
$$

(1.20)
For a near-circular orbit, one can expand the effective potential as a Taylor series about a circular orbit. A circular orbit in the meridional plane is, by definition, at the minimum in $\Phi_{\text{eff}}$, so if we define $R_c$ as the radius of a circular orbit, and $x \equiv R - R_c$

$$\Phi_{\text{eff}}(R, z) = \frac{1}{2} \left( \frac{\partial^2 \Phi_{\text{eff}}}{\partial R^2} \right)_{(R_c, 0)} x^2 + \frac{1}{2} \left( \frac{\partial^2 \Phi_{\text{eff}}}{\partial z^2} \right)_{(R_c, 0)} z^2 + \text{constant.} \quad (1.21)$$

The term $O(xz)$ is zero because a disc galaxy is symmetric about $z = 0$, and terms $O(xz^2)$ and higher are neglected. One can then define

$$\kappa^2 \equiv \left( \frac{\partial^2 \Phi_{\text{eff}}}{\partial R^2} \right)_{(R_c, 0)} = \left( \frac{\partial^2 \Phi}{\partial R^2} \right)_{(R_c, 0)} + \frac{3L_z^2}{R_c^3}, \quad (1.22)$$

and

$$\nu^2 \equiv \left( \frac{\partial^2 \Phi_{\text{eff}}}{\partial z^2} \right)_{(R_c, 0)} = \left( \frac{\partial^2 \Phi}{\partial z^2} \right)_{(R_c, 0)}. \quad (1.23)$$

The equations governing the behaviour of $x$ and $z$ then reduce to

$$\ddot{x} = -\kappa^2 x; \quad \ddot{z} = -\nu^2 z, \quad (1.24)$$

which are harmonic oscillations with frequencies $\kappa$ (in $x$, known as the epicyclic frequency), and $\nu$ (in $z$, known as the vertical frequency).

The movement of the star in the plane of the galaxy can thus be thought of as following an elliptical path around a guiding centre. This guiding centre follows a circular path of radius $R_c$, with the angular speed one would expect for a circular orbit at that radius, $\Omega_c$. The elliptical path is referred to as an epicycle, from the Greek, meaning “on the circle”.

The original use of epicycles was in the geocentric Ptolemaic system used to explain the movements of the planets in the solar system until Copernicus (1543) introduced his heliocentric model. The Ptolemaic system is based purely on circles, while in the modern epicycle approximation the axis ratio of the epicycle is $\kappa / 2\Omega_c$. In the Keplerian (point-mass potential) limit this tends to $\frac{1}{2}$, with the long axis parallel to the movement of the guiding centre.
1.6 Dynamical Friction

Finally in this introduction, it is important that I discuss the topic of dynamical friction. As I have already mentioned, collisions between stars in a galaxy are extremely rare, because the stars are vastly smaller than the distances between them. Galaxy mergers, such as the one between the Sagittarius dwarf elliptical and the Milky Way which is on-going, or the one between the Milky Way and M31 which is expected to occur in a few Gyr, are unlikely to cause any direct stellar impacts. However mergers, both major and minor, do cause changes within the galaxies that result in energy being transferred to the internal structure of the galaxies from kinetic energy of their relative motion. The physical process which causes this is referred to as dynamical friction.

The classic case (Chandrasekhar 1943) considers a massive body moving through an initially homogeneous distribution of stars. As the body moves through the medium, the individual bodies in the medium are deflected towards it. Since the massive body is moving with respect to the medium, this causes an over-density behind the body compared to the density in front of it. This results in a net force retarding the motion of the body, and the transfer of energy from the body to the medium.

This approximation is a reasonably accurate description of some cases in astrophysics (for instance a globular cluster orbiting in a galaxy), but does not describe, for example, the dynamical friction on a rotating bar (see section 3.1.2).

As the simulations of chapter 5 show, when two equal mass galaxies move through one another – on a path that would take single bodies of the same mass away to infinity – there is a transfer of energy from the mutual orbit to the internal dynamics of both which slows them. In the cases shown, this eventually causes the galaxies to merge. Again this is not equivalent to the situation considered by Chandrasekhar (1943), but it is a similar process, and has a similar net effect.
Chapter 2

Numerical methods

“Turning and turning in the widening gyre
   The falcon cannot hear the falconer;
   Things fall apart; the centre cannot hold”

W.B. Yates
2.1  $N$-body simulations

One of the most powerful tools for any stellar dynamicist is the $N$-body simulation. In an $N$ body simulation the motion of $N$ particles is followed under their mutual gravitational attraction and/or the influence of an external potential. It is a Monte-Carlo approach to solving the differential equations describing the dynamics of the system (section [1.4.1]). It is also an example of the “method of characteristics”, in that the CBE, which is a partial differential equation, degenerates into a series of ordinary differential equations along the particle trajectories.

The first $N$-body simulations were performed by Holmberg (1941), using 37 light-bulbs to represent two galaxies, and measuring the intensity at a point to determine the gravitational force. The first computational $N$-body simulations were performed by von Hoerner (1960), and were of a star cluster with $N = 8 - 16$. The lightbulb method may have allowed larger $N$ at that point in the development of computer technology, but it was relatively quickly overhauled, with $N = 100$ computational simulations being performed by 1963 (e.g. Aarseth 1963). Today the record for the largest $N$-body simulation is held by the “Millennium Run” (Springel et al. 2005) which used $2160^3 = 1.01 \times 10^{10}$ particles.

There are essentially 3 main steps in running a gravitational $N$-body simulation:

1. Setting up the initial conditions.
2. Determining the gravitational force on each body.
3. Integrating the equations of motion.

2.2  Initial conditions

The initial conditions of an $N$-body simulation are a representation of the mass distribution function $f(x, v)$ (section [1.4.1], typically (in the first approximation) by a set of delta functions

$$ f(x, v) \to \sum_{i=1}^{N} \delta^3(x - x_i) \delta^3(v - v_i). $$

This approximation improves as $N$-increases. This is a major reason why some accuracy in the force computation is generally traded for increased particle number in the methods described in section [2.4]. Increased $N$ also helps by reducing the effects of two-body
relaxation, as can be seen from the strong $N$ dependence of the relaxation time (e.g. equation [1.2]).

Where $f(x, v)$ is known, it is simplest to treat it as a probability distribution, and all bodies as having an equal mass $m_i$, and pick $x_i, v_i$ with a probability proportional to $f(x_i, v_i)$. There is a Poisson noise scatter inherent in drawing values truly randomly from this probability distribution, which affects the accuracy of the representation, and causes noise within the simulation. Methods exist to reduce this problem, such as “quiet start” techniques in disc simulations (see chapter [4] or Sellwood 1987), or the use of “quasi-random” numbers (which are defined in such a way as to provide nearly uniform coverage) in drawing values from the probability distribution.

In the case of Eddington inversion (section [1.5.1]), the density distribution $\rho(r)$ and $f(\epsilon)$ are known. It is, in practice, easiest to first choose $x_i$ from a probability distribution corresponding to $\rho(r)$, then find the particle speed using $f(\epsilon)$ via an “acceptance-rejection” method (e.g. Press, Flannery & Teukolsky 1986). For a particle at a radius $r$, a speed $v$ is chosen at random between 0 and the escape velocity at $r$. That speed is then accepted or rejected with a probability proportional to $v^2 f(\epsilon) \equiv v^2 f(\Psi(r) - v^2/2)$. Since the velocity dispersion is isotropic in this case, the direction of the velocity can then be chosen at random over a sphere.

2.3 Particle smoothing

It is simplest to think of an $N$-body simulation made up of $N$ point particles. Indeed, in collision-dominated systems, such as star clusters, where one genuinely wants to investigate the dynamics of $N$ bodies, this approach is useful (Leeuw, Combes & Binney 1993), though the infinitely deep potential well it defines can cause numerical problems. In the case of a collisionless system, such as an entire galaxy, it is better to represent the system as $N$ smoothed particles.

A reasonable sized galaxy, such as the Milky Way, contains $\sim 10^{11}$ stars, which is an order of magnitude more bodies than the largest $N$-body simulation ever performed (Springel et al. 2005), and several orders of magnitude more than an $N$-body simulation on a desktop computer can handle. As such, one cannot think of the bodies as representing individual stars, but rather as sampling the distribution function, so using smoothed particles is entirely valid.

In addition to removing singularities from the potential field, softening the particles has the important advantage of suppressing artificial two-body encounters, which would
otherwise require very careful and CPU intensive numerical integration of the equations of motion. In this sense, softening can be said to ensure that the simulation is close to collisionless.

The cost of softening is an inevitable loss of spatial resolution, and a bias in the force calculations. The standard approach to softening is to represent each body by a Plummer (1911) density model

\[
\rho(x) \propto \frac{\varepsilon^2}{(|x - x_i|^2 + \varepsilon^2)^{5/2}},
\]

where \(\varepsilon\) is the smoothing length. This introduces a bias in calculations at small separations (the force going to zero for \(|x - x_i| \ll \varepsilon\), rather than tending towards infinity), and indeed at large separations, as the potential only slowly converges to the Newtonian limit. Dehnen (2001) performed a thorough analysis of this problem. In all simulations in this thesis I follow Dehnen’s advice and use the stronger power law

\[
\rho(x) \propto \frac{\varepsilon^2}{(|x - x_i|^2 + \varepsilon^2)^{7/2}},
\]

which reduces the bias at large separations.

### 2.4 Force calculation

The most taxing problem in an \(N\)-body simulation is that of determining the gravitational force on each of the \(N\) bodies due to one another at every time-step. The most obvious approach is to compute the forces directly (e.g. Aarseth’s code reproduced in Appendix 4.B of Binney & Tremaine 1987). In the simplest case, for point particles

\[
\ddot{x}_i = -G \sum_{j \neq i}^N \frac{m_j (x_i - x_j)}{|x_i - x_j|^3}. \tag{2.4}
\]

This direct summation approach is slow, requiring \(N - 1\) calculations for each of the \(N\) particles, and thus scaling as \(O(N^2)\). Specialised hardware exists for this (the GRAvity PipE, or GRAPE computers, Kawai et al. 2000), but this approach is only normally used in cases with \(N \lesssim 10000\) and where the forces at close range are particularly important.

A common alternative are the so-called Particle-Mesh (PM) or grid codes. In these the gravitational field of the system is approximated on a mesh. The positions (and masses) of the particles are used to assign a density to every point on the mesh. Using a mesh means that the Poisson equation can be solved using a Fast Fourier Transform
Numerical methods: 2.4 Force calculation

(FFT) routine, under some assumption for the boundary conditions, usually that they are periodic. This gives the potential on the mesh, from which the force field can be calculated, and the force on each particle can be interpolated.

The simplest implementation of the PM method is on an unchanging Cartesian grid. In this case the computational cost scales as $O(N + N_g \log(N_g))$ where $N_g$ is the number of grid points. The disadvantage of this is that it makes it difficult to handle non-uniform particle distributions, which limits the simulation’s resolution. It is possible to alleviate this problem somewhat with alternative grid geometries in cases where the high density regions are known in advance, for example using a cylindrical polar grid in disc simulations (Pfenniger & Friedli 1991) or a spherical grid in halo or collapse simulations (van Albada 1982).

A number of refinements of the PM method have been developed. In particular it is possible to include an adaptive mesh refinement, such that after the initial grid potentials are calculated, sub-grids are introduced automatically in regions depending on their number density, allowing greater resolution in the calculation of the interparticle forces (e.g. Gelato, Chernoff & Wasserman 1997). Some simulations also (or alternatively) find the particle-particle forces directly for short range forces, and use the PM method for the long range forces. This is the Particle-Particle/Particle Mesh, or P$^3$M method (Hockney & Eastwood 1981).

Another alternative is the so-called Self-Consistent Field (SCF) method (Hernquist & Ostriker 1992a). In this method the density and potential are expanded as a series of orthogonal basis functions such that

$$\rho(x) = \sum_{n l m} A_{nlm} \rho_{nlm}(x) ;$$  \hspace{1cm} (2.5)

$$\Phi(x) = \sum_{n l m} A_{nlm} \Phi_{nlm}(x) ,$$  \hspace{1cm} (2.6)

where $n$ is the radial “quantum” number, and the $l$ & $m$ terms contain the angular dependence, expanded in the familiar $Y_{lm}$ spherical harmonics. The force on each particle can then be found from the analytic differential of the potential. In problems with a high degree of symmetry, and with a careful choice of basis sets it is possible to truncate the expansions in equations 2.5 & 2.6 at low order with little loss of accuracy. Computational cost scales as $\sim O(N n_{max} l_{max}^2)$, but the method is rather inflexible, and suffers badly if systems are off-centred, since it then requires computation of a large number of terms in equations 2.5 & 2.6 to find the true potential.
2.4 Force calculation

2.4.1 Tree-codes

Tree-codes (e.g. Barnes & Hut 1986) have no fixed grid which means they have no preferred geometry and do not waste time simulating regions without matter. They are particularly valuable for simulations in which the mass distribution changes significantly, such as galaxy mergers. The major idea of a tree-code is the recognition that the potential of a distant group of particles can be approximated by a low-order multipole expansion. The code starts with a single cell, containing all of the particles. This cell is then divided into 8 sub-cells (2 in each direction). Each of these is then subdivided in the same way, and so forth. For each cell the mass, centre of mass, and (in later implementations) low order multipole moments are computed. The force on a particle is found by a “tree walk”. Starting from the top cell in the tree, each cell is checked to see whether it is distant enough from the particle in question, according to a pre-defined criterion. If it is then it is put on the “interaction list”, if not, then it is opened and each sub-cell is checked in the same way, and so on until a complete interaction list of cells and individual particles is found. The force on the particle is then found by adding up the contributions of every cell or particle on the interaction list.

This reduces the number of interactions that need to be computed, thus speeding up the process. The computational cost, both of building the tree, and of determining the particle accelerations, scales as \( O(N \log(N)) \). One downside is that storing the tree structure costs in terms of memory usage. Building the tree takes only a few percent of the CPU time, so it is reasonable to rebuild it on every time-step. The process is sped up still further by exploiting the fact that nearby particles have very similar interaction lists (Barnes 1990).

2.4.2 falcON

All the simulations performed for this thesis use the \( N \)-body code gyrfalcON, which is based on Dehnen’s (2000, 2002) force solver falcON [force algorithm with complexity \( O(N) \)] and utilises the NEMO \( N \)-body package (Teuben 1995). Like the Barnes & Hut tree-code, falcON begins by building a tree of cells at each time-step. falcON determines the potential of the system using multipole expansions for cells, and exploits the similarity of the force from a distant cell upon cells that are near to each other. The multipole expansion is performed in Cartesian co-ordinates, rather than spherical harmonics.

For any particle, smoothed or otherwise, the potential at a body position \( x_i \) is gener-
alised by
\[ \Phi(x_i) = -\sum_{j\neq i} m_j g(x_i - x_j), \] (2.7)
where \( g(x) \) is a Greens function describing the potential of the particle. For a point particle \( g(x) = G/|x| \). For the smoothing kernel from equation 2.3
\[ g(x) = \frac{G(|x|^2 + \frac{3}{2} \varepsilon^2)}{(|x|^2 + \varepsilon^2)^{3/2}}. \] (2.8)

The approximation for gravity works as follows. Consider two cells A and B with centres of mass \( z_A \) and \( z_B \). The mutual interaction between a body at position \( x \) in cell A and a body at position \( y \) in cell B is can be expanded as a Taylor series about the separation \( R = z_A - z_B \):
\[ g(x - y) = \sum_{n=0}^{p} \frac{1}{n!} (x - y - R)^{(n)} \odot \nabla^{(n)} g(R) + \mathcal{R}_p(g), \] (2.9)
where I follow Dehnen's notation conventions in which \( x^{(n)} \) indicates the \( n \)-fold outer product of the vector \( x \) with itself, and \( \odot \) is a tensor inner product. \( p \) is the order of the expansion, and \( \mathcal{R}_p \) denotes the Taylor series remainder. In the gyrfalcON implementation \( p = 3 \).

To use this for whole cells the algorithm first calculates the multipole moments of cell B in Cartesian co-ordinates
\[ M_B^n = \sum_{y_i \in B} m_i (y_i - z_B)^{(n)}. \] (2.10)

In practice only the monopole (which is the mass of the cell, \( M_B^0 = m_b \)) and specific quadrupole moment (\( M_B^2 = M_B^2 / m_b \)) are needed by gyrfalcON. The dipole term is zero because the expansion is performed about the centre of mass of the cell; the octopole (\( M_B^3 \)) term only contributes to the potential (not its gradient, the force) in a 3rd order expansion, so can be ignored. This approximation is taken to be acceptable if the two cells are well separated. Whether or not two cells are well separated is determined by the condition:
\[ |Z_A - Z_B| > (r_{\text{max}A} + r_{\text{max}B})/\theta, \] (2.11)
where \( r_{\text{max}} \) is the radius about the centre of mass \( Z \), which encloses all the particles in the cell, and the opening angle \( \theta \) controls the accuracy of the code. \( \theta \) is not a constant, but is a weak function of the cell mass, in this case \( m_b \). This helps to ensure that the absolute error in the force calculation for an individual particle is not dominated by the
interaction with the most massive cells (as it would be for constant $\theta$). In all simulations in this thesis $\theta(M_{\text{tot}}) = 0.64$, where $M_{\text{tot}}$ is the total mass of the system. If the cells are well separated the potential due to cell B on cell A is then determined as

$$
\Phi_{B \to A}(x) \approx -\sum_{l=0}^{3} \frac{1}{l!} (x - z_A)^{(l)} \odot C_{B \to A}^l,
$$

(2.12)

where the field tensors $C_{B \to A}^l$ are given by

$$
C_{B \to A}^l = \sum_{n=0}^{3-l} \frac{(-1)^n}{n!} \nabla^{(n+l)} g(R) \odot M^n_{B}.
$$

(2.13)

Since the smoothed particles are spherically symmetric, $g$ is purely a function of radius, so the code can calculate the terms $\nabla^{(n+l)} g(R)$ through

$$
D^k \equiv \left( \frac{1}{r} \frac{\partial}{\partial r} \right)^k g(r) \bigg|_{r=|R|}.
$$

(2.14)

In practice the code calculates the coefficients

$$
\begin{align*}
    m_a C_{B \to A}^0 &= m_a m_b \left[ D^0 + \frac{1}{2} \hat{M}_{Bii}^2 D^1 + \frac{1}{2} R_i R_j \hat{M}_{Bij}^2 D^2 \right], \\
    m_a C_{B \to A, i} &= m_a m_b \left[ \delta_{ij} D^1 + R_i R_j D^2 \right], \\
    m_a C_{B \to A, ij} &= m_a m_b \left[ \delta_{ij} D^1 + R_i R_j D^2 \right], \\
    m_a C_{B \to A, ijk} &= m_a m_b \left[ \left( \delta_{ij} R_k + \delta_{jk} R_i + \delta_{ki} R_j \right) D^2 + R_i R_j R_k D^3 \right],
\end{align*}
$$

(2.15)

where I have used Einstein’s sum convention, $R_i$ is the $i$th Cartesian coordinate of the separation between the two centres of mass, and $\hat{M}_{Bij}^2$ is the $i,j$th Cartesian component of the specific quadrupole moment tensor. This has the benefit that $m_a C_{B \to A}^2 = m_b C_{A \to B}^2$, and $m_a C_{B \to A}^3 = -m_b C_{A \to B}^3$, which speeds up the computation. Also, using this for both $\Phi_{B \to A}$ and $\Phi_{A \to B}$ ensures that every action has an equal and opposite reaction.

If the two cells A and B are well separated, it is not necessary to separately determine the interaction coefficients of any of the sub-cells of A with B, or of A with any of the sub-cells of B, or of any of the sub-cells of A with any of the sub-cells of B. It is this, of course, which speeds up the force calculation when compared to the direct summation method. This ensures that there is no double counting in the summation over interaction coefficients. When all the necessary interaction coefficients have been determined, the
coefficients for each cell are summed over all interactions. For instance

$$C^l_A = \frac{1}{m_A} \sum_B m_a C^l_{B\rightarrow A}. \quad (2.16)$$

The force on an individual body can then be found by summing up the interaction coefficients of all the cells that the body is in, from smallest to biggest, once the Taylor expansion of all of them has been translated to a common centre.

Since the Taylor expansion was performed in Cartesian coordinates, this shift in centre is trivial. If $C^l_A$ are the coefficients about the centre of mass of cell A ($z_A$), then the coefficients after a translation to a new expansion centre ($z_t$) are

$$C^l_t = \sum_{n=0}^{3-l} \frac{1}{n!} (z_A - z_t)^{(n)} \odot C^{l+n}_A. \quad (2.17)$$

In addition to the forces due to the interaction between the $N$ bodies in a simulation, it is sometimes necessary to include the effect of some external potential on the bodies (e.g. chapters 3 and 4). This is straightforward; gyrfalcON calculates the force due to the external field (either analytically or numerically), and adds it to the inter-particle interaction force to find the total acceleration of the particle whenever the integrator requires it.

### 2.5 Integrator

The method that gyrfalcON uses to integrate the equations of motion for the particles is a block-step scheme based upon the leapfrog integrator.

In its simplest form the leapfrog integrator is as represented in figure 2.1. The particle positions (and thus accelerations) are determined at the start of every time-step, the velocities are determined at the halfway point in each time-step. A practical scheme of this kind, often described as a kick-drift-kick scheme, is represented mathematically as a loop over the operations:

$$v_i(t + \frac{\Delta t}{2}) = v_i(t) + a_i(t) \frac{\Delta t}{2}, \quad \text{(Kick)}$$
$$x_i(t + \Delta t) = x_i(t) + v_i \left(t + \frac{\Delta t}{2}\right) \Delta t, \quad \text{(Drift)}$$
$$a_i(t + \Delta t) = -\nabla_i \Phi(t + \Delta t),$$
$$v_i(t + \Delta t) = v_i \left(t + \frac{\Delta t}{2}\right) + a_i(t) \frac{\Delta t}{2} \quad \text{(Kick).} \quad (2.18)$$
Figure 2.1. Schematic representation of a simple leapfrog method. $x_1$ is determined from $x_0$ and $v_{1/2}$. $v_{3/2}$ is determined from $v_{1/2}$ and the acceleration determined from $x_1$, and so forth.

Force calculation (the third step in that loop) is carried out in the way discussed in section 2.4.2.

The leapfrog integration scheme is accurate to second order in $\Delta t$, but its main benefit is that it is time reversible. This ensures that the integration does not suffer from systematic errors, for instance one would expect it to prevent simple orbits from spiralling in or out, since that would break the time symmetry. It is “symplectic”, which means it corresponds to a Hamiltonian, one that only differs from the true Hamiltonian at third order in $\Delta t$. This ensures that it conserves phase space volume. It also ensures conservation of energy (time-averaged) over the long term, to within the accuracy of the falCON algorithm.

The leapfrog method is excellent for situations in which solving for the acceleration of every particle at every time-step is appropriate. However in situations with a large range of dynamical times, such as a cusped halo, this is not the case. In that case, far closer time integration is required for particles near the centre of the halo than those on the outskirts. Using a single time-step scheme would either mean that the innermost particles were not followed closely enough, rendering the simulation results invalid; or that the outermost particles were integrated unnecessarily closely, which can be a huge waste of CPU time.

In order to cope with systems with a large range of dynamical times gyrfalCOn uses a block-step schemes which is represented schematically in figure 2.2. The full time-step (block-step, $\tau_{max}$) is broken up on $n$ levels, each with time-steps half as big as the level before, allowing for a smallest time-step $\tau_{min}$ which is $2^n$ times smaller than $\tau_{max}$. The positions of every particle must be determined at every time-step, so all particles are “drifted” with time-steps $\tau_{min}$. However, the force computation (the “kick”)
Figure 2.2. Schematic representation of the block-step scheme. The diagram shows one full block-step with a scheme allowing time-steps 8 times smaller. Black squares correspond to the “kicks”. The filled black squares correspond to points where the accelerations need to be computed, the outline squares to points where the acceleration has already been computed (in the previous block-step).

Criteria for choosing time-steps are only useful if they are Galilean invariants. As such I use criteria based upon the acceleration $\alpha_i$ and potential $\Phi_i$ of a body. These are parametrised through $\text{fac}$ and $\text{fph}$, and the “target” time-step, $\tau_i$ which is defined as

$$
\tau_i = \min \left\{ \frac{\text{fac}}{|\alpha_i|}, \frac{\text{fph}}{|\Phi_i|} \right\}. 
$$

(2.19)

At the beginning of each loop, $\tau_i$ is determined for each particle and compared to the time-step of the regime the particle was in for the previous loop, $h_{n,i}$. The particle time-step is chosen such that $\tau_i$ is an average time-step.

The energy conservation in simulations using multiple time-steps is still excellent (as will be shown in chapters 3, 4 and 5), indicating that the integration scheme is perfectly acceptable.
Chapter 3

Halo evolution in the presence of a disc bar

“The great tragedy of Science – the slaying of a beautiful hypothesis by an ugly fact.”

Thomas Henry Huxley
In this Chapter I will investigate the interaction between the gravitation perturbation associated with barred galaxies and the galactic CDM haloes. In particular I will investigate the claim from Weinberg & Katz (2002) that angular momentum transport from the bar to the halo is a plausible mechanism for removing the CDM “cusp”.

3.1 Introduction

As discussed in section 1.3, CDM cosmological simulations consistently predict that the density profile of dark matter haloes increases sharply towards the centre, creating a ‘cusp’, where the density increases like a inverse power-law $\rho \propto r^{-\gamma_0}$ with $1 \lesssim \gamma_0 \lesssim 1.5$. This is typified by the NFW density profile, equation 1.1. However, a large number of observed galaxy rotation curves are inconsistent with this picture, and suggest that the true density profile has a constant density core.

3.1.1 Possible explanations for observed cored haloes

Numerous possibilities have been put forward in an effort to explain the difference between theory and observation. Many of these involve altering the assumed properties of dark matter, either the dark matter particle itself (e.g. Spergel & Steinhardt 2000, Kaplinghat, Knox & Turner 2000), or damping out small scale density fluctuations in the dark matter distribution (e.g. Colín, Avila-Reese & Valenzuela 2000, Bode, Ostriker & Turok 2001).

Since the central region of most galaxies are dominated by baryonic matter, it seems reasonable that the interaction between baryonic and dark matter will have a major effect upon the dark matter density in the inner region of the halo. If the effect of this interaction was to reduce the dark matter density at the centre of the halo, then this could potentially solve the problem. Several investigation of such processes have been undertaken, some pointing out effects that would increase the gradient of the density cusp still further, such as contraction associated with the dissipative infall of baryons to form a galactic disc. (Blumenthal et al. 1986). More recently there have been a number looking at mechanisms that could potentially remove the cusp:

- Binney, Gerhard & Silk (2001) proposed that an outflow of a large fraction of the baryonic mass of the galactic disc, after the initial collapse, could alter the density structure of the halo in such a way as to remove the cusp. However Gnedin & Zhao
Halo evolution in the presence of a disc bar: 3.1 Introduction

(2002) found that that, even in the case where 100% of the baryonic content was ejected, the central density was only reduced by a factor $\sim 2$.

- El-Zant, Shlosman & Hoffman (2001) present simulations in which the baryonic content of the galaxy forms “clumps” which can heat the cusp through dynamical friction thus erasing it.

- Weinberg & Katz (2002) pointed to an earlier study by Hernquist & Weinberg (1992) to suggest that angular momentum would be rapidly transferred from a rotating disc bar to the halo at an inner “Lindblad-like” resonance, and would be sufficient to remove the cusp.

3.1.2 Dynamical friction

Dynamical friction (as discussed in section 1.6) is the process by which a massive body loses energy as it moves through a diffuse, collisionless medium. In the case of a rotating bar the situation is complicated, because the diffuse medium is mainly made up of bodies that are on quasi-periodic orbits, which has important effects upon the nature of the bar’s wake. Angular momentum transfer occurs mainly around resonances between the orbital frequencies of the bodies, and the pattern speed of the bar. In the disc these resonances are the corotation resonance $\Omega_0 = \Omega_b$ where $\Omega_0$ is the circular frequency in the unperturbed (by the bar) potential, and $\Omega_b$ is the pattern speed of the bar; and the Lindblad resonances $m(\Omega_0 - \Omega_b) = \pm \kappa_0$ where $\kappa_0$ is the epicycle frequency (section 1.5.3) for the unperturbed potential, and $m$ is an integer. The cases when $m(\Omega_0 - \Omega_b) = +\kappa_0$ are known as the inner Lindblad resonances (ILR), typically occurring in the inner parts of the galaxy, as opposed to the outer Lindblad resonances (OLR) $m(\Omega_0 - \Omega_b) = -\kappa_0$.

In the case of the halo, the particle orbits are not approximately planar, nor generally close to circular, so the situation is somewhat different, however there are still resonances of the form $l_r \Omega_r + l_\phi \Omega_\phi = m \Omega_b$ where the radial and azimuthal frequencies of the orbit are $\Omega_r$ and $\Omega_\phi$ respectively, and $l_r, l_\phi$ and $m$ are integers.

Since Lynden-Bell & Kalnajs (1972) it has been known that, in a disc, stars at ILR lose angular momentum while those at corotation and OLR gain it. More recent work by Athanassoula (2003) has extended this work to a spheroidal component, such as a halo. She found that, in the case where the distribution function of the halo depends only on energy, all halo resonances absorb angular momentum.
3.1.3 Previous work

Weinberg & Katz (2002) suggested that the angular momentum transport from a large primordial bar to the halo could provide a plausible mechanism by which the cusped haloes seen in cosmological simulations could evolve to produce the cored density profiles inferred from observational evidence (section 5.1). Simulation of galaxy formation which include a dissipative component have suggested that early in the life of a galaxy it is easy to form a large, strong bar (Steinmetz & Navarro 2002).

Weinberg & Katz supported their hypothesis with analytical calculations and greatly simplified \(N\)-body simulations. In the simulations, the gravitational potential of a rotating bar, with its centre pinned at the initial centre of the density distribution, was imposed upon an \(N\)-body representation of a CDM halo with an NFW profile. Their results showed the initially cusped halo density profile seemed to have become cored within just a few bar rotations.

The simulations of Weinberg & Katz (2002) were re-examined by Sellwood (2003) who found the same reduction in central density, but argued that the deliberately artificial perturbation applied to the system lead to misleading results. Sellwood (2003) ran fully self-consistent simulations, which showed a slight steepening of the halo’s inner profile from the action of a bar. This result was also seen in the simulations of Valenzuela & Klypin (2003). However the self consistent simulations of Holley-Bockelmann, Weinberg & Katz (2005) showed a clear flattening in the cusp, as do those of Weinberg (2004). A discussion of this apparently irreconcilable discrepancy was given by Athanassoula (2004).

The simulations of Weinberg & Katz (2002) and Sellwood (2003), and also the original study of Hernquist & Weinberg (1992) all employed spherical harmonics to compute the gravitational potential and accelerations. Sellwood used a hybrid spherical harmonics – particle mesh code, the other two papers used SCF codes (section 2.4). The main motivation for this choice was the high symmetry of the problem, which allows the restriction to a low number of harmonics. This in turn reduces the computational effort and enables a larger number \(N\) of bodies and hence higher numerical resolution than with other methods, such as the Barnes & Hut (1986) tree code or \texttt{gyrfalcON}. In fact, Weinberg & Katz (2002) even claim that cusp removal cannot be successfully simulated by a tree code (unless \(N \gg 10^6\)), because the noise in the forces would scatter bodies off the resonant orbits which are driving the angular momentum transfer. Because of this argument, it was not entirely clear whether the aforementioned results of fully self-consistent simulations are realistic or not.
There is, however, a price to pay for using an spherical harmonics. Most importantly, any method based on spherical polar coordinates is vulnerable to off-centring: if the modelled density distribution is not centred on the origin of the coordinate system, many terms are required for the series to converge. Conversely, when off-centring occurs, the gravity obtained from only the first few terms is strongly biased, which may induce artificial $m = 1$ instabilities and/or interfere with natural $m = 1$ instabilities. The only proper way to avoid this problem is to enforce centring, i.e. not allowing coefficients with odd $l, m$ to carry any weight. Clearly, this makes systems with any properties that are not reflection symmetric, including instabilities, impossible to simulate in this way.

Weinberg & Katz (2002) included coefficients of odd $(l, m)$ and hence their simulations were vulnerable to artificial effects due to off-centring. Sellwood (2003) could reproduce their results, but if he suppressed the coefficients with odd $(l, m)$, the results were completely altered in that more than hundred instead of a few bar rotations were required to remove the density cusp. Sellwood also pointed to a previous study by White (1983) which reported that the $l = 1$ term can cause numerical artifacts. Thus, apparently, the simplified bar-halo model of Weinberg & Katz (2002) undergoes a non-reflection symmetric evolution. However, owing to the problems inherent to the use of spherical harmonics, it is unclear from the studies of Weinberg & Katz and of Sellwood whether this evolution is an artifact of their algorithm (i.e. the usage of a potential expansion in spherical harmonics) or not.

In an effort to resolve this question I performed similar simulations with gyrfalcON, which does not rely on an expansion in spherical harmonics about the origin and hence avoids these problems. Technical details of the models used are given in section 3.2, my results are presented in section 3.3. I discuss the results and draw conclusions in section 3.4.

### 3.2 Modelling

The models used by Weinberg & Katz (2002) were deliberately designed to be very simple. I use the same simulation set-up, which consists of an $N$-body representation of a CDM halo, and an external potential to represent a rotating bar with its centre at the initial centre of the halo.

For the vast majority of simulations I model the halo as an isotropic spherical Hern-
Halo evolution in the presence of a disc bar: 3.2 Modelling

Hernquist (1990) model, which has density

$$\rho(r) = \frac{M_{\text{halo}} r_s}{2\pi r(r_s + r)^3},$$

(3.1)

with $M_{\text{halo}}$ and $r_s$ the total mass and scale radius, respectively. This profile has the same ‘cuspy’ behaviour as the NFW profile (equation 1.1) in the limit $r \to 0$, but has the virtue of having less mass in the outer regions of the halo. This means that a simulation with the same particle number will have greater resolution in the cusp.

I introduce the bar as an external potential applied to the $N$-body simulation. Only the quadrupole moment of the bar is included in the force calculation, since the monopole term would have the effect of adding mass at the centre of the halo, altering the equilibrium. The higher order terms have little effect on the evolution (Hernquist & Weinberg 1992), and are ignored for simplicity, as they were in Weinberg & Katz (2002), and Sellwood (2003). A convenient fitting formula, which behaves properly as a quadrupole for both $r \to 0$ and $r \to \infty$ is provided by Hernquist & Weinberg (1992):

$$\Phi_{\text{bar}} = -\frac{GM_{\text{bar}}}{a} \frac{\alpha r_s^2}{(\beta r_s^2 + r^2)^{5/\gamma}} \sin^2 \theta \cos 2(\phi - \Omega_p t).$$

(3.2)

Here, $r_s = r/a$ with $a$ the semi-major axis of the bar, $r$, $\theta$, and $\phi$ are the usual spherical polar coordinates, and $M_{\text{bar}}$ and $\Omega_p$ are the mass and pattern speed of the bar, respectively. $\alpha$, $\beta$ and $\gamma$ are dimensionless parameters determined by best fit to the chosen bar density profile and shape. The bar is ‘turned on’ adiabatically over $\sim 10$ rotation periods to minimise transient effects.

In the vast majority of cases I follow Hernquist & Weinberg (1992) and Sellwood (2003) and model the bar as an $n = 2$ Ferrers (1877) bar with axis ratio $1:0.5:0.1$. This is fitted by the parameters $\alpha \simeq 0.1404$, $\beta \simeq 0.4372$ and $\gamma = 2$.

The simulations of Weinberg & Katz (2002) were of a NFW density profile halo (equation 1.1), truncated at large radii (as the NFW model has infinite mass, if considered for $r \to \infty$). The bar density profile used was that of a homogeneous ellipsoid of axis ratio $1:0.5:0.05$. This is well fitted by the same fitting formula used for the $n=2$ Ferrers bar (equation 3.2), with the parameters $\alpha \simeq 0.1227$, $\beta \simeq 0.6288$ and $\gamma \simeq 5.314$.

In Section 3.3.2 I present simulations with the same parameters as Weinberg & Katz (2002) in an effort to compare my simulations with theirs. The truncated NFW halo was defined as having the density

29
Halo evolution in the presence of a disc bar: 3.3 Results

\[ \rho(r) = \frac{\rho_c}{(r/r_s)(1 + r/r_s)^2} \text{sech}(r/r_t), \quad (3.3) \]

with a truncation radius \( r_t = 10r_s \).

3.2.1 Technical details

I generate the initial positions and velocities from the density profile (equation 3.1 or 3.3) with an isotropic distribution function. In the case of the Hernquist halo this is known (Hernquist 1990), while in the case of the truncated NFW halo it is found numerically using Eddington’s formula (equation 1.10). Quasi-random numbers were employed in order to suppress Poisson noise. The \( N \)-body simulations were performed using gyrfalcON.

For the Hernquist halo I used units of time, mass, and length such that \( G \equiv 1 \), \( r_s \equiv 1 \) and \( M_{\text{halo}} \equiv 1 \). For the truncated NFW halo I used units such that \( G \equiv 1 \), \( r_s \equiv 1 \), and \( v_{\text{circ, max}} \equiv 1 \). The equations of motion were integrated using the leap-frog integrator with minimum time step \( 2^{-7} \) and a block-step scheme allowing steps up to four times larger. Individual particle time steps were adjusted in an (almost) time-symmetric fashion such that on average

\[ \tau_i = \min \left\{ \frac{0.01}{|\alpha_i|}, \frac{0.01}{|\Phi_i|} \right\}, \quad (3.4) \]

with \( \Phi_i \) and \( \alpha_i \) the gravitational potential and acceleration of the \( i \)th body. A fiducial simulation with \( N = 3 \times 10^5 \) and no imposed bar conserved energy to within 0.2% over 1500 time units corresponding to \( \sim 120 \) bar rotations.

Throughout this chapter lengths will be quoted in terms of halo scale lengths.

3.3 Results

Initial simulations of a Hernquist model halo used \( N = 300,000 \) and a bar of length (semi-major axis) \( a = 0.7 \). Bar mass was 30% of the halo mass interior to \( r = a \), as it was for all simulations in this halo type. The bar rotates with a fixed pattern speed with corotation at \( r_s \), in this experiment and all others. These experiments showed a rapid change in the 1% Lagrange radius (the radius within which 1% of the total mass of the halo is contained, Fig. 3.1). This only demonstrates that the density at the origin was reduced by the action of the bar, which is not necessarily the density peak of the halo. The spherically averaged density profile (Fig. 3.2) seems to show the formation of a core
Halo evolution in the presence of a disc bar: 3.3 Results

**Figure 3.1.** Time evolution of the 1% Lagrange radius in flawed simulation.

after just 8 bar rotations. Neither of these measures take into account any possible change in the position of the centre of the distribution. Similar simulations with $N$ ranging from $10^4$ to $10^6$ showed essentially the same behaviour.

These results are in agreement with the apparent rapid evolution seen in simulations including the $l = 1$ term by Sellwood (2003), though the timescale is slightly longer. It is similar to the evolution observed by Weinberg & Katz (2002), though the halo model is different (see Section 3.3.2).
FIGURE 3.2. Spherically averaged density profile (inner part) for the same simulation as in Fig. 3.1. Shown are the initial profile (dashed); and the profile after 8 bar rotations, considered about the original centre (solid), or about the centre of the cusp (dotted). The profile appears to be cored after 8 bar rotation times but this is because the implicit assumption of conserved (or at least approximately conserved) spherical symmetry is incorrect, see Fig. 3.3. The profile considered about the centre of the cusp shows that the cusp is as strong as ever, but the force applied by the bar has become wildly unphysical, so this is result is, in any case, insignificant.
3.3 Results

3.3.1 Unstable evolution

This reduction in density is, however, not demonstrative of the creation of a cored density profile. Simply plotting the particle positions in the $x$-$y$ plane (Fig. 3.3) shows that within a few bar rotations the densest part of the halo is no longer at the origin. This creates the appearance of a cored density profile if analysis is performed with respect to the origin. This effect occurs even if the initial distribution of bodies is made exactly symmetrical initially (for each body at $x, v$ there is another one at $-x, -v$).

This off-centring means that these simulation results are invalid. The bar’s centre is fixed to the origin and is not allowed to react to the off-centring of the halo. This leads to a wildly unphysical interaction between the bar and the halo which rapidly transports energy and angular momentum to the halo, and invalidates all results found this way. I strongly suspect that the rapid evolution reported by Weinberg & Katz (2002) and its $l = 1$ dependence seen by Sellwood (2003) are actually this off-centring. Since this effect is reproduced in my simulations, it is clear that it is a real instability of the simplified bar-halo system investigated, and is not entirely due to the artifact of the spherical harmonic approach noted by Sellwood (2004).

The initial conditions with the applied quadrupole potential are entirely point symmetric about the origin. However they are unstable to motion of the cusp away from the centre of the bar. This occurs even when the initial particle distribution is defined to be exactly symmetric, because slight asymmetries caused by, for example, truncation errors break the imposed symmetry.

The cause of this instability can be understood by considering the external quadrupole potential applied to the $N$-body system. Figure 3.4 shows the position of the central density in the frame of the rotating bar, overlaid upon a contour plot of the quadrupole potential. The cusp of the halo is initially at a saddle point of the external potential. This is unstable to a small movement of the cusp. One can think, in a simplistic picture, of the ‘cusp’ falling into a potential well. The restoring force due to the bulk of the halo is too weak to counter this. Also, note that the cusp does not fall to the exact centre of the well, which can presumably be thought of as being due to the Coriolis force.

3.3.2 NFW halo

In an effort to re-examine the results of Weinberg & Katz (2002), I ran simulations of an NFW halo with the same parameters they used. These simulations had a comparatively short, light bar with semi-major axis $a = 0.5r_s$; and a mass equal to 15% of the halo.
FIGURE 3.3. Positions in the $x$-$y$ plane of bodies with $|z| < 1$ after 8 bar rotations. The plot shows that the centre of the distribution has shifted from the origin, thus invalidating the simulation setup.
FIGURE 3.4. Motion of the density centre in the plane, with the frame co-rotating with the bar, during the first 11 bar rotations (solid line) for the same simulation as in previous Figures. The thick dashed line indicates the orientation of the bar, which rotates anticlockwise. Contours of the sum of the effective bar quadrupole potential in this frame and an approximation for the halo-bulk potential (halo without cusp, using a $\gamma = 0$ Dehnen (1993) model) are shown as dotted curves. Stationary points perpendicular to the bar are maxima, origin is a saddle point, other stationary points along the bar axis are minima.
mass within the bar radius. Since Weinberg & Katz (2002) suggested that high numerical resolution was required to accurately simulate the resonant dynamics, I modelled the halo with particles of varying mass, with the lowest mass particles in the inner regions where the important resonant processes occur. This allowed an increase in the resolution of the inner regions of the NFW halo by up to a factor of 5 for constant particle number without loss of stability.

Experiments utilising gyrfalcON, and with the effective value of $N$ varying between 30,000 and 6,000,000 showed no significant evolution of the halo under these conditions. The bar potential was too weak to destabilise the halo and move its centre. This is in contrast to simulations with $N$ between 10,000 and 10,000,000, run with the potential expansion codes of Weinberg & Katz, and Sellwood which did become unstable rapidly (within $\sim 10$ bar rotations). The very low $N$ at which this instability is observable in the potential expansion case, and the absence of instability even in the high $N$ gyrfalcON case, indicates that this difference is not caused by particle noise “drowning out” the effect of the bar in the case presented here. The potential expansion approach is clearly more susceptible to the instability of the system than gyrfalcON.

It should be noted that there are two separate instabilities. The genuine instability of the artificial bar-halo system was responsible for the shift of the cusp away from the origin in the simulations with the Hernquist halo. In the NFW case the bar is too weak to destabilise the halo when gyrfalcON is used, but using a potential expansion approach exacerbates the problem because of the $l = 1$ instability referred to by Sellwood.

### 3.3.3 Preventing instability

#### Forcing symmetry

In the simplified model, both the halo and the bar are entirely point symmetric about the origin. The particle distribution would remain so for all time (thus preventing the instability) if it were not for truncation errors in the properties of the particles as determined during the simulation. In order to investigate the situation without the confusion caused by the instability, I ran a number of simulations in which the distribution of bodies was kept point symmetric about the origin at all times. To this end the bodies were treated as pairs, and for each pair $a$ and $b$, we enforced that $x_a + x_b = 0$ and $v_a + v_b = 0$ in the initial conditions and after each time step.

This is effectively the equivalent of simulations using an expansion in spherical harmonics which exclude all the terms with odd $(l, m)$. Such experiments were carried out
As Figure 3.5 shows, the ability of a rotating bar to remove a cusp is genuine, and not reliant upon instability. Even when the simulation is completely symmetric, angular momentum transferred from the bar to the halo eventually removes the cusp. This process is far slower than the unstable evolution seen in section 3.3.1. The dependence of the effect on \( N \) and bar strength are comparable to the results obtained by Sellwood (Figure 3.5). The major difference to note is that my results do not yet seem to converge at high \( N \). This is likely caused by the force solver I used allowing stochastic jitter in the body distribution to be passed to the forces more readily than a method involving only low-order spherical harmonics does.

Bars with mass

An alternative method for stabilising the simulation is to allow the bar to “have mass”, and thus provide a restoring force on the halo. This was achieved in my simulations by including the monopole component of the bar potential in the force calculation. The initial distribution function of the halo with the additional monopole potential can be determined using Eddington’s formula, by taking the density \( \rho \) in equation 1.10 to be the density of the halo, and the reduced potential \( \Psi \) to be the reduced potential of the halo and additional monopole combined.

In these simulations the monopole component of the potential was included at all times, the quadrupole was added adiabatically as in all previous simulations. Any off-centring of the halo causes an artificial transfer of linear momentum from the bar to the halo, so in addition the bar was constrained to move in such a way that the linear momentum of the bar-halo system was conserved.

Figure 3.5 shows a comparison of the results from this approach to those from the symmetrised simulations. The movement required of the bar in order to conserve the linear momentum of the system was extremely small, and only carried the bar a distance \( \sim 10^{-5} r_s \), which is several orders of magnitude smaller than the movement of the cusp seen in Figure 3.4. Simulations which included the monopole in the bar potential, but in which the bar was pinned to the origin produced results which were nearly indistinguishable from those with a moving bar. It is clear that the restoring force provided by inclusion of the monopole term is sufficient to prevent the growth of the instability in this case. The presence of the monopole and removal of the artificially enforced symmetry prevent the rapid “runaway” evolution seen at \( t \simeq 170 \) in the corresponding symmetrised simulations.
Figure 3.5. The 1% Lagrange radius plotted against time for many different simulations. Top: the effect of increasing bar size with fixed $N = 300,000$ and enforced symmetry. Bar mass was defined as 30% of the halo mass within the bar radius for all cases. Also shown are the results of a fiducial run with $N = 300,000$ and no imposed bar. Bottom: the effect of increasing resolution (particle number), for fixed bar size $a = 0.7$ and enforced symmetry also showing (Dotted) simulation with conserved linear momentum with same bar size, and $N = 1,000,000 (*)$
Halo evolution in the presence of a disc bar: 3.4 Conclusions

simulation. However the bar still removes the cusp of the halo over approximately the same timescale in the two simulations.

3.3.4 Angular momentum conservation

These experiments are deliberately kept as simple as possible, which leaves them unrealistic. One major difference between the simulations described thus far and the true situation is the apparent infinite supply of angular momentum for the bar, which keeps it rotating with the same pattern speed for the entirety of the simulation, however much angular momentum has been transferred to the halo.

Weinberg & Katz (2002) pointed to a “suite of simulations with a slowing bar, whose pattern speed follows from the conservation of angular momentum of the combined bar-halo system” as further evidence pointing to the effectiveness of this process in removing the cusp. However these results are tainted by the instability of the system, so I repeated the experiments with a slowing bar and the symmetrisation method to prevent instability. This is somewhat similar to experiments performed by Sellwood (2003), with the $l = 1$ spherical harmonic suppressed. Since initially the bar has zero mass, and the mass (and thus angular momentum) is increased adiabatically, angular momentum is introduced to the system by the growing bar. I calculate the angular momentum of the halo after each block-step and use that to calculate the change in the pattern speed of the bar at every point, so that the total angular momentum added to the halo is no more than that of the bar at its full mass, rotating with its original pattern speed.

As Figure 3.6 shows, when the bar is not given an infinite supply of angular momentum, the situation is somewhat different. Essentially all the angular momentum of the bar is absorbed by the halo within 15 initial rotation periods (in fact the bar has made ten complete revolutions at that point). This absorbed angular momentum is insufficient to remove matter from the cusp to any great extent. In experiments carried out without symmetrisation, the central density did fall to approximately that seen in the experiments with a bar rotating with a constant pattern speed. This was again because the cusp of the distribution shifted from the origin.

3.4 Conclusions

I find that over-simplifying the interaction of a barred galaxy with a cusped CDM halo may lead to substantial over-estimation of the reduction in central density. The cause is
FIGURE 3.6. Density profile for a slowing bar in a $N = 300,000$ representation of a Hernquist halo initially (dashed) and after 16 initial bar rotation periods (solid). The bar has not gone through 16 rotations, because it has decelerated, and by this point is rotating at 10% of its initial pattern speed.
an instability of the simplified model: a slight off-centring of the CDM cusp from the origin results in a net force from the bar, which is fixed to stay centred, further driving the cusp away from the origin. The instability is implicit in the model and is not simply a result of determining the potential as a spherical harmonic expansion about the origin, though that approach does seem to exacerbate the problem. It seems very likely that the rapid evolution reported by Weinberg & Katz (2002) and Sellwood (2003, when odd $l$ terms where included in the computation of the forces) were an artifact of this instability in conjunction with averaging over spherical shells concentric with the origin (rather than the density centre).

If this artificial instability is suppressed, either by enforcing reflection symmetry with respect to the origin, or by including the monopole of the bar potential and conserving linear momentum, the removal of the CDM cusp is a very slow process, requiring $\mathcal{O}(100)$ bar-rotations, which represents (at least) a significant fraction of a Hubble time. Moreover, this simplistic model contains an additional artifact: the infinite supply of angular momentum from the ever torquing bar. If this is removed by controlling the total angular momentum, the cusp is at most slightly modified. This means that, under these conditions, the amount of angular momentum needed to remove a CDM cusp is far larger than that present in the inner parts of disc galaxies.

Of course, this work makes a large number of assumptions. The halo has been assumed to be spherically symmetric, isotropic, without substructure, isolated and initially in equilibrium. The bar was assumed to be rigid, and gas physics and bar-disc interaction were neglected. These are the same assumptions made by the paper that proposed this mechanism (Weinberg & Katz 2002). These simulations cannot tell the whole story, but they do provide an insight into the fundamental physical mechanism.

I conclude that, contrary to the original proposal, these simulations suggest that angular momentum (and energy) transport from a disc-bar is not an effective way to destroy CDM cusps, in agreement with studies using less simplistic models (Valenzuela & Klypin 2003, Athanassoula 2004). If galactic scale CDM haloes truly are cored, it seems that some other process is responsible.
Chapter 4

Initial conditions for simulations of multi-component galaxies

“All models are wrong, but some are useful”

George E. P. Box
In this chapter I present a very general recipe for constructing N-body realizations of galaxies comprised of more than one component. I go on to implement and test this model. The equilibrium halo (and bulge) distribution function is determined exactly with the monopole of the disc component’s potential included in the calculation. The full disc potential is then adiabatically grown from the monopole. The disc component is occupied by particles drawn from a distribution function determined in a similar way to that described in Dehnen (1999b). This recipe avoids the need for the “local-Maxwellian” approximation.

4.1 Introduction

Generating an equilibrium $N$-body representation of a multi-component galaxy is of importance for any number of applications, for example the study of bars (Debattista & Sellwood 2000, Athanassoula 2002), warps (Ideta et al. 2000) and galaxy mergers, both minor (Mihos et al. 1995, Walker, Mihos & Hernquist 1996) and major (chapter 5 Heyl, Hernquist & Spergel 1996, Naab, Burkert & Hernquist 1999). The typical method used in the past is that of Hernquist (1993), which uses a Maxwellian approximation. This method is not rigorous and Hernquist (1993) suggested that “In the future, it will likely be necessary to refine the basic approach as computer hardware and software permit simulations with particle numbers significantly in excess of those discussed here.” (49,152 particles were used in his empirical tests).

This work is motivated by the growing realisation that this approach is no longer appropriate as simulation quality improves. Kazantzidis, Magorrian & Moore (2004) looked at $N$-body haloes (with no disc component) and compared the properties of those modelled using the Maxwellian approximation with those in which the distribution function was found exactly, through Eddington inversion (section 1.5.1). They showed clearly that the haloes formed using the Maxwellian approximation were out of equilibrium and, as an example, showed that the difference was important in the tidal stripping of substructure in CDM haloes.

Eddington inversion only works for spherically symmetric systems, so cannot be straightforwardly applied to the problem of finding the self-consistent distribution function of the multi-component system.

In section 4.2 I provide an outline of other methods of creating multi-component galaxies, focusing on the work of Hernquist (1993); in section 4.3 I describe my new method in detail and in section 4.4 I test this new method.
4.2 Existing methods

Barnes (1988) introduced a method for constructing a multi-component galaxy which he referred to as being “rather ad hoc”. He constructed separate equilibrium $N$-body King (1966) models for the halo and bulge, then superposed them and allowed them to relax into a new equilibrium in a simulation over several dynamical times. Next the potential field of the disc was slowly imposed on the halo, this causes both a flattening of the halo and bulge, and a radial contraction. Finally the disc was populated with particles.

The method of Hernquist (1993) has been enormously successful and influential. It superseded the method of Barnes (1988) because it makes it possible to specify the density and velocity profiles of the various components in a straightforward way.

First, the positions of all the particles, in all components, are determined. This can be carried out trivially since the desired density profiles are known, and can be used as probability distributions.

To find particle velocities for the halo component, the approximation is made that the mass distribution is spherical, and that the velocity distribution is isotropic, (therefore $v_r^2 = v_\theta^2 = v_\phi^2$). Thus it is possible to use the one of the Jeans equations (e.g. Binney & Tremaine 1987) to determine the velocity dispersion in the halo through integrating

$$\frac{d(\rho_h v_r^2)}{dr} = -\rho_h \frac{d\Phi}{dr},$$

and recognising that $\Phi$ is the potential of the whole system. This leads to the result (equation 2.14 in Hernquist 1993)

$$\overline{v_r^2}(r) = \frac{1}{\rho_h(r)} \int_r^\infty \rho_h(r') \frac{GM(r')}{r'^2} dr',$$

where $r'$ is a dummy variable, and $M(r)$ is the cumulative mass distribution.

This can be solved through numerical integration, under the approximation that the bulge and disc components can be binned in spherical shells, which are then used in computing $M(r)$. In practice Hernquist made an ad-hoc reduction in the disc radius of $\sim 35\%$ in computing the disc’s contribution to $M(r)$. This was in an effort to reduce the redistribution of the halo mass that was found when the galaxy was allowed to evolve dynamically.

This provides the velocity dispersion, but not the shape of the velocity distribution. It is here that the local-Maxwellian approximation is used. The speeds of the particles
are chosen from
\[ F(v, r) = C_{\text{norm}} v^2 \exp(-v^2/2v_r^2), \] (4.3)

where \( C_{\text{norm}} \) is a normalisation constant. Athanassoula & Misiriotis (2002) used an alternative functional form with \( F(v, r) \propto (1 - v^2/C_v^2) \) and \( F = 0 \) for \( v > C_v \). This was in an effort to reduce the number of particles given velocities greater than their local escape velocity.

Hernquist (1993) also demonstrates how to find velocity dispersions for the disc and bulge. These are axisymmetric, so the dispersions are found in cylindrical coordinates. For each particle \( v_R, v_\phi \) and \( v_z \) are drawn independently from Gaussians of standard deviation \( \sqrt{v_R^2}, \sqrt{\sigma_\phi^2} \) and \( \sqrt{v_z^2} \) respectively. Only the random component of the azimuthal motion is found in this way, as the disc has a non-zero azimuthal streaming (rotational) velocity \( \overline{v_\phi} \), which is also added.

Boily, Kroupa & Peñarrubia-Garrido (2001) extended this approach to include a non-symmetric halo, but maintained the approximation to a Gaussian or Maxwellian distribution for the particle velocities. This approximation (or that of Athanassoula & Misiriotis) is convenient, but inadequate for a number of applications (e.g. Kazantzidis, Magorrian & Moore 2004).

Another alternative method is provided by Kuijken & Dubinski (1995), and expanded upon by Widrow & Dubinski (2005). They determine an analytical form for equilibrium distribution functions for each component in isolation, written in terms of energy and angular momentum. In the case of the disc, the distribution function is also a function of the “vertical energy” \( E_z = \frac{1}{2}v_z^2 + \Phi(R, z) - \Phi(R, 0) \). These distribution functions are found for a predetermined isolated density distributions. In the case of Kuijken & Dubinski (1995) these were a King model bulge (King 1966), a lowered Evans model halo (Kuijken & Dubinski 1994) and an exponential disc (see section 4.3.3). The combined model is then found by using these distribution functions in the combined gravitational potential of the entire model.

The major problem with this approach is summed up by Widrow & Dubinski (2005) when they state that the combined model constructed in this way “may bare [sic] little resemblance to the corresponding isolated components, a situation which is cumbersome for model building”. This approach is also rather inflexible.
4.3 The new method

My method bears some relation to that of Barnes (1988), in that the non-spherically symmetric component (the disc), is grown adiabatically within the halo. The bulge is assumed to be spherically symmetric, though it would be a relatively straightforward step to include a non-spherical bulge in a similar way to the disc. The process has three stages:

1. Creating an equilibrium initial $N$-body representation of the spherically symmetric components of the galaxy (the halo and bulge) in the presence of an external potential field corresponding to the monopole (spherical average) of the desired disc potential.

2. Evolving the $N$-body system by growing the non-monopole components of the potential adiabatically in a simulation using gyrfalcON, allowing halo and bulge particles time to relax into the new potential.

3. Replacing the external potential field representing the disc component with an $N$-body representation.

4.3.1 Creating the initial $N$-body halo/bulge

First I create an $N$-body realization of the spherically symmetric components of the system using an extension of Eddington’s approach derived by Cuddeford (1991). This is a further extension from the Osipkov-Merritt model (section 1.5.2).

Again, one typically works in terms of the the “relative potential” $\Psi \equiv -\Phi + \Phi_0$ and the “relative energy” $\varepsilon \equiv -E + \Phi_0 = \Psi - \frac{1}{2}v^2$, where $\Phi_0$ is some constant defined such that $f(\varepsilon \leq 0) = 0$ (so in practice $\Phi_0 = 0$ is always used here).

Cuddeford considered distribution functions of the form

$$f(\varepsilon, L) = f_0 \left( \varepsilon - \frac{L^2}{2r_a^2} \right) L^{2\alpha} = f_0(Q)L^{2\alpha},$$

(4.4)

where $Q$ and the anisotropy radius $r_a$ are familiar from Osipkov-Merritt models, and where $\alpha > -1$ and real. For a specified spherically symmetric density profile $\rho(r)$ in a relative potential $\Psi(r)$ this choice has the general solution

$$f_0(Q) = \frac{\sin(n-1/2-\alpha)\pi}{\pi \lambda(\alpha) \eta(\alpha)} \frac{\mathrm{d}Q}{\mathrm{d}\Psi} \int_0^Q \frac{d^n\rho_{\text{cud}}}{d\Psi^n} \frac{d\Psi}{(Q - \Psi)^{\alpha+3/2-n}},$$

(4.5)
where \( n \) is defined as the largest integer equal to or less than \( \alpha + \frac{3}{2} \), and where

\[
\eta(\alpha) = \begin{cases} 
(\alpha + \frac{1}{2})(\alpha - \frac{1}{2})...(\alpha + \frac{3}{2} - n) & \alpha > -\frac{1}{2} \\
1 & -1 < \alpha \leq -\frac{1}{2},
\end{cases}
\]  

(4.6)

and \( \lambda \) is

\[
\lambda(\alpha) = 2^{\alpha+3/2} \pi^{3/2} \frac{\Gamma(\alpha + 1)}{\Gamma(\alpha + 3/2)},
\]  

(4.7)

with the standard Gamma function \( \Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt \). The “reduced density” \( \rho_{\text{cud}} \) is

\[
\rho_{\text{cud}} = \frac{(1 + r^2/\alpha a)^{\alpha+1}}{r^{2\alpha}} \rho.
\]  

(4.8)

This distribution function produces a spherically symmetric system with a velocity distribution such the \( \beta \) parameter (equation (4.9)) is given by

\[
\beta(r) = \frac{r^2 - \alpha r_a^2}{r^2 + r_a^2}.
\]  

(4.9)

As one would expect, in the case where \( \alpha = 0 \) this becomes the anisotropy corresponding to an Osipkov-Merritt model. In the case where \( r_a \to \infty \), the anisotropy of the halo is the same at all radii, \( \beta = -\alpha \).

This approach has the advantage that the distribution function is exact for a spherically symmetric system, and thus remains in equilibrium, maintaining its original density profile. At no point is the assumption of a Maxwellian velocity distribution made. The only restrictions to this method are that \( \Psi = \Psi(r) \) is a monotonic function of \( r \); that \( \rho = \rho(r) \) and that the solution to equation (4.5) must be physical, i.e. \( f_0(Q) > 0 \). This allows me to use a wide range of different halo (or bulge) density profiles (see, for example, section 5.2.2).

It is extremely useful that the derivation of equation (4.5) does not make the assumption that the potential of the system is that due to the density profile through the Poisson equation (equation 1.6). This means that it is straightforward to generalise this approach to find the distribution function of a spherically symmetric component of a larger, spherically symmetric, system. The term \( \rho_{\text{cud}} \) in equation (4.8) is replaced by \( \rho_{\text{cud},i} \), the reduced density of the component \( i \); the term \( \Psi \) in equation (4.8) always refers to the relative potential of the entire system.

In the model the presence of the disc component is taken into account when calculating the distribution function of the halo and bulge. It is impossible to do this exactly using this method, as the disc is not spherically symmetric.
The best approximation available using this method is to use the spherical average of the disc potential. Then the distribution functions of the bulge and halo are found in the total potential of the halo, bulge and ficticious spherically averaged disc. This prevents the radial contraction seen in the models constructed by Barnes (1988).

4.3.2 Evolving the halo and bulge to adapt to the disc

The second stage of creating the model is to evolve the particle distribution of the halo and bulge so that they are in equilibrium with the full disc potential, rather than with its spherical average. I use gyrfalcON for this, though the axisymmetry of the system means that an SCF code (section 2.4) would be well suited to this purpose.

The full potential of the disc, $\Phi_{\text{disc}}(x)$ is grown from its spherical average $\Phi_{\text{disc},0}(r)$ according to the formula

$$
\Phi(x, t) = \Phi_{\text{disc},0}(r) + A(t) \left[ \Phi_{\text{disc}}(x) - \Phi_{\text{disc},0}(r) \right],
$$

where $A(t)$ is a growth factor that goes from 0 at $t = 0$ to 1 smoothly in a timescale, $t_{\text{grow}}$, far longer than the dynamical time in the region of the disc. The halo and bulge are then allowed further time to relax completely under the influence of the disc potential.

This process causes changes to the density distribution of the halo. The spherically averaged density distribution remains the same, but the halo and bulge are somewhat flattened, and the iso-density contours of both the halo and bulge become oblate spheroids. The degree of flattening is dependent on the details of the various components. Details for a typical case are given in Section 4.4.

This stage is by some distance the slowest in the creating of the initial conditions. However it is still takes a relatively small fraction of the total CPU time of almost any scientifically interesting simulation. It should also be noted that up until this point the velocity distribution of the disc has not been a factor in the calculations, so the halo and bulge created by this process can be re-used in simulations with identical disc density profiles, but different kinematics.
4.3.3 Populating the disc

I start from the assumption that the distribution function of the disc can be decomposed into its components in the plane of the disc, and perpendicular to it, i.e.

$$f_d \simeq f_d(E_p, L_z, E_z),$$

(4.11)

where the “vertical (z-component) energy” $E_z = \frac{1}{2}v_z^2 + \Phi(R, z) - \Phi(R, 0)$, and the “planar energy” $E_p = E - E_z = \frac{1}{2}(v_R^2 + v_\phi^2) + \Phi(R, 0)$. Because the potential of the halo and bulge must be determined from the evolved system, it is necessary to approximate the potential of the $N$-body distribution as being cylindrically-symmetric. To do this I find the potential of the system as an expansion about the centre of the galaxy in spherical harmonics. This is the same method used in the SCF force calculation method. Ensuring cylindrical symmetry corresponds to using only $m = 0$, $l$ even terms in the expansion.

For the planar components of particle position and velocity I broadly follow the procedure outlined in Dehnen (1999b) for populating the disc using the distribution function described in that paper as $f_{\text{new}}$.

The simplest distribution function to use is that of a completely cold disc, in which all particles are on circular orbits. This can be written (e.g. Dehnen 1999b) as

$$f_{\text{cold}}(E_p, L_z) = f_{\text{circ}}(E_p) \delta(E_p - E_{p,c}(L_z)) = \frac{\Omega(R_{Lz}) \Sigma(R_{Lz})}{\pi \kappa(R_{Lz})} \delta(E_p - E_{p,c}(L_z)),$$

(4.12)

where $E_{p,c}(L_z)$ is the energy of a circular orbit in the galaxy midplane with angular momentum $L_z$. $R_{Lz}$ is the radius of that orbit, $\Omega$ its circular frequency and $\kappa$ its epicycle frequency. $\Sigma(R)$ is the surface density of the disc.

The most common choice for a warm disc distribution function is to replace the $\delta$-function in equation 4.12 with an exponential in energy, which corresponds to a Maxwellian velocity distribution (e.g. Shu 1969, Kuijken & Dubinski 1995). In this case the distribution function $f_{\text{shu}} \propto \exp((E_{p,c}(L_z) - E_p)/\sigma_R^2)/\sigma_R^2$.

The distribution function used in my models is based upon the alternative description of the cold disc distribution function $f_{\text{cold}}(E_p, L_z) = f_{\text{circ}}(E_p) \delta(\Omega(R_{Ep})[L_{z,c}(E_p) - L_z])$, where $R_{Ep}$ is the radius of a circular orbit in the disc midplane with planar energy $E_p$, and $L_{z,c}(E_p)$ is its angular momentum.

As before, the warm disc distribution function is found by replacing the $\delta$-function
with an exponential, giving

\[ f_{\text{disc,p}}(E_p, L_z) = \frac{\Omega(R_{Ep}) \Sigma'(R_{Ep})}{\pi \kappa(R_{Ep}) \sigma_R^2(R_{Ep})} \exp \left( \frac{\Omega(R_{Ep}) [L_z - L_{z,c}(E_p)]}{\sigma_R^2(R_{Ep})} \right), \]  

(4.13)

where \( \Sigma'(R) \) and \( \sigma_R^2(R) \) are defined such that the true surface density \( \Sigma(R) \) and radial velocity dispersion \( \sigma_R^2(R) \) of the \( N \)-body representation are those desired, to within an appropriate degree of accuracy.

Dehnen argued that equation (4.13) was a more useful distribution function than that of Shu (1969). He argued that the warming that occurs in a disc can be described as an exponential in the radial action \( J_R = \frac{1}{2\pi} \oint p_R dR \), where \( p_R \) is the generalised momentum (\( p_R \equiv \partial L / \partial \dot{R} \), with \( L \) the Lagrangian), and the integral is over one complete radial oscillation.

The radial action is more closely approximated by \( \Omega(R_{Ep})(L_{z,c}(E_p) - |L_z|) / \omega_R \) than by \( (E_p - E_{p,c}(L_z)) / \omega_R \), where \( \omega_R \) is the frequency of radial oscillations (Dehnen 1999a). More practically, this distribution function has the advantage that the value of \( R_{Ep} \) is generally a far better approximation to the mean radius of an orbit than \( R_{Lz} \), which ensures that \( \Sigma(R) \) and \( \sigma_R^2(R) \) closely resemble \( \Sigma'(R) \) and \( \sigma_R^2(R) \). This choice of distribution function also extends to negative \( L \), unlike that of Shu. This allows for a “tail” of counter-rotating stars.

When sampling from this distribution function, I attempt to minimise noise in the particle distribution by sampling points more regularly than at random in phase space. I accomplish this by sampling orbits from the target density distribution, then placing (approximately, see below) \( N_{sam} \) particles at points on the orbit, where \( N_{sam} \ll N_d \) (the total number of disc particles).

To sample from this distribution function I choose a radius \( R \) randomly from the cumulative distribution

\[ P(R) = \int_0^R R' \Sigma(R') dR', \]  

(4.14)

and determine \( E_p = E_{p,c}(R) \), the planar energy of a circular orbit in the disc midplane with radius \( R \). I then choose \( \xi \in (0, 1) \) randomly and determine

\[ L_z = L_{z,c}(E_p) + \sigma^2(R) \ln \xi / \Omega(R). \]  

(4.15)

Providing that these values are not unphysical (i.e. \( |L_z| < L_{z,c}(E_p) \)) I integrate the orbit with these values of \( E_p \) and \( L_z \) to find the values of \( R \) and \( \dot{R} \) as a function of time over one complete radial oscillation period \( T_R \). From this I compute the fre-
Initial conditions for simulations of multi-component galaxies: 4.3 The new method

quency of radial oscillations \( \omega_R(E_p, L_z) \equiv 2\pi/T_R \), and evaluate the “correction factor”
\[ g_{\text{corr}} = \kappa(R)/\omega_R(E_p, L_z) \]
I then sample \( N_{\text{sam}} \) phase-space points \( \{ R_i, \phi_i, \dot{R}_i, \dot{\phi}_i \} \) randomly for this orbit. Here \( N_{\text{sam}} \) is either of the two integers next to \( g_{\text{corr}} N_{\text{orb}} \), chosen with probabilities such that the mean is equal to \( g_{\text{corr}} N_{\text{orb}} \).

To find these points only two random picks are required. I randomly find \( \phi_i \in [0, 2\pi] \) and \( \tau_i \in [0, T_R] \). The values of \( R(\tau_i) \) and \( \dot{R}_i(\tau_i) \) are found from the table made during orbit integration. \( \dot{\phi}_i(\tau_i) \) is then found from the fact that \( \dot{\phi}_i = L_z/R_i^2 \).

This approach has the flaw that the surface density and radial velocity dispersion profiles of the particle distribution produced are not those defined in equations 4.14 and 4.15, because I have (effectively) sampled in energy, rather than in particle radius. I mitigate this through an iterative scheme.

Consider three surface density profiles: \( \Sigma_{\text{tar}}(R) \), the target surface density; \( \Sigma_{\text{in}}(R) \), the surface density used as the input to eq. 4.14; and \( \Sigma_{\text{out}}(R) \), the surface density of the resulting particle distribution. My first estimate is to use \( \Sigma_{\text{in}} = \Sigma_{\text{tar}} \). I then use the above method and determine \( \Sigma_{\text{out}} \), sampling more points than will be used in the final particle distribution to reduce sampling error. I then make the replacement \( \Sigma_{\text{in}} \rightarrow \Sigma_{\text{in}}\Sigma_{\text{tar}}/\Sigma_{\text{out}} \), and use this in equation 4.14. Simultaneously I perform the same process for the radial velocity dispersion profiles \( \sigma_{\text{tar}}(R) \), \( \sigma_{\text{in}}(R) \) and \( \sigma_{\text{out}}(R) \). I repeat this iteration until the particle distribution has properties as close to the targets as I require and am able to achieve with such a simple scheme.

This process, while somewhat laborious, is relatively rapid on modern machines, taking less than a minute for a \( 10^6 \) particle disc on an ordinary desktop machine. It has the advantage of having a “quiet-start” (e.g. Sellwood 1987). The particles are arranged far more evenly in phase space than would be the case if they were simply placed at random, thus minimising shot-noise.

Once this is done I determine the vertical component of each disc particle’s position and velocity. I assume that the local structure in the z-direction corresponds to that of an isothermal sheet with a constant vertical scale height, \( z_d \) (Spitzer 1942). This leads to a density throughout the disc \( \rho(R, z) \propto \text{sech}^2(z/z_d) \). \( v_z \) is then drawn randomly from a normal distribution with \( \sigma_{z}^2 = \pi G \Sigma(R) z_d \), where \( \Sigma(R) \) is the disc surface density.
4.4 Testing

To test this approach I use a model based loosely upon the Milky Way as modelled by Klypin, Zhao & Somerville (2002). The disc is defined as having a surface density profile which is exponential in the cylindrical radius $R$, i.e. $\Sigma(R) \propto e^{-R/R_d}$. With a vertical structure modelled by isothermal sheets, this has a density distribution

$$\rho_d(R, z) = \frac{M_d}{4\pi R_d^2 z_d} \exp \left( -\frac{R}{R_d} \right) \text{sech}^2 \left( \frac{z}{z_d} \right),$$

where $M_d$ is the total disc mass, $R_d$ is the disc scale radius, and $z_d$ the scale height.
The halo model is a truncated NFW halo (equation 3.3), the bulge model has a Hernquist density profile (equation 3.1). I choose units such that Newton’s gravitation constant \( G = 1 \), \( R_d = 1 \), and \( M_d = 1 \). I take the disc scale height \( z_d = 0.1 \), the bulge mass \( M_b = 0.2 \), and bulge scale length \( r_b = 0.2 \). Scaling these values to the Milky Way, taking \( R_d = 3.5 \text{kpc} \), \( M_d + M_b = 5 \times 10^{10} M_\odot \) (as in Klypin et al. 2002) gives a time unit \( \simeq 14 \text{Myr} \), and thus a velocity unit \( \sim 250 \text{km s}^{-1} \).

For the halo I take the scale radius \( r_h = 6 \), the truncation radius \( r_t = 60 \), and the halo mass \( M_h = 24 \). 79\% of this total mass is within the truncation radius. The rotation curve for this model (with velocities given in code units) is shown in Figure 4.1.

In all tests in which they were populated, the halo was populated with 1200000 particles, the bulge with 40000 and the disc with 200000. That corresponds to each halo particle being 4 times more massive than a stellar particle. Particles had a smoothed density profile as per equation 2.3 with a smoothing length \( \varepsilon_s = 0.02 \) for disc and bulge particles, and \( \varepsilon_h = 0.04 \) for halo particles. The maximum force exerted by a single particle \( i \) is proportional to \( m_i / \varepsilon_i^2 \), so this choice of smoothing parameter ensures that the maximum force from all particles is the same.

Simulations were performed with gyrfalcON with a minimum time-step of \( 2^{-7} \), and a block-step scheme with largest timestep \( 2^{-4} \). Individual particle time-steps were adjusted such that on average the time-step

\[
\tau_i = \min \left\{ \frac{0.01}{|\mathbf{a}_i|}, \frac{0.05}{|\Phi_i|} \right\},
\]

(4.17)

with \( \Phi_i \) and \( a_i \) the gravitational potential and acceleration of the \( i \)th body.

The first tests were to ensure that the spherically symmetric components remain in equilibrium in the spherically-symmetrised potential. This was tested with various different \( r_a \) and \( \alpha \) (equation 4.4) in the halo. This was to ensure that Cuddeford inversion had been implemented correctly. The halo and bulge profiles remained consistent to within a few smoothing lengths of the centre, where smoothing and two-body relaxation have some small effect. This testing was also used to assess whether the time integration parameters were appropriate. Since energy was typically conserved in the simulation to within 0.05\% over 200 time units, it can be assumed that the time integration is sufficiently accurate. For all further tests I only consider the isotropic (\( \alpha = 0 \), \( r_a \rightarrow \infty \), so \( \beta = 0 \) throughout the halo) case, for simplicity.
4.4 Testing

The next step was to ensure that the growth of the full disc potential from the spherically-symmetrised potential occurs without significant effect upon the density profile of the halo or bulge, or upon their kinematics, and to quantify the effect upon the shape of the resultant density distribution. The disc potential is grown as per equation 4.10, with a total growth time $t_{\text{grow}} = 40$. This is substantially longer than the dynamical time $t_{\text{dyn}}$ in the vicinity of the disc (for instance at $R = 3$, $t_{\text{dyn}} \sim 6$).

Figures 4.2 & 4.3 show the density profiles of the halo and bulge respectively. Both
Initial conditions for simulations of multi-component galaxies: 4.4 Testing

**Figure 4.3.** Spherically averaged density profile of the bulge in the same simulation as figure 4.2, shown before growth of the full disc potential ($t = 0$), and at two times after that growth ($t = 100, 200$). The disc component was not populated in this simulation.

![Figure 4.3](image)

The halo density is slightly raised in the inner $\sim 0.02r_h = 0.12$ and the bulge density is slightly lowered in the inner $\sim 0.3r_b = 0.06$. In each case this corresponds to only a few smoothing lengths, and is approximately the same as that observed in simulations with no disc growth.

Figure 4.4 shows the velocity anisotropy parameter, $\beta$ as a function of radius for both the halo and bulge. $\beta$ was determined for particles in spherical shells. Both halo and bulge are initially isotropic and, to within numerical accuracy, remain so during and well after the growth of the full disc potential.
Initial conditions for simulations of multi-component galaxies:

4.4 Testing

Figure 4.4. Anisotropy parameter $\beta$ (equation 1.17) against radius for the halo and bulge in the same simulations as figures 4.2 & 4.3. $\beta$ was determined for particles in spherical shells. Initial conditions were defined as having $\beta = 0$, any variation from that at $t = 0$ is due to discreteness noise.
To find a quantitative measure of the expected flattening of the system, due to the growth of the full disc potential, I determine the axis ratios of the halo and bulge as a function of radius (Figure 4.5).

In order to determine the axis ratios of a system when they are expected to vary, it is necessary to order the particles in some variable, and then split the particles into groups and determine the axis ratios of each group individually.

The axis ratios of a group of particles is found from a diagonalisation of the moment of inertia tensor

\[
I_{i,j} = \sum_{\alpha} m_\alpha x_{i,\alpha} x_{j,\alpha},
\]

where the sum is over all particles considered, and \( x_{i,\alpha} \) is the \( i \) coordinate of the \( \alpha \)th particle. The eigenvalues of the moment of inertia tensor \( \lambda_a \geq \lambda_b \geq \lambda_c \) define the axis ratios through \( b/a = \sqrt{\lambda_b/\lambda_a}, \ c/a = \sqrt{\lambda_c/\lambda_a}; \) with \( a, b, c \) the major, intermediate, and minor axes respectively.

The most common method is to rank the particles in radius (e.g. Kazantzidis et al. 2004), but this suffers from a heavy bias, it tends to drastically overestimate \( b/a \) and \( c/a \), especially at small radii (Athanassoula, private communication). Sorting by potential is better, but still biased. The most accurate way to determine the axial ratios of a triaxial body is to sort the particles by local density, and this is what is shown in Figure 4.5. The radius of the group of particles (the density “shell”) is defined as the median radius of the particles within it.

I determine the local density for particles using the NEMO program hackdens, which uses the Barnes & Hut (1986) tree-code to speed up the calculation of density through a “nearest neighbours” approach (e.g. Casertano & Hut 1985). This approach estimates a particle’s local density by finding the distance to its \( K \)th nearest neighbour. In this case I use the 15th nearest neighbour.

Figure 4.5 shows that there is some small effect on the bulge and halo shape from the growth of the disc. In both cases the shortest \((c)\) axis is very close to the \( z \)-axis, which means that the naïve assumption that the halo and bulge would be squashed perpendicular to the disc by the growth of the full potential is correct. The degree of flattening is relatively small, with \( c/a \sim 0.8 \) in the most flattened shells. In the case of the halo this is within the inner \( \sim 0.3r_h = 1.8R_d \). In the case of the bulge the most flattened shells are at \( \sim 3r_b = 0.6R_d \). The bulge is less flattened at smaller radii because the disc is of finite thickness (scale height \( z_d = 0.1R_d \)), which means that the full potential is less flattened – when compared to the spherical average – in the inner parts of the bulge than in the outer. The same effect would be observable in the halo with sufficient resolution.
4.4 Testing

4.4.2 Testing the disc model

Before testing the full compound galaxy, I test the disc model in isolation. To do this I follow the evolution of the populated disc component in the static external potential $\Phi(R,z)$, the cylindrically-symmetrised potential in which the distribution function of the disc was found.

The main difficulty when investigating the stability of a disc model is that one wants it to be stable to axisymmetric perturbation, but not to non-axisymmetric instabilities, such as spiral waves and bar modes, both of which one would expect to see. These instabilities cause redistribution of the mass of the disc in both the $R$ and $z$ directions (e.g. Hohl 1971, Athanassoula & Misiriotis 2002).

In an effort to determine that the disc model is stable, I use a technique that is commonly used to prevent non-axisymmetric modes from growing. After every block-step in the force integration the position and velocity of each body in the disc is arbitrarily rotated about the $z$-axis by an angle chosen randomly in the range $\{0, 2\pi\}$, thus destroying any coherent non-axisymmetric perturbation.

The distribution function defined by equation 4.13 allows for a huge range of possible $\sigma_R$. In practice my implementation restricts the possibilities to either $\sigma_R \propto e^{-R/R_\sigma}$, or $\sigma_R$ is such that the Toomre (1964) stability parameter $Q$ is constant throughout the disc,
where
\[ Q \equiv \frac{\sigma R \kappa}{3.36G \Sigma}. \] (4.19)

This should not be confused with the Osipkov-Merritt \(Q\) (equation 1.11). A stellar disc is known to be unstable to axisymmetric waves if Toomre’s \(Q < 1\).

I test two models, one with constant \(Q = 1.2\), and one with \(\sigma_R \propto e^{-R/2R_d}\) (i.e. \(\sigma_R^2 \propto \Sigma\)), with the constant of proportionality defined such that \(Q(R = R_0) = 1.2\). The two give similar quality results, so only the results from the latter distribution function are shown (this makes it substantially easier to show the desired \(\sigma_R\)).

The surface density of the disc model is preserved to within the inner \(\sim 0.1 - 0.2R_d\), and out to beyond \(6R_d\) (figure 4.6). The radial velocity dispersion \(\sigma_R\) is preserved over a similar range.

I measure the disc thickness in terms of the root-mean-squared (r.m.s.) value of \(z\). For a \(\text{sech}^2\) disc of scale height \(z_d\), one expects \(z_{\text{r.m.s.}} \approx 0.91z_d\). The disc thickness remains near constant in the range \(0.3 \lesssim R \lesssim 4\), with a slight warming which can reasonably be attributed to particle smoothing. In the inner \(\sim 0.3R_d\) the disc becomes somewhat thinner. This is understandable because in the inner parts of the disc the approximation that the planar and vertical components of the distribution function can be decoupled (equation 4.11) is particularly bad. The outer parts of the disc \((R > 4)\) are also slightly thinner. The outer disc is sparsely populated, which may well contribute to this.

### 4.4.3 Testing the full model

#### A constrained model

Finally I test the complete compound model. First I wish to examine the behaviour of the system in the absence of bar instabilities. In order to do so, I utilise the same “randomised azimuth” method as in section 4.4.2. Using this method with a live halo and bulge (i.e. ones populated with particles) is not straightforward, however. As discussed in chapter 3, a live N-body simulation will tend to “drift” slightly from its original centre. In this case, when that happens the disc will tend to drift in same the direction as the halo. However, since the particle positions in the disc are being rotated about the \(z\)-axis after every block-step, this drift is continually checked. This then affects the properties of the disc in a way that has no connection to the stability of the model.

In order to prevent this I utilise the same method I used in section 3.3.3: enforcing symmetry in the halo and bulge at all times (during disc growth and the fully populated
Initial conditions for simulations of multi-component galaxies: 4.4 Testing

**Figure 4.6.** Surface density profile for an isolated disc model with the randomised-azimuth method used to prevent the growth of bar modes or spirals. Both linear (upper), and log-linear profiles (lower) are shown in order to demonstrate how closely the profile retains its shape at both large and small radii. The colour code is the same in both graphs, the dotted line indicates the desired surface density profile.
Figure 4.7. Disc thickness (upper) and radial velocity dispersion (lower) for an isolated disc model with the randomised-azimuth method used to prevent the growth of bar modes or spirals. The colour code is the same as in figure 4.6.
Figure 4.8. Spherically averaged density profile of the halo in a fully populated, but constrained, simulation. The profile is shown at $t = 0$, after the growth of the full disc potential, and the replacement of the external potential with an $N$-body system, but before the full $N$-body system has evolved; and at two times during the full $N$-body simulation ($t = 100, 200$). Symmetry about the origin was enforced, to avoid numerical difficulties relating to off-centring.

Simulation). In this case I define the velocity dispersion of the disc $\sigma_R$ such that $Q = 1.2$ for all $R$.

The full disc potential was grown over a period $t_{\text{grow}} = 40$, as it was in section 4.4.1 and then kept in place for a time $t_{\text{hold}} = 20$ to ensure that the halo and bulge had fully relaxed in its presence. Only then was the disc populated with particles. Simulation were then run to observe the evolution over 200 time-units. Times quoted in figures 4.8-4.13 take $t = 0$ to be the time when the fully populated simulation begins (not the start of the growth of the full disc potential, as in section 4.4.1).
FIGURE 4.9. Spherically averaged density profile of the bulge in the same simulation as figure 4.8, shown at the same points in the evolution. Symmetry about the origin was enforced in the bulge, as well as in the halo.
4.4 Testing

**Figure 4.10.** Anisotropy parameter $\beta$ (equation 1.17) against radius for the halo and bulge in the same simulations as figures 4.8 & 4.9. $\beta$ was determined for particles in spherical shells.
Initial conditions for simulations of multi-component galaxies: 4.4 Testing

**Figure 4.11.** Minor and intermediate axis ratios (c/a and b/a respectively, c/a always being the lower of the two lines), for both the halo and bulge in the same simulations as figure 4.8. Particles are divided into shells by density, and the values plotted here are the axis ratios of the shells against their median radius.

Figures 4.8-4.11 show that the properties of the halo and bulge are maintained to within a high degree of accuracy throughout the simulation. The halo density is slightly raised in the inner \( \sim 0.02r_h = 0.12 \) (figure 4.8) and the bulge density is slightly lowered in the inner \( \sim 0.3r_b = 0.06 \) (figure 4.9), much as they were in simulations with an unpopulated disc (figures 4.2 & 4.3). Both components remain isotropic (with some noise, figure 4.10), and while they are non-spherical, which is not ideal, they do not become significantly more or less aspherical over the course of the simulation (figure 4.11).

Figure 4.12 shows that the surface density of the disc is nearly unchanged at all radii throughout the entire simulation. Figure 4.13 (upper) shows that the disc thickness \( z_{r.m.s.} \) remains near constant over the full radial range of the disc. The small deviations from the original near constant value of \( z_{r.m.s.} \) can be attributed to the same causes as in the isolated disc case.

The radial velocity dispersion in the disc (\( \sigma_R \) figure 4.13 bottom) is nearly unchanged in the range \( 0.5 < R < 3 \). In the outer parts of the disc there is some variation, but no consistent warming or cooling. In the inner parts of the disc there is a clear increase in \( \sigma_R \). This warming clearly does not affect the density distribution. It can reasonably be attributed to the facts that the decoupling of the vertical and planar motion is a poor approximation in the inner disc, and that, while the choice of a distribution function with constant \( Q \) can be useful in that it avoids the inner parts of the disc becoming very hot, it does lead to the velocity dispersion being unrealistically low at the very inner radii.

65
4.4 Testing

**Figure 4.12.** Surface density profile for an disc model embedded the full $N$-body halo and bulge, in the same simulation represented in figures 4.8 & 4.9. The randomised-azimuth method was used to prevent the growth of bar modes or spirals. Both linear (upper), and log-linear profiles (lower) are shown in order to demonstrate how closely the profile retains its shape at both large and small radii. The colour code is the same as in figure 4.6.

\[
\sum_d \quad \begin{array}{ll}
0.15 & \frac{0.1}{0.05} \\
0.1 & \frac{0.05}{0.01} \\
0 & \frac{0.01}{0.001} \\
\end{array}
\]

\[
R \\
0 & 1 & 2 & 3 & 4 & 5 & 6
\]
Initial conditions for simulations of multi-component galaxies: 4.4 Testing

**Figure 4.13.** Disc thickness (upper) and radial velocity dispersion (lower) for a disc model embedded the full $N$-body halo and bulge, in the same simulation represented in figures 4.8-4.12. The randomised-azimuth method used to prevent the growth of bar modes or spirals. The colour code is the same as in figure 4.6.
Figure 4.14. Surface density profiles for the discs in unconstrained simulations. For each simulation the surface density is plotted on both a linear (upper) and logarithmic (lower) scale. The red line corresponds to the desired profile; the black lines correspond to the surface density in the models at $t = 0$ (solid), $t = 100$ (dotted) and $t = 200$ (dashed).

Unconstrained tests

Finally I run simulations in which the galaxy model is unconstrained from the moment the disc component is populated. This means that the disc component is free to develop bar modes and other instabilities.

In figure 4.14 I plot the surface density profiles of discs with various initial $\sigma_R$ profiles. In each case $\sigma_R$ was defined such that $Q$ was constant across the disc. Simulations are shown with $Q = 1.2, 2, 3, 4$. As the disc gets hotter, it becomes increasingly difficult to iterate the particle positions such that the surface density is close to that desired, us-
ing the method described in section 4.3.3. This is why the initial surface density of the $Q = 4$ disc (and to a lesser extent the $Q = 3$ disc) is not an exponential in radius. It should also be noted that the motivation of warming an initially cold distribution function is increasingly invalid as the disc becomes hotter.

It has long been recognised that numerical simulations of cool discs are much more susceptible to bar formation than those of warm discs (e.g. Hohl 1971). This is what I observe in my models. The scatter plot of particle positions for the $Q = 1.2$ case, figure 4.15, shows spiral structure forming within 50 time units, and the formation of a strong bar within 100. This has the effect of substantially altering the surface density profile, dramatically increasing the disc’s surface density in the inner $\sim 0.3$ scale lengths (as seen in figure 4.14). The $Q = 3$ disc shows no signs of developing spiral structure within 200 time units (figure 4.16), and the density profile remains nearly unchanged throughout the simulation.

## 4.5 Conclusions

I have described and tested a new method for constructing an equilibrium $N$-body representation of a galaxy with halo, disc and bulge components.

I have used a distribution function for the halo and bulge based upon Cuddeford (1991) inversion, while the distribution function of the disc is based on the work of Dehnen (1999b). This avoids the use of a Maxwellian approximation, and produces models which tend to stay very close to their original states, though non-axisymmetric instabilities develop in the disc if it is unconstrained and reasonably dynamically cold.

This method has the advantage that the density distributions of the various components are relatively straightforward to prescribe. While the initially spherically symmetric components are somewhat flattened by the growth of the full disc potential, there is little to no radial contraction or expansion.

The choice of disc distribution function, while physically motivated, and clearly in equilibrium if non-axisymmetric modes are not allowed to grow, causes problems when creating a warm disc (Toomre’s $Q \gtrsim 3$) since it becomes increasingly difficult to tailor the distribution function to the desired density and velocity dispersion profiles.
Figure 4.15. Scatter plot of disc particle positions in the $x$-$y$ plane at various times in an unconstrained simulation with initial $Q = 1.2$. Only 1 in every 20 particles is plotted, in the interests of clarity.
Figure 4.16. Scatter plot of disc particle positions in the $x$-$y$ plane at various times in an unconstrained simulation with initial $Q = 3$. As in figure 4.15, only 1 in every 20 particles is plotted.
Chapter 5

The haloes of merger remnants

“Unless th’ Almighty Maker them ordain
His dark materials to create more worlds”

John Milton
In this chapter I perform simulations of 1:1 galaxy mergers, using models constructed as described in chapter 4. I focus on the behaviour of the halo component. I investigate the effect of varying the galaxy orientation; the orbit on which the galaxies approach one another; and the anisotropy or cusp strength of the haloes. I examine the effects upon the halo density profiles and shape, as well as on their kinematics.

### 5.1 Introduction

As discussed in section 1.3, simulations of hierarchical structure formation in a $\Lambda$CDM universe consistently produce cusped haloes, with spherically-averaged density profiles in the inner halo following a power law, $\rho \propto r^{-\gamma}$ with $1.0 \lesssim \gamma_0 \lesssim 1.5$, while in the outer parts of the halo $\rho \propto r^{-3}$ (e.g. Navarro et al. 1996, Moore et al. 1998, Power et al. 2003).

A further generic prediction of $\Lambda$CDM simulations is that dark matter haloes are triaxial simulations (e.g. Dubinski & Carlberg 1991, Allgood et al. 2006), with typical minor-to-major axes ratios $c/a \sim 0.6 - 0.7$ for a Milky Way sized galaxy. However, observations of the tidal stream of the Sagittarius dwarf spheroidal have been used to argue that the halo of the Milky Way is nearly spherical, with $c/a < 0.7$ ruled out to a high degree of confidence for Galactocentric distances in the range $16\text{kpc} < r < 60\text{kpc}$ (Ibata et al. 2001, Majewski et al. 2003). This is not necessarily a surprising result, since simulations which take into account the effect of gas cooling on dark matter haloes find that it results in haloes that are significantly more spherical than those found in pure CDM simulations (e.g. Katz & Gunn 1991, Kazantzidis et al. 2004).

Hansen & Moore (2006) claimed that there is also a universal relationship between the local density slope and velocity anisotropy of dark matter haloes. They argued that this could be understood through a recognition that the density slope in the tangential direction was zero (for a spherical halo), and that the shape of a velocity distribution is directly dependent on the density slope (which is known for simple structures; e.g. Hansen et al. 2005). The anisotropy of the halo is parameterised by $\beta$ (equation 1.17) and the local density slope as

$$\gamma(r) = - \frac{d \log \rho(r)}{d \log r}. \quad (5.1)$$

Hansen & Moore suggested that the anisotropy relates to the density profile as

$$\beta(\gamma) = 1 - \xi (1 - \gamma/6). \quad (5.2)$$
They used data from collapse, merger and cosmological simulations to support their hypothesis, and found that the best fit to their data come from taking the free parameter $\xi \approx 1.15$.

Dehnen & McLaughlin (2005) showed that assuming a linear $\beta - \gamma$ relationship, spherical symmetry, and the relationship $\rho/\sigma^3 \propto r^{-\alpha}$ where $\alpha$ is some constant (e.g. Taylor & Navarro 2001), defines a single analytical dynamical equilibrium for a dark matter halo with a density profile which closely resembles that found in simulations, independent of the value of $\alpha$.

Since the work of Toomre & Toomre (1972) on the tidal origin of bridges and tails, it has been recognised that elliptical galaxies could be formed by the merger of two disc galaxies. Various authors have conducted numerical ($N$-body) simulations of collisionless mergers, some focusing on the stellar component (e.g. Negroponte & White 1983, Barnes 1988, Hernquist & Ostriker 1992b, Naab & Burkert 2003), some on the dark matter halo (e.g. White 1978, Villumsen 1983, Boylan-Kolchin & Ma 2004, Kazantzidis et al. 2004, Aceves & Velázquez 2006, Kazantzidis, Zentner & Kravtsov 2006).

The studies of White (1978), Villumsen (1983) and Boylan-Kolchin & Ma (2004), and most other studies of the halo component used idealised, spherically symmetric systems to represent the galaxies. There are relatively few studies of the remnant haloes of mergers between haloes containing disc galaxies.

Kazantzidis et al. (2004) looked at the shapes of dark matter haloes, primarily considering the effect of gas cooling, but also reporting simulations of collisionless mergers of galaxies with disc components. They stated that the shape of the remnant halo depends sensitively on the relative inclination of the discs. Aceves & Velázquez (2006) conducted simulations of mergers of galaxies with mass-ratios of 1:1, 1:3 and 1:10, determining that in each case the initial “cuspy” density profile was preserved. Kazantzidis et al. (2006), as part of a larger study of dark matter halo mergers, included simulations of binary (1:1 mass-ratio) mergers of haloes with disc components. They too found that the presence of a disc did not affect the conclusion that “dissipationless mergers result in remnants that are practically scaled versions of their progenitors”.

All three of these papers create their initial conditions using the prescription of Hernquist (1993), (section 4.2) which makes the approximation of a locally Maxwellian velocity. This method is far from rigorous as the true equilibrium velocity distribution can be strongly non-Maxwellian.

Novak et al. (2006) looked at the shapes of both the stellar and dark matter components of the remnants from hydrodynamic simulations of major mergers. They found
The haloes of merger remnants: 5.1 Introduction

that the stellar remnants were generally oblate, while the haloes were prolate or triaxial. The initial haloes in their simulations were spherical, and the majority of remnant haloes were still relatively close to spherical, having $c/a > 0.75$, with no haloes from any of the 58 simulations having $c/a < 0.6$.

The most useful observational work on individual dark matter halo shapes for elliptical galaxies (Buote & Canizares 1996, Buote & Canizares 1998, Buote et al. 2002) uses measurements of the flattening of X-ray isophotes to place constraints on the ellipticity of the dark matter haloes of elliptical galaxies NGC1332, NGC3923 and NGC720 under the (questionable) assumption that flattening due to gas rotation is negligible. They find that all the haloes are substantially flattened, with $0.28 < c/a < 0.65$, and place tight constraints on the ellipticity of NGC720, which has $c/a = 0.38 \pm 0.05$ for any of their oblate density models, and $c/a = 0.37 \pm 0.04$ for any of their prolate models.

In section 5.2.1 I describe the galaxy models used in my simulations; in section 5.2.2 I describe the various suites of simulations I perform; in section 5.3.1 I examine the density profiles of the remnant haloes; in section 5.3.2 I look at their kinematics and in section 5.3.3 I consider their shape. Finally I discuss my results in section 5.4.

5.1.1 Dynamical processes of importance

There are certain dynamical processes that have significant effects in these simulations, so I will introduce them here. The two major relaxation processes that occur in a merger are violent relaxation and phase mixing.

Violent relaxation (e.g. Lynden-Bell 1967) is a process that changes the energy of individual particles. It occurs because the potential of the system varies with time. A common example of the way in which this alters the energy of an object is to consider a star at rest in the centre of a system undergoing gravitational collapse. In a perfectly symmetrical system the star will remain at rest, but the potential well that it is in will become deeper, thus decreasing the energy of the star. Mergers have a more complicated geometry than this, but similar fates can befall objects in the interaction. Conversely objects can also gain energy if they fall into and move out of the potential wells of the merging galaxies at the opposite times.

What will happen to an individual object due to violent relaxation is very hard to determine a priori, but the overall effect is to produce a greater spread in the energies of the objects.
Phase mixing is best explained through a simple case, analogous to the more complicated case of a stellar system, but simpler to visualise, for instance Dehnen (2005) uses a system with the Hamiltonian $H = p^2 + |q|$. It relates to what happens when a large number of objects have a similar position in phase space. These objects can be pendulums with similar initial positions and velocities, or stars on similar orbits. Figure 5.1 shows the position in the phase space of $\sim 10^5$ particles evolving with time in a 1-dimensional system. Initially the particles are tightly bunched in $p$ and $q$, but over time they move out of phase with one another, and the particles are left spread more evenly over phase space. The collisionless Boltzmann equation (equation 1.4) tells us that phase space density is conserved, which of course it is for the infinitesimal volume which follows each individual particle. However, the average over a finite volume, known as the coarse grained phase space density $\bar{f}$, is clearly reduced in this case.

In fact the CBE ensures that $\bar{f}$ cannot increase, and any mixing process will cause a decrease in $\bar{f}$. In a full 3-dimensional case (such as a stellar system), this mixing process is far quicker than the 1-dimensional process shown in figure 5.1, especially if there is also violent relaxation, as there is in a merger. Dehnen (2005) introduced the excess mass function $D(f) = \int_{\bar{f}}^{\bar{f}_{\star}} f(x, v) - f \, d^3x \, d^3v$, which he demonstrated always
decreases or is unchanged in any mixing process. Dehnen showed that the merger of two “pure cusps” cannot produce a remnant cusp steeper (or, he argued shallower) than the steepest of its progenitors.

Another process that may well be important in these simulations is the radial-orbit instability (e.g. Antonov 1973, Barnes 1985). Systems with large numbers of radial orbits are known to be unstable to deformation towards a barred or triaxial shape. This is due to the fact that, in a well behaved potential, near-radial orbits are nearly closed, so they act like slowly precessing rods. In the presence of a bar-like potential, these “rods” tend to bunch up towards the bar, thus increasing its strength (Polyachenko 1989). It has been argued that the radial-orbit instability in cosmological simulations is instrumental in driving dark matter haloes towards the observed universal density profile (e.g. Huss, Jain & Steinmetz 1999, MacMillan, Widrow & Henriksen 2006).

5.2 Galaxy models and simulations

5.2.1 Initial Conditions

The galaxy models used in these simulations are constructed using the method of chapter 4. As in the tests of these initial conditions, the disc has a surface density which decreases as an exponential in cylindrical radius (eq. 4.16), with the distribution function defined by equation 4.13. The bulge has a Hernquist density profile (equation 3.1). For the halo I use a generalised spherically-symmetric truncated NFW-like profile:

$$\rho_h(r) = \frac{\rho_c}{(r/r_h)^{\gamma_0}(1 + r/r_h)^{3-\gamma_0}} \text{sech}(r/r_t),$$  \hspace{1cm} (5.3)

where $\rho_c$ is a scale density, $r_h$ is the halo scale radius, $r_t$ is the halo truncation radius. The generalisation of this profile as compared to the previously used truncated NFW profile (equation 3.3) is in the new parameter $\gamma_0$ which describes the inner slope of the density profile. As $r \rightarrow 0$, the halo density $\rho_h \propto r^{-\gamma_0}$

In all my models (as in section 4.4) I defined the disc scale length, $R_d = 1$, disc mass $M_d = 1$, and disc scale height $z_d = 0.1$. The velocity dispersion of the disc was defined such that the Toomre (1964) stability parameter $Q = 1.2$ (equation 4.19) at all radii. The bulge mass $M_b = 0.2$, and bulge scale length $r_b = 0.2$. I choose units such that the constant of gravity $G = 1$. I chose the scale radius of the halo $r_h = 6$, the truncation radius $r_t = 60$, and the halo mass $M_h = 24$. In simulations with $\gamma_0 = 1.0$, 79% of this
total mass is within the truncation radius. The rotation curve for this model is shown in Figure 4.1.

The stellar components were populated with 150,000 equal mass particles (i.e. 125,000 in the disc, 25,000 in the bulge) with a smoothing length $\epsilon_{\text{stellar}} = 0.02$. The halo component was populated with 750,000 equal mass particles with a smoothing length $\epsilon_{\text{halo}} = 0.04$. That corresponds to each halo particle being 4 times more massive than a stellar particle. A small number of simulations with 4 times as many particles were performed for comparison. These demonstrated that the numerical resolution used was sufficient.

The N-body simulations were performed using gyrfalcON, with a minimum time step $2^{-7}$ and a block-step scheme allowing steps up to eight times larger. Individual particle time steps were adjusted in an (almost) time-symmetric fashion such that on average

$$
\tau_i = \min \left\{ \frac{0.01}{|a_i|}, \frac{0.05}{|\Phi_i|} \right\},
$$

with $\Phi_i$ and $a_i$ the gravitational potential and acceleration of the $i$th body. With these parameters, energy was conserved to within 0.1% over the full time span (1000 time units) in a typical simulation, approximately corresponding to a Hubble time. The time span was chosen such that there is sufficient time for the remnant to reach a dynamical equilibrium.

In all simulations the equal mass galaxies were placed on a mutual orbit in the x-y plane. In all cases this orbit was parabolic and, in the vast majority of cases, corresponded to one that would have a pericentre separation of $8 R_d$ if the galaxies were point masses; I refer to this distance as the “impact parameter” $d$ of the merger. In some cases the galaxies were placed on parabolic orbits with smaller pericentre separations, or on a radial orbit with zero net energy (i.e. a parabolic orbit in the limit where impact parameter $d \to 0$). The galaxy centres were initially separated by $200R_d$.

### 5.2.2 Suites of simulations

I perform four suites of simulations of equal-mass mergers, to examine the effects of varying different parameters on the mergers. The suites vary in

1. Orientation of the disc galaxy.

2. Pericentre separation of the galaxies’ mutual orbit.
3. Initial velocity anisotropy of the halo component.

4. Cusp strength of the haloes ($\gamma_0$ in Eq. 5.3).

In addition, I ran control simulations, which were stand-alone simulations of the same galaxy models, run for the same length of time, with no merger. The galaxy disc is bar-unstable, which in these simulations causes a steepening in the density profile of the inner parts of the disc, which also causes a steepening in the halo profile in the inner $r \sim 0.2r_h$. As discussed in chapter 3, this effect is a common observation in simulations (e.g. Sellwood 2003, Athanassoula 2004)

**Suite 1: Orientation**

My first suite of simulations consisted of 16 mergers which vary only in the orientation of the discs with respect to one another, and to the angular momentum vector of their mutual orbit. Following the example of Barnes (1988), I define four different orientations for the spin vector of each galaxy disc, corresponding to pointing towards the four vertices of a regular tetrahedron. The orientations of disc galaxies in mergers are usually described in terms of the disc inclination relative to the orbital plane $i$, and the argument of pericentre $\omega$ (see figure 5.2 or Toomre & Toomre 1972). The inclinations ($i_1$ & $i_2$), and arguments of pericentre ($\omega_1$ & $\omega_2$) for the suite of mergers is shown in Table 5.1.
The haloes of merger remnants: 5.2 Galaxy models and simulations

In all cases the initial halo has an inner density slope $\gamma_0 = 1.0$ (Eq. 5.3), and an isotropic ($\beta = 0$) velocity distribution. The galaxies were put on a parabolic orbit with a pericentre separation of $8R_d$.

<table>
<thead>
<tr>
<th>Disc 1</th>
<th>Disc 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>#</td>
<td>$i_1$</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>-109</td>
</tr>
<tr>
<td>6</td>
<td>-109</td>
</tr>
<tr>
<td>7</td>
<td>-109</td>
</tr>
<tr>
<td>8</td>
<td>-109</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Disc 1</th>
<th>Disc 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>#</td>
<td>$i_1$</td>
</tr>
<tr>
<td>9</td>
<td>-109</td>
</tr>
<tr>
<td>10</td>
<td>-109</td>
</tr>
<tr>
<td>11</td>
<td>-109</td>
</tr>
<tr>
<td>12</td>
<td>-109</td>
</tr>
<tr>
<td>13</td>
<td>-109</td>
</tr>
<tr>
<td>14</td>
<td>-109</td>
</tr>
<tr>
<td>15</td>
<td>-109</td>
</tr>
<tr>
<td>16</td>
<td>-109</td>
</tr>
</tbody>
</table>

**TABLE 5.1**

**Suite 2: Impact parameter**

I performed two sets of simulations with galaxy models identical to those used in Suite 1, varying the separation at pericentre (impact parameter) $d$ between $8R_d$ and 0 (i.e. radial impact) in increments of $2R_d$. The two sets of simulations differ in the galaxy orientations, which correspond to those of simulations 4 & 7 in suite 1 (Table 5.1).

**Suite 3: Halo anisotropy**

My third suite of simulations consisted of mergers of galaxies with varying initial halo velocity dispersions and were performed to examine the effect that has on remnant properties. The haloes had identical density profiles to those in suites 1 and 2, but rather than an isotropic velocity distribution, they were initialised with constant radial anisotropy ($\beta = 0.35$ throughout the halo), constant tangential anisotropy ($\beta = -0.4$ throughout the halo), or Osipkov-Merritt-model-like anisotropy (Eq. 1.15) with $r_a = 24 = 4r_h$. Simulations were performed with galaxies orientated as per orientation 7 (table 5.1), and with pericentre separations of 8, 3, and $0R_d$. 
**Suite 4: Cusp strength**

I ran a series of simulations with varying halo cusp strengths. The inner slope of the density distribution ($\gamma_0$ in Eq. 5.3) varied between $\gamma_0 = 0.1$ and $\gamma_0 = 1.6$. In all cases I retained the parameters $r_h = 6$, $r_t = 60$ and $M_h = 24$. The galaxy discs had the same orientations as simulation 4 in table 5.1. The haloes all had isotropic velocity distributions, and the mergers trajectories all had impact parameter $d = 8$.

**5.3 Results**

**5.3.1 Halo density profiles**

The haloes of the merger remnants are not spherically symmetrical (see Section 5.3.3), however insight, and a comparison with previous work, can be found from looking at spherically averaged density profiles, and local density slopes (eq. 5.1).

Figure 5.3 shows the density profiles and slopes for the 16 merger remnants from suite 1 (Section 5.2.2), with the equivalent for the stand-alone simulations. The disc orientations have no effect on the density profile, even in the regions where they dominate the galaxies’ density. The shape of the remnant density profile is very similar to that of the initial haloes, with an almost identical scale length. There are, however differences in the outer parts: $\sim 6\%$ of the halo mass becomes unbound during the merger, and the mass within $10r_h$ is $\sim 50\%$ of the total mass of the original haloes (as compared to 79\% in each halo initially). Hence the density within the inner $\sim 10r_h$ of the merged haloes is closer to that of one of the initial haloes than to double that (as it would be if the merger was equivalent to “freezing” the haloes and placing them on top of one another).

The simulations of suite 2 show that varying the impact parameter of mergers also has little effect on the density profile (Figure 5.4). Results for suite 4 showed the same trends for all cusp strengths. In the interests of brevity, only the results for $\gamma_0 = 0.1$ and $\gamma_0 = 1.4$ are shown (Figure 5.5).

Results from suite 3 do show a slight trend in remnant cusp strength with initial halo anisotropy. The trend is observable for all orbital separations simulated, but only results from simulations with orbital separation $d = 8R_d$ are shown in figure 5.6 for clarity. Haloes which have initially constant tangential anisotropy have sharper remnant cusps than initially isotropic haloes, while haloes with initial constant radial anisotropy have weaker remnant cusps. Giving the haloes Osipkov-Merritt anisotropy does not alter the
Figure 5.3. Suite 1: Spherically averaged density profile (lower) and density slope profile ($\gamma$, see eq. 5.1, upper) for remnant haloes from the 16 simulations in suite 1 (solid lines). Also shown, for comparison, is the density profile and slope of the halo of the equivalent stand-alone simulation (dotted), and the same stand-alone density profile with the density doubled - i.e. what one would expect if the galaxies were simply stacked on top of one another (dashed).
FIGURE 5.4. Suite 2: Spherically averaged density profile and density slope profile for remnant haloes, as in figure 5.3, for simulations with impact parameters $d = 8, 6, 4, 2, 0 R_d$ (solid lines).
cusp strength. Compared to the isotropic case, cusp strength – measured in terms of $\gamma(r)$ at small radii – is $\sim 0.1$ greater for the $\beta = 0.35$ case than for the isotropic case, and is $\sim 0.1$ less in the $\beta = -0.4$ case. This trend is not seen in the stand-alone profiles, which are nearly identical in all three cases.

### 5.3.2 Halo velocity distributions

I investigate the velocity distribution of the haloes by looking at $\beta$ as a function of radius, and as a function of the local density slope $\gamma$. $\beta$ is determined for particles in spherical shells, with the median particle radius for the shell being the value plotted in figures 5.7 & 5.8.

Error estimates, as shown in figures 5.7 & 5.8, were found for the spherical shell of particles through the standard propagation of errors formula, such that the error $\Delta \beta$ is given by

$$
(\Delta \beta)^2 = \left( \frac{\partial \beta}{\partial \sigma_0^2} \right)^2 (\Delta \sigma_0^2)^2 + \left( \frac{\partial \beta}{\partial \sigma_0^2} \right)^2 (\Delta \sigma_0^2)^2 + \left( \frac{\partial \beta}{\partial \sigma_r^2} \right)^2 (\Delta \sigma_r^2)^2;
$$

(5.5)

The remarkable aspect of the observed $\beta$ profiles of the remnant haloes is its near independence from the parameters varied in suites 1–3 (Figure 5.7, top to bottom). In all cases examined $\beta \approx 0.1$ at small $r$, and increases to $\beta \sim 0.2 - 0.3$ at $r \sim 10r_h$. This shows that while the remnant halo has a clear “memory” of the density profiles of the
Figure 5.6. Suite 3: Spherically averaged density profile and density slope profile for remnant haloes from $d = 8R_d$ mergers in which the initial haloes had constant radial anisotropy ($\beta = 0.35$: solid line); constant tangential anisotropy ($\beta = -0.4$: short-dashed); Osipkov-Merritt anisotropy ($r_a = 24$: long-dashed) Also shown are the density profile and slope for the equivalent mergers with initially isotropic haloes, as in all other simulations dotted.)
Figure 5.7. $\beta$ as a function of radius for the simulations of suites 1–3 (top to bottom). Error bars shown are typical of all simulations in respective plot, found using equation 5.5. In the bottom plot remnants of mergers with haloes with different initial anisotropies are represented by different line types. Dotted lines plot $\beta$ for remnants of initial halo $\beta = 0.35$ mergers, dashed lines for initial halo $\beta = -0.4$, solid lines for the Osipkov-Merritt haloes.
initial haloes, it has little memory of the velocity distribution of the initial haloes. There is some difference observable between the $\beta = 0.35$ and $\beta = -0.4$ cases, perhaps most noticeable in the range $0.5r_h < r < r_h$, but it is vastly smaller than the difference in the initial halo conditions.

One would also naively expect the impact parameter of the merger to have a significant effect, but it does not show any greater effect on the $\beta$ profile than the orientation of the galaxies does. The independence from disc orientation is unsurprising at large radii, but also continues to small radii, where one might have expected the disc orientation to play a role.

The $\beta$ profiles from the simulations of suite 4 are shown in Figure 5.8. As $r \rightarrow 0$, the value of $\beta$ tends to 0 for the less strongly cusped initial conditions ($\gamma_0 = 0.1 - 0.6$). As the cusp strength increases, the radial anisotropy of the inner parts of the halo increases. At larger radii the anisotropy tends towards $\beta \sim 0.2 - 0.3$ for all cusp strengths, with the less strongly cusped haloes tending to have higher anisotropy.

An & Evans (2006) showed that in a cusped halo the value of the central velocity anisotropy $\beta_0$ is limited such that $\beta_0 \leq \gamma/2$. This is supported by Figure 5.8, in which we see $\beta_0 \sim \gamma_0/10$ for the cusp strengths $\gamma_0 \geq 0.6$.

Almost all of the $\beta$ profiles share a similar non-monotonic shape. $\beta$ increases at small $r$, reaching a peak at $r \sim 1 - 1.5r_h$, then falls off, reaching a local minimum at $r \sim 2 - 3r_h$, before rising again. This apparent oscillatory variation of $\beta$ is not associated with any tidal arms, shells, or any other observed features of the particle distribution; it is also not seen to disappear or vary to any great extent with time. A similar $\beta$ profile can be seen in Figure 11 of Boylan-Kolchin & Ma (2004). Further work is required to determine its cause and importance.

Hansen & Moore (2006) argued for a universal relationship between local density slope $\gamma$ and velocity anisotropy $\beta$, given in equation 5.2, which I plot with their fit for the free parameter $\xi = 1.15$ (solid line) in each of the graphs in figure 5.9. The data fit the suggested relationship for $\gamma \lesssim 2$, but do not fit for $\gamma \gtrsim 2$, where $\beta$ is systematically lower than the fitting function. While it is the outer parts of the halo which have $\gamma \gtrsim 2$, and the dynamical time is longer here than in the inner parts of the halo, this is not simply because the outer halo has yet to reach an equilibrium. Simulations which ran for twice as long (until $t = 2000 \equiv 28\text{Gyr}$ scaled to the Milky Way) showed identical results. The $\beta = 0.35$ initial conditions produce remnants which are a close fit to the suggested relationship, but the $\beta = -0.4$ initial conditions produce remnants which have systematically slightly lower values of $\beta$ for the same $\gamma$. The fitting formula equation 5.2.
Figure 5.8. $\beta$ as a function of radius for the simulations of suite 4. Error bars shown are typical. Simulation results shown are for haloes with density cusps of strengths as indicated.
The haloes of merger remnants:

5.3 Results

FIGURE 5.9. Anisotropy parameter $\beta$ plotted as a function of the local density slope $\gamma(r)$ for mergers from suite 2 (top), suite 3 (bottom-left), and suite 4 (bottom-right). $\beta$ and $\gamma$ are determined for each simulation at many points within the radial range $0.1r_h < r < 10r_h$. The lines drawn in each plot correspond to equation 5.2, with $\xi = 1.15$ (solid line) and $\xi = 1.05$ (dotted). In the central plot, filled squares correspond to mergers with $\beta = 0.35$ initial conditions, open squares are from the $\beta = -0.4$ initial conditions, and crosses for the Osipkov-Merritt halo initial conditions.

with the free parameter $\xi \simeq 1.05$ (plotted as a dotted line in figure 5.9) does appear to provide an upper limit for $\beta$.

5.3.3 Halo Shape

I calculate the overall shape of the remnant halo through an iterative procedure based on diagonalising the moment of inertia tensor (equation 4.15). I start with a spherical window, centred on the origin, which contains 50% of the mass of the halo, then determine and
Figure 5.10. Minor ($c/a$) and intermediate ($b/a$) axis ratios for the simulations of all suites of simulations (as labelled). In all figures the solid diagonal indicates the position on the plot of a completely prolate figure. The dashed lines are of constant triaxiality, as indicated. For suites 2 (top-right), 3 (bottom-left) and 4 (bottom-right) the different symbols represent different simulation parameters as labelled. The three points for each initial anisotropy (suite 3) correspond to simulations with impact parameters $d = 0, 3, 8$ with, in each case, $b/a_{d=0} < b/a_{d=3} < b/a_{d=8}$.
diagonalise its moment of inertia tensor. This determines the principal axes and the axis ratios of an ellipsoidal window, centred on the origin, which is scaled to contain 50% of the halo mass. This is then repeated to find new ellipsoidal windows until the results converge such that the volume of the window is conserved to within a fractional difference of \(10^{-3}\) between iterations. From the ellipsoidal window I determine the axis lengths \(a, b, c\) with \(a \geq b \geq c\), and thus the intermediate \((b/a)\) and minor \((c/a)\) axis ratios. I can then define an ellipsoidal “radius”

\[
\zeta = a \sqrt{\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2}}.
\]

(5.6)

The question of whether an ellipsoidal body is prolate, oblate or triaxial can be quantified through the parameter

\[
T = \frac{(a^2 - b^2)}{(a^2 - c^2)}.
\]

(5.7)

I use this method to enable a like-for-like comparison with the simulations of Novak et al. (2006), though starting with a spherical window can cause a bias of the kind discussed in section 4.4.1.

**Trends in halo shapes**

The graphs of figure 5.10, which show the ratios \(c/a\) and \(b/a\) (determined as described above) for simulations from all of our suites, demonstrate the following things:

1. Remnant haloes are significantly non-spherical \((c/a \lesssim 0.55\) in all cases shown). They are generally prolate, or at least tending towards prolate, in my simulations.

2. The triaxiality of the halo at this scale \((\zeta \sim 10r_h\) for outermost particles) is barely affected by the initial inclination of the disc.

3. More radial mergers (i.e. mergers with lower values of \(d\)) tend to have more prolate remnant haloes. This is true whether or not the initial haloes are isotropic.

4. Tangential anisotropy in the initial halo tends to produce haloes that are more prolate, and have a higher value of \(c/a\) than in the isotropic case. Conversely, radial anisotropy in the initial halo tends to produce haloes that are more triaxial, and have a lower value of \(c/a\) than in the isotropic case.

5. In the range \(0.4 < \gamma_0 < 1.6\), an increasing cusp strength tends to produce increasing far from prolate remnants (i.e. \(b/a\) increases). This is not simply due to the
decreasing size of the window containing 50% of the mass in haloes with increasing cusp strength, and is also seen if the physical size of the window is defined such that it is the same for all cusp strengths.

**Ellipticity profiles**

I wish to examine the effect of the orientation of the disc components upon the ellipticity of the remnant halo as a function of radius. In an effort to do this I use the method described in section 4.4.1, dividing the particles into density shells, and determine the axis ratios and median ellipsoidal radius of each shell. For comparison to Kazantzidis et al. (2004) I also divide the merger remnant halo into spherical shells, and find the axis ratio of the particles in those shells. I defined the radius of the spherical shells as being the median radius of the particles in it. The two approaches produced qualitatively similar results, though the axis ratios found for the spherical shells were, unsurprisingly, systematically larger (see section 4.4.1). In figure 5.11 I plot the results from the both analyses.

Considered in terms of $c/a$ at the innermost point determined, the flattest merger remnant comes from the co-planar merger, and the next four flattest remnants are from mergers with one of the discs in the orbital plane. The difference in minor axial ratio is relatively small. Using the density shells values, at the innermost point for the co-planar merger remnant $c/a$ is $\sim 0.62$, for mergers with one galaxy disc in the orbital plane it’s $\sim 0.67$, and for mergers with neither disc in the orbital plane it’s $\sim 0.72$. The simulation with orientation 16 (see table 5.1) produces a notably more spherical remnant than all the other simulations ($c/a \sim 0.83$). Beyond $r \sim 1.5r_h$ the difference in $c/a$ is negligible. There is no trend apparent in the intermediate axis ratio, $b/a$. These trends are exactly as one would expect.

It is reasonable to ask whether the fact that these haloes are so far from spherical invalidates the approach in section 5.3.1 of using a spherically averaged density profile. In an effort to investigate this I created density profiles based upon the density shells, found through the method described above. These profiles, while suffering somewhat from increased noise, showed the same behaviour as that described in section 5.3.1. This indicates that the use of spherically averaged profiles is sufficient and appropriate for the analysis of the haloes of merger remnants.
The haloes of merger remnants: 5.3 Results

**Figure 5.11.** Minor axis ratios as a function of radius $r$ (upper) or ellipsoidal radius $\zeta$ (lower) for the simulations of suite 1. The dotted lines correspond to the simulation in which galaxy discs are both oriented in the merger plane, the dashed lines correspond to the simulations in which one of the discs was oriented in the merger plane.
5.4 Discussion

I have performed a number of suites of simulations, covering a number of different merger parameters. Across all these simulations I have seen that the halo cusp strength is extremely robust against changes caused by major mergers, even when there is a centrally concentrated stellar component. This is in keeping with previous results from halo only simulations (e.g. Kazantzidis et al. 2006); simulations with disc components, but initialised using a Maxwellian approximation (e.g. Aceves & Velázquez 2006) and analytical arguments (Dehnen 2005). Thus mergers, at least of gas-less equal mass galaxies, can not solve the persistent discrepancy between observations of nearby galaxies, which imply that galactic dark matter haloes have a density profile with a flat core, and the cosmological standard model, which predicts that haloes should have a cusp.

In contrast, the velocity anisotropy of the halo has very little “memory” of the initial conditions of the merger. In all cases examined, the velocity anisotropy of the remnant followed the relationship (eq. 5.2) suggested by Hansen & Moore (2006) relatively closely for $\gamma \lesssim 2$, but $\beta$ was systematically lower than predicted for $\gamma \gtrsim 2$. That these results are nearly independent of initial anisotropy is very encouraging because the overwhelming majority of merger simulations to date have been performed with isotropic haloes, while the haloes found in cosmological simulations are typically radially anisotropic. A strong dependence on anisotropy would suggest that the approximation of isotropy in the halo was a poor one, which would bring into question the results from simulations using it. The anisotropy of the halo is clearly a less robust property than its density profile.

Given the shape of the dark matter haloes inferred by, for example Buote et al. (2002), the results of section 5.3.3 are of particular importance. Axis ratios $c/a$ in the range $0.37 \pm 0.04$ (as determined for NGC720) are seen for all initial halo anisotropies except $\beta = -0.4$, in cases with sufficiently small impact parameters. Simulations which incorporate the effects of gas physics (e.g. Novak et al. 2006) find remnant haloes which are significantly more spherical. It is likely, therefore, that galaxies with such highly flattened haloes were formed from near-radial trajectory, very gas-poor mergers in which collisionless dynamics dominate.

That the remnant halo is closest to completely prolate for mergers with low impact parameters is unsurprising, and was recognised in the literature as far back as Villumsen (1983) in the case of completely spherical initial models. In the $d = 0$ case, the major axis of the remnant is in the same direction as the initial motion of the galaxy centres.

The dependence of remnant shape on initial anisotropy of the haloes is less triv-
ially understandable, but it could be related to the radial-orbit instability (section 5.1.1). Spherical systems with large numbers of radial orbits are known to be unstable to deformation towards a barred or triaxial shape. The radial-orbit instability is also known to reduce the central concentration of models (e.g. Merritt & Aguilar 1985), which could explain the slight decrease in cusp strength seen in the $\beta = 0.35$ model remnants (figure 5.6).

It is known that a steep cusp in a triaxial model causes chaotic orbit-scattering which can have a significant effect on its shape (Valluri & Merritt 1998), reducing the triaxiality. However, the effect of increasing the cusp strength of our haloes is to make the remnants less prolate (i.e. increasing $b/a$), which means that the most strongly cusped haloes are more triaxial than the least strongly cusped ones.

Increasing the cusp strength of the halo puts more of the mass of the halo at radii smaller than the impact parameter of the merger. This is similar to the effect of increasing the impact parameter of the merger, which we have shown increases $b/a$ for the remnant. It is likely that this explains the trend in remnant shape with cusp strength.

5.5 Conclusions

I have performed a thorough global study of the properties of the haloes of merger remnants, covering their structural and kinematic aspects. I followed collisionless $N$-body simulations of 1:1 mass ratio galaxy mergers, using models which include a galaxy halo, disc and bulge. I have investigated the effect of varying the galaxies’ orientation, their mutual orbit, and the initial velocity anisotropy or cusp strength of the haloes upon the remnant halo density profiles and axis ratios, as well as on the velocity anisotropy of the halo.

I have shown that the halo density profile (determined as a spherical average, an approximation I found to be appropriate) is exceptionally robust in mergers. This means that mergers, even with a centrally concentrated stellar component, do not appear to be a potential mechanism for producing cored dark matter haloes from ones which are initially cusped.

The velocity anisotropy of the remnant haloes is nearly independent of the orbits or initial anisotropy of the haloes. The remnants follow the halo anisotropy - local density slope ($\beta - \gamma$) relation suggested by Hansen & Moore (2006) in the inner parts of the halo, but $\beta$ is systematically lower than this relation predicts in the outer parts.
The axis ratios of the remnant haloes are strongly dependent on the initial parameters of the haloes and their orbits, in particular increased radial anisotropy in the initial halo tends to produce more aspherical remnants, and haloes which approach one another with small impact parameters tend to be close to prolate. I also find that the remnant haloes are significantly less spherical than those described in studies of simulations which include gas cooling. This suggests that the far from spherical haloes observed around elliptical galaxies (e.g. Buote et al. 2002) may well have been produced by very gas poor mergers.
Bibliography


'Abd Al-Rahman Al Sufi (964), Book of Fixed Stars, Isfahan, Persia.


Bibliography


Bibliography


Kazantzidis, S., Kravtsov, A. V., Zentner, A. R., Allgood, B., Nagai, D. & Moore, B.
611, L73–L76.


AJ 71, 64–++.

logical Baryon Density from the Deuterium-to-Hydrogen Ratio in QSO Absorption


Kuo, C. L., Ade, P. A. R., Bock, J. J., Cantalupo, C., Daub, M. D., Goldstein, J.,
Holzapfel, W. L., Lange, A. E., Lueker, M., Newcomb, M., Peterson, J. B., Ruhl, J.,
Runyan, M. C. & Torbet, E. (2004), ‘High-Resolution Observations of the Cosmic


Lynden-Bell, D. & Kalnajs, A. J. (1972), ‘On the generating mechanism of spiral struc-


Bibliography


