The Evolution of Disk Galaxies in Cold Dark Matter Halos

by

Andreea S. Font
B.Sc., University of Bucharest, 1996
M.Sc (Astronomy) Saint Mary's University, 2000

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Physics and Astronomy

Acknowledgements

I would like to thank my supervisor, Julio Navarro, for suggesting the topic of this thesis, for providing the material and logistic support necessary to carry it on, and for detailed and careful feedback along the way. I also thank Heather Morrison and Arif Babul for sharing their expertise on this topic and for useful comments and suggestions. I am indebted to Chris Pritchet, Ann Gower and David Hartwick for their advice on professional and personal matters. I have fond memories of the morning coffee breaks which often were a fertile learning ground and which helped refresh my commitment to my work. I thank Doug Johnstone for sharing his enthusiasm for Astronomy and for reminding me that doing science is also about having fun. Many thanks also to Rosemary Barlow and Geri Blake for their kind help in all administrative matters.

During my stay in Victoria I have shared the space and personal experiences with many fellow students. I cannot list them all here and it would be unfair to single one out. I thank them for making my graduate life more interesting and balanced and by enriching it with their various experiences.

I would also like to express here my gratitude for my two mentors, Göran Sandell and Mircea Rusu. Through his integrity, strong work ethic and uncompromising honesty Göran has been a great role model, albeit an impossible one to emulate. Nevertheless, knowing that I had the appreciation of this person helped me persevere in difficult moments. Mircea Rusu made a fabulous entry in my life (or rather I stumbled into his) more than fifteen years ago. Since then, he worked his magic at several critical turning points in my life. As a Physics professor he gave me and countless of other Romanian students an unforgettable introduction into the wonders of the universe and has diligently built up our confidence to carry on.

Words are never enough to thank those people who were close to me in this journey. I thank Ian McCarthy for never tiring of helping me out, for supplying me with strength and reason at times when I lacked them, and for enlightening my life with his witty humour. This thesis also benefited from his careful reading, insightful

comments and patient spellchecking.

I also want to thank my mother, Ada Maria Ghiuzeli, who despite thousands of miles of land and water that separated us (notwithstanding the inconvenient time lag), has always been in sync with my foreign exploits and who, whenever I was uncertain, reminded me of what really matters. Distance however had its benefit, revealing what at close approach was not visible or rather what I took it for granted. That she is a strong woman who worked hard in order to open up for me a realm of possibilities and who supported me in whatever decisions I took. With hopes that my realization does not come too late, I dedicate this work to her.

Contents

List of Tables

List of Figures

Supervisor: Dr. Julio F. Navarro

ABSTRACT

We use high resolution N-body simulations to investigate the dynamical effects that substructure in Cold Dark Matter (CDM) halos have on galactic disks, with particular emphasis on their secular evolution, heating, tilting and warping. The simulations analyzed here are some of the largest and most realistic simulations of disk heating/ warping available in the appropriate cosmological context. Our detailed treatment of the dark matter distinguishes them from previous numerical simulations that have focused on the interaction with a single satellite.

Our study shows that substructure halos with masses, densities and orbits expected in the CDM paradigm typically play only a minor dynamical role in the heating of the disk over several Gyrs, and thus do not typically pose a danger to the stability of thin disks. This is largely because the most massive dark satellites, which dominate the secular heating, seldom approach the disk, where tidal effects are strongest. Occasionally, however, massive subhalos couple effectively with the disk, resulting in noticeable tidal effects on the structure of the stellar disk, including: (i) tilting and ii) the forcing of short-lived, asymmetric warps as a result of tidal impulses that arise during each pericentric passage. I show that this is a viable mechanism for creating asymmetric disk warps such as those observed in the local Universe. Moreover, the fact that a satellite can have recurrent interactions with the disk suggests a natural explanation for the observed frequency of the warps, which would otherwise be very short lived.

I conclude that dark matter halo substructure does not preclude virialized CDM halos from being acceptable hosts of thin stellar disks like that of the Milky Way and that the ubiquity of minor stellar warps may be associated with the recurrent tidal influence on the disk of the most massive substructure halos.

Chapter 1

Introduction

1.1 The "Concordance" Model of Structure Formation

Over the past few decades, substantial progress has been made in our understanding of the structure and evolution of the Universe. The past recent years have witnessed a further breakthrough, with the emergence and establishment of the so-called "concordance" ΛCDM model (Wang et al. 2000). The ΛCDM model is consistent with most observational constraints: the age of the universe; the primordial element abundances predicted by nucleosynthesis calculations; the overall abundance and motions of galaxies, groups and clusters; as well as the large scale streaming motions in the local Universe (Bahcall et al. 1999). In addition, recent observations with the WMAP (Wilkinson Microwave Anisotropy Probe) satellite have revealed that the properties of cosmic background radiation (CBR) are also in astonishing agreement with the concordance model.

The latest surveys (Bennett et al. 2003) tell us that the total density of the universe is $\Omega_{tot} = 1.02 \pm 0.02$, of which approximately 73 % is in the form of vacuum (or dark) energy ($\Omega_\Lambda = 0.73 \pm 0.04$), ~ 27 % is in the form of matter ($\Omega_m = 0.27 \pm 0.04$),

mostly dark and $\Omega_b = 0.044 \pm 0.004$, baryonic.[1] Other fundamental parameters of the cosmological model have also been constrained: h, the Hubble constant in units of 100 km s^{-1} Mpc^{-1}, is 0.71 ± 0.03; and σ_8, the rms fluctuation on the scale of 8 h^{-1} Mpc, is 0.84 ± 0.04 (Bennett et al. 2003). A full list of the cosmological parameters is given by Bennett et al. (2003). During the next decade, efforts will be focused on refining the measurements of the parameters that define the cosmological paradigm, and thus deriving detailed information about the various components that make up the Universe. These studies hopefully will help unravel the nature of two important components of the universe whose nature still eludes us: the dark matter and the dark energy.

Dark matter is a crucial component in the formation and evolution of cosmological structure. Although it cannot be directly detected in observations, its presence is felt through its gravitational pull on cosmological structures. We also have good indications that this unseen matter behaves like an ensemble of collisionless particles interacting only through the force of gravity. These particles are believed to be cold at early times (hence "Cold Dark Matter", or CDM), meaning they have a very small velocity dispersion. The dark energy, one example of which is provided by the cosmological constant Λ, is the other unknown ingredient of the model. It essentially acts as a negative pressure which causes the expansion of the universe to accelerate[2] (Perlmutter et al. 1999; Riess et al., 2001). Cold dark matter and a non-zero cosmological constant are the centerpiece of the concordance model, which is why it is often referred to as the ΛCDM cosmology.

[1]The densities are expressed in units of the critical density of the universe, $\Omega_0 = \frac{3{H_0}^2}{8\pi G}$, where H_0 is the present-day value of Hubble's constant.

[2]Until recently, the most widely held model of the universe was the so-called "Standard Cold Dark Matter" (SCDM) cosmology. Its main assumption, that there is enough mass to prevent the runaway collapse or expansion of the universe ($\Omega_m = 1$), has been disproved by the recent observational results.

1.2 Galaxy Formation and Evolution

1.2.1 Hierarchical Galaxy Formation

The current concordance model envisions a hierarchical sequence of structure formation. Structure in a Universe dominated by Cold Dark Matter develops as gravitational instability drives the growth of small primordial density fluctuations. Fluctuations on small scales have larger amplitudes and therefore are the first to collapse. Dark halos grow progressively larger as they acquire mass through accretion and mergers with other halos. The baryonic component of these halos follows this process but it may also cool and collapse towards the center of the dark matter halos, where it may start forming stars. Thus, according to the CDM paradigm, systems like the Galaxy form through a succession of mergers and accretion events. The process of accretion is likely to continue even today, as illustrated by the Sagittarius dwarf, which is currently being cannibalized by the Milky Way (Ibata et al. 1994).

1.2.2 Milky Way Formation Models

This CDM -inspired model for the formation of the Galaxy blends features of two competing models that have been intensely debated over the past four decades. The first model, developed by Eggen, Lynden-Bell & Sandage (1962, henceforth ELS), proposes that the Galaxy formed through the smooth monolithic collapse of a protogalactic cloud shortly after it decoupled from the universal expansion. Time scaling arguments imply that the collapse might have occurred very rapidly, on a time scale of about 10^8 years. ELS supported their model with data from high velocity stars, which appear to indicate that the lower the metal abundance of a star, the higher the energy and the lower the angular momentum of its orbit.

In a contrasting model, Searle & Zinn (1978, hence SZ) pointed out that Galactic globular clusters have a wide range of metal abundances, independent of Galactocentric distance. They argued that these data suggest that the halo was built through

the assembly of many small fragments with masses of about 10^8 $M_\odot$, on a longer time scale, of several Gyr.

The observational data suggests that the process of galaxy formation includes elements from both the above scenarios. According to Freeman & Bland-Hawthorn (2002), the process generally follows three distinct stages: The first one is dominated by dark matter, which virializes into a main halo. This stage may be accompanied by a process of active star formation, although this is not a necessary ingredient. In the second one, the gas dissipates energy and accretes smoothly into the center of the halo, forming a thin disk (in the case of a spiral galaxy) and, in some cases, the bulge. Finally, the accretion of small neighboring sub-halos contributes stars predominantly to the halo, but may also trigger new episodes of star formation.

Clues to the history of the Milky Way may be found in many parts of the Galaxy. Every component of the Galaxy: the halo, the bulge, and the disk preserve information about past events in the Galaxy's history. Within each component, the accretion history can be deciphered from the metallicity, age, spatial distribution and dynamics of different stellar populations, of globular clusters, and of the disrupted satellite galaxies orbiting in the Galactic field. For example, the disk of the Galaxy reflects mainly the epoch of smooth accretion of the gas, together with minor merger events that may be responsible for the morphological peculiarities of the disk (thickening, flaring, warping or the formation of a bar). The bulge may also contain fossil information; however, its history is more difficult to recover.

I will focus below on a few of such evolutionary signatures, with special emphasis on those closely related to the interaction between neighboring satellite galaxies and the disk. The purpose of reviewing these observational signatures is that they may be compared to similar features found in the numerical experiments presented in this thesis. Such comparison may shed light on the process of galaxy formation in CDM cosmogonies.

1.2.3 Accretion Signatures in the Milky Way's Halo

The stellar halo contains a detailed record of the assembly history of the Galaxy in the form of debris from past accretion events. Examples are the newly discovered coherent moving groups of stars in the halo (Majewski et al. 1994; Helmi et al. 1999), as well as the mean retrograde motion of some young globular clusters (Zinn 1993) and halo stars (Norris & Ryan 1989). Globular clusters are particularly intriguing accretion tracers. For example, ωCen, the most massive globular cluster in the Galaxy, may have been associated with a gas-rich dwarf galaxy (Lee et al. 1999). Recent modeling of its properties and evolution indicate that ωCen may have been the core of a dwarf elliptical accreted by the Milky Way more than 10 Gyr ago (Tsuchiya, Dinescu & Korchagin 2003).

The most conspicuous signatures in the halo are, of course, those of satellite galaxies. The majority of them today are low-mass, dwarf galaxies (van den Bergh 2000), but those that survive today may be just a small fraction of the galaxies that have been accreted during the entire lifetime of the Galaxy. Satellite accretion probably occurred mostly before the formation of the thin disk, but a few mergers are known to occur even today. One example is the Sagittarius dwarf, a new satellite in the process of full tidal disruption, recently discovered in our "neighborhood" (Ibata et al. 1994). Sagittarius moves on a polar orbit which intersects the disk frequently (the last passage was about 1 Gyr ago). Its current appearance is extremely distorted (Majewski et al. 1999), suggesting that Sagittarius was totally disrupted during its last pericentric passage. Several globular clusters and a long stellar stream accompany the debris of galaxy in its orbit (Bellazzini, Ferraro & Ibata 2003).

New observations continue to add evidence of the accretion history of the Galaxy. New tidal streams are discovered at an ever faster pace, in part due to the advent of new techniques and instruments. One of the latest discoveries is that of a ring of stars coplanar with the disk and very thin (less than 2 kpc in height), encircling the disk at a distance of about 18 kpc from the center (Newberg et. al 2002, Ibata et al.

2003). With a total stellar mass estimated to be about $10^8 - 10^9\ M_\odot$, it is likely the remnant of a disrupted satellite which started on an low inclination orbit (Yanny et al. 2003; Ibata et al. 2003; Helmi et al. 2003).

Not all accretion events leave a decipherable record. Some of the debris from the process of accretion is expected to survive in the plane of the orbit only for a few Gyrs (Johnston et al. 1996), after which it disperses and faints in surface brightness. However, if the CDM cosmology is correct, many hundreds of streamers should exist in the Galactic halo at the present time (Helmi & White 1999).

This is currently a very active observational field, both from the ground and from space. The advent of several satellite missions will make it possible to map in unprecedented detail the stellar structure in the halo of our Galaxy. Two upcoming satellite missions, the Space Interferometry Mission (SIM) and the Global Astrometric Interferometer for Astrophysics (GAIA) satellite, will measure positions of billions of stars in the halo with microarcsecond precision. This will allow for the precise measurement of distances to objects throughout the Galaxy and, in particular, to constrain the phase-space information for stars in tidal streams.

1.2.4 Accretion Signatures in Disk Components

Many spiral galaxies have two kinematically distinct disk components: the thin and the thick disks. The thin disk is most likely the end product of the quiescent, dissipative settling of gas at the center of the Galaxy and its subsequent transformation into stars. The origin of the thick disk is not fully understood, and several formation mechanisms have been proposed (a review of these mechanisms will be presented later in this Section). The exponential scale height [3] of the thin disk in the Milky Way is of order 300 pc, whereas that of the thick disk is $\sim$ 1 kpc (see Buser 2000 and references therein).

[3] Scale heights are defined as the vertical distances at which stellar densities decrease to a factor $1/e \simeq 0.37$ of their initial values on the surface of the disk.

Although dynamical information is lost as gas dissipates and settles into the thin disk, it still retains important clues to the conditions prevailing during and before its formation. The main signatures are preserved in the fundamental properties of the disk: its size, surface brightness, luminosity and rotation speed. These properties can be altered through later interactions with neighboring satellites, either through direct mergers or through resonant dynamical heating. Such processes may lead to structural changes in the disk, in the form of flaring, warping or bar formation, although many such effects may be short-lived, due to rapid orbital mixing in the disk. Major accretion events are thought to heat thin disks into thick disks.

Thick disks are, however, generally not easy to detect. Even in the case of the Milky Way, the thick disk has been discovered only recently, with the help of deep surface photometry and dynamical information (Gilmore & Reid 1983). In spite of the difficulty in detecting them, thick disk components have also been detected in several external galaxies, like IC 5249 (Abe et al. 1999), NGC 4565 (van der Kruit & Searle 1981a), NGC 891 (van der Kruit & Searle 1981b, Morrison et al. 1997), NGC 4762 (Tsikoudi 1980), in several edge-on S0 galaxies observed by Burstein (1979), and are ubiquitous in the sample of edge-on galaxies of Dalcanton & Bernstein (2002).

The complex history of the Galactic thick disk is reflected in its mixture of old and young stars and in its wide range of metallicities. The oldest stars are about 10 -12 Gyr (ages are usually inferred from different methods of stellar dating, like white dwarf cooling, radioactive decay rates or isochrone fitting), whereas their metallicities span $-2.2 \leq [Fe/H] \leq -0.6$ (Chiba & Beers 2000).

The Age-Velocity Dispersion as a Measure of the Heating History of the Milky Way Disk

A good tool for measuring dynamical heating in a disk component is the velocity dispersion of stars, measured in any of the three Galactic coordinates: radial ($u = dR/dt$), azimuthal ($v = R\,d\phi/dt$) and vertical ($w = dz/dt$). Kinematic data compiled

for several thousands of stars in the solar neighborhood (within $\sim$ 20 pc of the Sun) show that the velocity dispersion of the disk stars in all three directions (u, v, w) increases monotonically with the age of the stars (Wielen 1974; Dehnen & Binney 1998; Fuchs 2001, Nordström et al. 2004). Thus, younger stars, such as early-type stars, supergiants and early-type giants, have velocity dispersions in the range $\sigma_z \sim 10 - 20$ km s^{-1}, whereas older stars, like planetary nebulae, subgiants or white dwarfs, have velocity dispersions $\sigma_z \sim 40$ km s^{-1} (Mihalas & Binney 1981).

Thus, there seems to be unequivocal evidence for a strong correlation between stellar age and velocity dispersion. It is therefore important to investigate the cause of this relation and what it tells us about the different evolutionary processes operating in the disk. I review below the most important mechanisms that have been proposed as explanations for this correlation.

The proposed heating mechanisms can be divided in two major categories:

i) global mechanisms. The most notable example of this category is the heating by transient spiral arms. This mechanism suggests that the velocity dispersions of stars can be suddenly increased when the stars pass through a density enhancement in the disk, such as the spiral arms (Barbanis & Woltjer 1967; Sellwood & Carlberg 1984; Carlberg & Sellwood 1985; de Simone, Wu & Tremaine 2004). In this model, the increase in the random motion of the stars can be sustained through a continuous regeneration of spiral density waves in the disk. However, this mechanism fails to match quantitatively the observed increase in σ_z (Mihalas & Binney 1981).

ii) local mechanisms. These mechanisms involve, one way or another, gravitational interactions between stars in the disk and objects which can perturb their orbits. These encounters are believed to stochastically heat the stellar disk, such as:

- star-star collisions (however, they are ineffective in heating the disk; the relaxation time in star-star collisions is about the age of the Universe (Chandrasekhar 1960)),

- encounters between stars and molecular clouds (Spitzer & Schwarzschild 1951;

1953),

- indirect encounters with the satellite population in the galactic halo (resonant heating), or

- direct encounters with infalling satellites.

Theoretical studies suggest that the Galaxy has not undergone a major merger for quite a long time, possibly not since redshift $z \sim 1.5$ - 2 (Tóth & Ostriker 1992), and there is some evidence that suggests a major merger just before then (Gilmore, Wyse & Norris 2002).

All of the above mechanisms cause fluctuations in the gravitational potential and thus lead to heating of the disk. Freeman (1991) has argued that the age-velocity dispersion relation is consistent with the operation of multiple heating mechanisms (for example, the transient spiral arms can cause an increase in σ_z up to ~ 20 km s^{-1} for stars about 3 Gyr old, and an ancient merger could have brought the older stars to $\sigma_z \sim 40$ km s^{-1}. Are there other less degenerate signatures of past tidal accretion events?

Warp Signatures

Warps are probably the most intriguing morphological features in disk galaxies: despite their ubiquitousness in the Universe, we still lack a clear understanding of how they form and evolve. Many disk galaxies display warps, either in the traditional, symmetric shape - one side upwards, the other side downwards - or asymmetric (one sided warps). In the Local Group, for example, all spiral disk galaxies (the Milky Way, Andromeda, M33) are warped. Gas warps are relatively more conspicuous, as revealed in observations of HI emission at 21 cm (Sancisi 1976; Bosma 1981). Warps in stellar disks are in general more difficult to detect. This is because warps are usually found to form between R_{25} and the Holmberg radius $R_{\mathrm{H_0}}$ (i.e. where the surface brightness in B-magnitude reaches 25 and 26.5 mag arcsec^{-2}, respectively) (Briggs 1990), where the optical disk is very faint. However, the evidence seems to

suggest that stellar warps are about as frequent as gas warps; large surveys of edge-on disk galaxies show that approximately half of all disk galaxies have optical warps (Sánchez-Saavedra et al. 1990).

The mechanism for generating and maintaining warps remains elusive. Several explanations have been proposed, but no consensus has been reached. It is well known, however, that once created, warps "wind up" quickly and do not survive for a long time (Binney & Tremaine 1987, p. 346; henceforth BT87). This is because different regions in the disk precess at different rates, and therefore warps gradually wind up into a spiral corrugation wave and disperse into random motions in the disk (hence leading to some thickening of the disk). Thus, in order to explain the ubiquity of warps in the nearby Universe, either mechanism that generates them has to operate continuously or, alternatively, a way must be found to sustain them over an extended period of time.

I now briefly review some of the most popular mechanisms which have been proposed to explain galactic warps:

- *non-spherical halos:* One possibility is that warps arise from the interaction between the disk and the massive dark matter halo in which it resides. One of the most promising mechanisms investigated along this line assumes that the halo is non-spherical. This model finds support in numerical simulations which show that dark halos are generally flattened (Frenk et al. 1985; Dubinski & Carlberg 1991). Observational studies also suggest that some galaxy halos are flattened (for a review on this subject, see Sackett 1999). According to this model, a misalignment between the direction of the principal symmetry axis of the halo and that of the rotational angular momentum of the disk, induces a torque in the disk that may be able to lift some of the disk stars in a warp. Studies which employed a fixed potential of a non-spherical halo suggest that warps can be stabilized (Sparke & Casertano 1988). On the other hand, simulations using a "live" halo (made of self-gravitating particles) have

not confirmed this result (Binney et al. 1998).

- *disk - satellite interactions:* Another mechanism suggests that satellite-disk interactions may create the necessary tidal torque. In particular, it was proposed that the warp in the Milky Way might have been triggered by a close-by passage of the most massive satellite galaxy, the Large Magellanic Cloud (LMC) (Murai & Fujimoto 1980; Gardiner & Noguchi 1996). This proposal, however, runs counter to the calculation of Hunter & Toomre (1969), who show that the LMC does not have enough mass to generate a warp of the right magnitude [although see Weinberg (1989) for a different argument which suggests an amplification of the effect through a wake in the halo]. The failure to connect the Galactic warp with the LMC has led to a decrease of the interest in this mechanism. In addition, observations of warps in external galaxies, seemed to suggest that some warped spirals do not have any detectable companions (Sancisi 1976).

  Recently, however, this mechanism has received renewed attention due to some surprising results. The galaxy NGC 5907, considered as the favorite example of a warped spiral without a companion, has recently been found to have one (Shang et al. 1998). Moreover, a new satellite companion has been discovered in the Milky Way, the Sagittarius Dwarf Galaxy, and several studies indicate that, with some tuning of its mass and orbital parameters, this satellite can produce the warp (Velázquez & White 1995; Ibata & Razoumov 1998; Bailin 2003).

  A related mechanism has also been recently reintroduced by Binney and collaborators (eg. Binney 1992; Jiang & Binney 1999): these authors envisage that the continuously infalling material during the epoch of galaxy formation may induce misalignments between different shells in the halo with the consequence of creating tidal torques in the disk that may produce a warp. Of course, in a realistic cosmological setting, a galactic disk experiences gravitational torques from *any* non-spherically symmetric distribution of mass around it, be it in a

form of a triaxial halo or a surrounding distribution of satellite galaxies.

In this thesis, we intend to address only the disk - satellite interaction mechanism and analyze in depth the effects that substructure in the halo might have on the warping of a galactic disk. More details on the motivation for pursuing this test are presented in Section §1.4.2.

1.3 Problems with the CDM Paradigm

In contrast with the many successes of the CDM paradigm on large scales alluded to in §1.1, several problems appear when one tries to match observations on the scale of individual galaxies. A few of the main discrepancies between the theory and observations are the following:

- The density profile of cold dark matter halos is expected to rise sharply towards the center of the system (this feature is also known as the dark matter central "cusp"; see, for example, Navarro, Frenk & White 1996). Rotation curves of low surface brightness galaxies suggest that the central dark matter density profiles are not cuspy, but have a core (Flores & Primack 1994; Moore 1994), with an approximative size estimated to be around a few kpc. There is still some controversy around the issue of observational effects, like beam smearing, that could, in some cases, mimic a core. When these effects are taken into account, there seems to be some marginal consistency with cuspy dark matter profiles (van den Bosch and Swaters, 2001), although in some conspicuous cases cuspy profiles can be excluded (Coté et al., 2000; van den Bosch et al. 2000; de Blok et al., 2001). Also, recent observations have revealed that in several clusters of galaxies (Tyson, Kochanski & Dell'Antonio 1998; Sand et al. 2003), the dark matter distribution in the inner regions is flatter than expected. The same effect is observed for some spiral galaxies (Binney & Evans 2001; Davé et

al. 2001), massive ellipticals (Keeton 2001), as well as for some dwarf galaxies, like Sculptor and Draco (Stoehr et al. 2002; Kleyna et al. 2003).

- Recent numerical simulations based on the CDM cosmologies, in which halos assemble through a sequence of mergers, have difficulty accounting simultaneously for the mass, luminosity, rotation speed, and angular momentum of galaxy disks accreted in these halos (Navarro & Steinmetz 1997, 2000).

- Another problem, which was first identified in theoretical and semi-analytical models of galaxy formation (White & Rees 1978, Kauffmann & White 1993), and has been re-emphasized by the advent of high resolution numerical simulations (Moore et al. 1999, hereafter M99; Klypin et al. 1999, hereafter K99) is the "dwarf satellite problem". Simply stated, virialized DM halos are abundant in substructure halos that outnumber low mass luminous satellites in the local universe by a least an order of magnitude. For example, in a recent Local Group census, Mateo (1998) identifies only about 40 such objects and suggests that observations may miss, at most, other $15 - 20$ dwarf galaxies; however, CDM simulations suggest that 10 or 100 times as many dwarf DM substructure halos may be present on the same scale.

- Finally, a problem that stems directly from the one above, is that the dark matter satellites may dramatically affect the internal structure of stellar galaxy disks, in contradiction with observational constraints (M99). This particular aspect will be investigated in detail in this thesis.

An intense effort over the past few years has been directed towards finding the cause of these (and other) discrepancies, with proposals covering a wide range of possibilities. The more exotic ones advocate entirely different cosmological models, but others try, more conservatively, to refine the current paradigm. For example, since the satellite problem is a direct consequence of the power on small scales in the CDM cosmology, several alternative models have been proposed, where certain

physical processes erase the amplitude of density perturbations on small scales and, consequently, reduce the number of low mass halos. These scenarios invoke a different nature of the dark matter particles, in the form of warm, repulsive, self-interacting, annihilating, decaying or fluid dark matter (Hogan and Dalcanton, 2000; Goodman 2000; Davé et al. 2001; Cen 2001; Peebles 2000) or of a broken scale invariance obtained by modifying the process of inflation (Kamionkowski & Liddle, 2000).

The more conservative alternatives suggest, rather, that we have failed to detect the dark matter satellites in the Local Group due to their lack of luminous matter. The idea is that dark matter satellites may be actually present in the number predicted by the CDM model but that they are dark because either the baryonic (hence luminous) material was removed long time ago through supernova winds (Dekel & Silk 1986; Mac Low & Ferrara 1999) or because the gas has been prevented from settling in. The last argument relies on the fact that the gas needs to cool in order to fall deep into the potential well of dark matter halos and form stars. External radiation, such as the high energy background radiation produced during the epoch of reionization, may heat the gas efficiently in low mass halos (Rees 1986; Babul & Rees 1992; Efstathiou 1992). Both mechanisms may in principle explain the discrepancy between the number of observed satellites and the dark matter halos, as well as why it occurs in halos with circular velocities of order $v_c \sim 10 - 30$ km s^{-1} (Bullock et al. 2000).

It therefore appears that the overabundance of low mass satellites can be accommodated within the CDM paradigm or at least, that in itself, it does not automatically invalidate the cosmological model. It would certainly be a major breakthrough in this problem if the low-mass dark matter halos could be detected directly – through gravitational lensing, for example (Dalal & Kochanek 2002; Ostriker & Steinhardt 2003). For the time being, a less stringent conjecture would be to assume the validity of the CDM theory and to test the consequences of the "satellite problem", looking for disagreement with the observations. One useful such test is to study the changes in the structure of thin stellar disks, as a result of interaction with the dark matter

satellites. I present below a detailed motivation for pursuing such a test, with particular emphasis on its application to the Milky Way (referred also from now on as "the Galaxy").

1.4 Goals of the Present Study

1.4.1 Heating of Stellar Disks by CDM substructure

The first goal of this study is to establish whether the substructure in galactic cold dark matter halos is consistent with the existence of thin disks. Thin stellar disks are fragile structures, prone to substantial perturbations in a dynamically evolving potential. Strong tidal interactions with dark matter satellites are one avenue for creating a fluctuating potential. As previously discussed, these interactions can increase the velocity dispersion of disk stars and, in extreme cases, induce severe morphological changes leading to the total disruption of the disk. However, we do not have, currently, any stringent constraints on the frequency of these interactions with the disk. The ubiquitousness of galactic disks (see van der Kruit & Searle 1982, for example) argues against this being a dominant mechanism in the evolution of spiral systems, as spiral galaxies are seen up to redshift of $z \sim 1$ (Lilly et al. 1998). The best documented example of a thin stellar disk is, of course, that of our own galaxy, the Milky Way. The existence of the thin component of the Galactic disk precludes any recent catastrophic mergers. However, several examples of minor accretion events ($M_{sat}/M_{disk} < 0.2$) are documented in the Galaxy. The Sagittarius dwarf galaxy provides an example of an ongoing minor merger. Although this satellite galaxy is not very massive (its mass is estimated to be around $10^9 M_{\odot}$), it seems to be in an advantaged stage of merging and might leave a signature on the disk structure.

Large scale structure cosmological simulations support the idea that mergers are important events in the lifetime of a galaxy. Mergers with satellites of masses comparable with that of the disk are more likely to occur at early stages of a galaxy (Abadi

et al. 2003). However, minor mergers (in the range $M_{sat}/M_{disk} \sim 0.05 - 0.2$) are also thought to be common events at the present time ($z \sim 0$). Important questions that need to be answered concern the frequency of major and minor mergers in the lifetime of the Galaxy and the impact they have on the disk.

Because of the complexity of interactions, numerical simulations are the ideal tool for investigating the gravitational interaction between substructure and disks and to gauge the effect of substructure on disk heating. Numerous simulations performed to date have shed light on a number of issues concerning the relation between the orbital parameters of satellites and the heating of the disk. I summarize below the main results of these simulations. I also discuss briefly their limitations and provide motivation for pursuing improvements.

Prior work

Quinn & Goodman (1986) performed one of the first numerical studies of this kind, by constructing a model of the Milky Way galaxy interacting with satellites with mass ratios ranging between $0.01 - 0.04$. They provide an estimate of the decay rates as a function of satellite mass and orbital parameters and show that orbital decay occurs faster for prograde and low-inclination orbits. These authors also investigate the energy transfer between satellites and disk and find that the kinetic energy is deposited mainly in the plane of the disk (i.e. radially), rather than in the vertical direction. This result has important implications for disk heating, suggesting that satellites might not be too efficient in increasing the disk height. Although the study of Quinn & Goodman (1986) captures the basic properties of satellite dynamics, their quantitative predictions are likely to be affected by the simplified methods they employed (eg. the galactic components and the satellites are modeled as rigid bodies, a procedure which neglects the effect of dynamical friction). Quinn, Hernquist & Fullagar (1993, hence QHF93) presented an improved study, which includes self-consistent models (i.e. N-body models) of both the disk and the satellite. The halo, however, is still modeled with a fixed potential, its size is too small to be realistic ($\sim$

21 kpc), and the bulge component is neglected. They perform several simulations by varying the mass of the satellite (assuming mass ratios M_{sat}/M_{disk} as 0.04, 0.1 and 0.2) and considering circular orbits at different inclinations (0^o, 30^o and 60^o) with respect to the plane of the disk. Overall, the average disk height increases roughly by a factor of 2, with most of the heating occurring at large radii.

The authors also investigate the effect of multiple mergers and find an important result, namely that subsequent mergers become less and less efficient at increasing the disk height (for example, they find an $\sim$ 115% increase in the mean scale height after first impact, in contrast with about 16% increase after the second impact). It thus appears as if the heating process achieves a "saturation" level, after which disks are more difficult to heat up (Freeman 1991; Quillen & Garnett 2001). Walker, Mihos & Hernquist (1996) refined one of QHF93's simulations (having as initial conditions a 10% M_{disk} satellite which evolves on a circular orbit inclined at 30% from the plane of the disk), by modeling all galactic components self-consistently. They also improved the numerical resolution by increasing the number of disk particles by a factor of $\sim$ 7 and the number of satellite particles by a factor of $\sim$10 (the total number of particles used in the simulation was 500, 000, of which 45% in the disk, 45% in the halo and 10% in the satellite). The heating of the disk obtained in this simulation was only slightly lower than the value measured by QHF93. They conclude that the satellite with the above chosen orbital parameters can, by itself, reproduce the heating observed in the solar neighborhood. The main limitation of this simulation is that the cut-off radius of the dark halo and the initial starting point of the satellite are still unrealistically small ($\sim$ 21 kpc). Thus, the evolution of the satellite (in terms of tidal stripping and orbital decay) prior to reaching the vicinity of the disk is neglected.

Huang & Carlberg (1997, hence HC97) construct a self-consistent model of the Galaxy, in which satellites start at larger distances (about three times the initial radius of the disk), on orbits with eccentricity[4] $e = 0.2$. However, the size of the halo

[4]For an ellipse, the eccentricity is defined as $e \equiv \sqrt{1 - \frac{b^2}{a^2}}$, where a and b are the the semi-major and semi-minor axes of the ellipse, respectively.

is relatively small, with a half mass radius of 16 kpc. Also, they choose to study the evolution of low-density satellites, in order to match the surface brightness of dwarf galaxies (in contrast with previous studies which construct high-density satellites that match the constraints of dark matter profiles). In general, the disk is found to respond to the infalling satellites mainly by tilting rather than heating. Satellites with masses 10% and 20% M_{disk} are found not to thicken the disk significantly, whereas the satellite with 30% M_{disk} mainly thickens the disk in the outer parts. This result is not surprising since low density satellites will suffer a more accelerated process of tidal stripping and, therefore, lose more mass before colliding with the disk. This process, in turn, can lead to a reduced disk heating rate.

Finally, the most recent numerical study of disk heating is that of Velázquez & White (1999, hence VW99). This study includes several notable improvements. Firstly, the authors explore a large parameter space that characterizes the initial conditions of the satellites by varying their masses, concentrations, orbital inclinations and eccentricities. Secondly, the disk, bulge, halo and satellites are all modeled self-consistently. Thirdly, the halo has a larger cut-off radius, of 84 kpc (although this value is still too small in comparison with the values predicted by cosmological simulations for the size of the dark matter halo of the Galaxy). The authors find that the heating of the disk is less significant when the satellite has a lower concentration; when it is on a retrograde orbit; or when a massive bulge is present.

Although the results of the above simulations reveal important aspects of the disk heating process, putting it all together in a consistent picture of disk evolution is not straightforward. In the first place, a comparative analysis of these simulations is difficult to make given the large differences in numerical resolution or in the orbital parameters and masses of satellites and other galactic components. It is therefore unclear to what extent the discrepancies among different results are real or, rather, depend on the particular choices of initial conditions. It is also difficult to obtain a general picture of the disk heating from a series of fragmented analyses which consider the interaction of the disk with one satellite at a time. Some studies try to

alleviate this problem by adding another merger to the final state of the first merger (QHF93), or by analyzing the effect of varying the inclination, eccentricity and mass of the satellites (VW99). However, up to the present, no study has attempted to construct a realistic distribution of satellites with initial conditions consistent with CDM models.

The analytical study of Tóth & Ostriker (1992, hence TO92) offers a more general approach, by incorporating an estimate of the rate of satellite accretion in cosmological models, combined with an appropriate model for the dynamical evolution of satellites (which includes dynamical friction). The results of this study have placed severe constraints on the amount of mass that the Galaxy may have accreted. Namely, these authors find that the Galaxy could not have accreted more than 4% of its mass interior to the solar circle ($R_\odot = 8.5$ kpc) during the last 5 Gyr. The authors conclude that this prediction is in apparent contradiction with the high merging rate expected in an $\Omega_m = 1$ (SCDM) universe, and argue that their analytical constraints favor an $\Omega_m < 1$ Universe.

The conclusion reached by TO92 has been intensely debated. HC97 point out that a low accretion rate near the solar circle does not necessarily imply a low accretion at larger distances. Also, VW99 compare the increase in the disk heating in their simulations with the prediction derived from TO92's formalism and argue that TO92's analytical argument overestimates the heating in solar neighborhood by a factor of $2 - 3$. Several simplifications in TO92's analytical treatment may be responsible for this discrepancy – such as the dynamical modeling that does not take into account the coherent response of the disk and the halo, or the too strict assumption that the energy is deposited locally by the satellite. Most importantly, TO92's modeling of satellites as rigid potentials neglects the effect of tidal stripping, which can destroy some satellites even before they can reach the disk.

A more accurate estimate of the accretion rate is provided by cosmological simulations. It is only in the past few years that cosmological simulations have achieved enough numerical resolution to enable us to identify the accretion rate and orbital

parameters of the substructure in a galaxy-sized halo (M99; K99). Based on statistics of orbital parameters (pericenter-to-apocenter ratios) of dark matter satellites in high resolution simulations, M99 suggested that the heating of the Milky Way disk by the CDM substructure would be inconsistent with observational constraints. In particular, the authors estimate that the energy pumped by subhalos into a disk like that of the Milky Way may add up to a significant fraction of the binding energy of the disk over a Hubble time. Estimating analytically the heating rate due to the cumulative effect of disk/subhalo collisions is, however, quite uncertain, and in practice simple calculations that neglect the self-gravitating response of the disk often overestimate the effects of heating by infalling satellites. A more promising approach to this problem requires a self-consistent numerical simulation that includes, along with the predicted CDM substructure, a realistic galaxy disk.

The aim of this study is to determine the effect that a cosmological realization of dark matter satellites has on thin stellar disks and to estimate its contribution to their heating rates. This work revisits the M99 conclusion using self-consistent numerical simulations, by taking into account the mass spectrum, orbital distribution, and internal structure of subhalos identified in cosmological simulations of galaxy-sized CDM halos. As in previous studies of disk heating, this study chooses initial conditions so as to reproduce the case of the Milky Way, where we have fairly good observational constraints on the internal structure of the Galactic components. With this realistic numerical model, it becomes possible for the first time to address a number of questions, such as:

- is disk heating produced mainly by a few massive satellites?
- what is the cumulative effect of many minor mergers?

and

- how does the disk re-adjust after recurrent interactions with substructure halos?

1.4.2 Disk - Satellite Tidal Interactions as a Mechanism for Generating Warps

Another question addressed in this study relates to the role of CDM substructure in triggering the formation of warps. As described in Section §1.3.2, the idea that satellites can induce warps is not new (QHF93; HC97; VW99). However, no detailed investigation of the properties of these warps in the context of the CDM cosmology has been attempted to date. Our aim is to follow up this line of investigation by modeling the complete distribution of satellites around a typical spiral disk in order to determine if CDM substructure can, in fact, trigger warps of comparable magnitudes to those observed in the local Universe.

Our analysis suggests that this mechanism is well suited for explaining the origin of transient asymmetric warps, such as that of the Milky Way. The Galactic warp is indeed slightly asymmetric – extending $\sim$ 2 kpc on one side and less than 1 kpc on the other side (Burton 1988). The connection between the Galactic warp and the Milky Way satellites is not yet fully understood, although, as mentioned above, the Large Magellanic Cloud is an unlikely culprit. The most promising candidate remains the Sagittarius dwarf galaxy. In this scenario, the asymmetry is supposed to have been created through the last two consecutive interactions between this satellite and the disk (Lin 1996).

The other large spiral galaxy in the Local Group, Andromeda (or M31), displays a very well defined asymmetric warp (Choi, Guhathakurta & Johnston 2002). Although the mechanism for generating this warp has not yet been elucidated, both theoretical and observational studies have narrowed the search down to the two closest satellites of M31: NGC 205 and M32 (although see Morrison et al. (2003) for a suggestion that the newly discovered satellite And VIII may be the culprit). The answer to the question of which of these satellites is responsible for the warp awaits accurate determination of their orbital parameters. Finally, also in the list of famous warped galaxies, we find NGC 5907. This edge-on spiral galaxy also has an asymmetric warp

which may originate from the interaction with its newly discovered companion, a small satellite in a polar orbit, which is severely disrupted along a large tidal stream (Shang et al. 1998). However, this galaxy also has a flattened luminous halo (Sackett et al. 1994). If the dark matter halo of NGC 5907 is also flattened, this in principle can offer an alternative explanation for the warp. We note, however, that the stellar light detected by Sackett et al. (1994) is not a direct tracer of the dark matter content in this galaxy (H. Morrison, private communication).

Although the simulations presented in this thesis have not been tailored to match any of the above systems, I will discuss how our numerical results can be rescaled to these particular situations and what lessons may be learned from such rescaling. In addition, I emphasize that the present numerical simulations represent a significant improvement over previous studies [eg. Velázquez & White (1995); Ibata & Razoumov (1998)], in that they span a much longer time scale, of $\sim$ 10 Gyrs, making it possible to follow the evolution of a satellite for many orbits. This is important because we have clear evidence that some of the Local Group satellites have been through multiple pericentric passages – for example, Sagittarius (Lin 1996) – and therefore, may have induced several transient warps in the past.

All simulations presented in this thesis were run with PKDGRAV (Stadel 2001), a parallel N-body tree code well suited for this problem, being specifically constructed to adapt spatially to a large range in particle densities and, temporally, to a large range of dynamical timescales. The runs were carried out on the 40-node Beowulf PC cluster at the Department of Physics and Astronomy of the University of Victoria.

Chapter 2

Dark Matter Substructure in ΛCDM and SCDM Galactic Halos

Abstract

In order to build realistic models of spiral galaxies, one needs to know the distribution and the physical properties of the population of dark matter satellites which orbit in the galactic halo. In this chapter, we make use of a group finder algorithm (SKID) in order to identify the dark matter substructure in several representative galactic halos extracted from large cosmological simulations, in both SCDM and ΛCDM models. Our results confirm previous findings in the literature, namely that the number of dark matter sub-halos exceeds the number of visible satellite galaxies by a large factor (M99; K99). In addition, we investigate the orbital motion of galactic dark matter sub-halos and find that, in general, there is no significant excess of satellites in either prograde or retrograde orbits. This result is in good agreement with recent observational studies.

2.1 Introduction

In this Section, we will briefly revisit the "dwarf satellite problem" alluded to in Chapter 1, with particular emphasis on the population of substructure halos found

in N-body cosmological simulations.

In the last decade, advances in computational power, coupled with the advent of commodity parallel computers, have opened the way to running simulations with millions of particles. The high numerical resolution achieved in recent years has brought to light a very important result for CDM halos: during the merger events that characterize the assembly and growth of dark matter halos, the accreted halos may survive as self-bound entities in the central regions of the parent halo for several orbits (Ghigna et al. 1998; Klypin, Gottlöber, Kravtsov & Khokhlov 1999). The recent high resolution simulations show not only that dark matter satellites survive in the tidal field of the galaxy, but that they also exceed in number the known Galactic satellites (K99; M99). Some of the consequences of this new problem, known as the "dwarf satellite problem", will be discussed further in the current and following chapters.

Section §2.2 presents information about the dark matter halos used in this study as well as the large scale simulations from which they were extracted. Section §2.3 describes the procedure which identifies the substructure in these halos and presents a comparison with other results in the literature. Section §2.4 investigates the orbital motions of the dark matter satellites and finds that they are in good agreement with observations.

2.2 The Cosmological Simulations

The present study draws its initial conditions for the dark matter satellites from several large scale cosmological simulations. Two different cosmologies are examined: the so-called "standard" CDM (hence SCDM) and the ΛCDM models. These simulations were performed prior to this thesis and the data were provided in electronic format by their authors (S. Ghigna (SCDM); J. F. Navarro (ΛCDM)).

The cosmological parameters (Ω_{tot}, Ω_M, Ω_Λ, h and σ_8), as well as the simulation parameters (the size of the simulation box and the total number of particles of high

resolution particles in the simulation, $N_{tot}^{high\,res}$) are shown in Table 2.1.

Table 2.1: Parameters of the cosmological simulations

Model	Ω_{tot}	Ω_M	Ω_Λ	h	σ_8	Box ($h^{-1}\ M_\odot$)	$N_{tot}^{high\,res}$
SCDM	1	1	0	0.5	0.7	50	7,450,276
ΛCDM (128^3)	1	0.3	0.7	0.65	0.9	32.5	980,416
ΛCDM (64^3)	1	0.3	0.7	0.65	0.9	32.5	262,144

The SCDM simulation follows a "high-resolution" region in a cosmological box 100 Mpc on a side that forms two large halos, similar in size and kinematics to the two large galaxies in the Local Group, the Milky Way ("MWy") and Andromeda ("M31"). In the case of the ΛCDM simulation, the system was selected from a $32.5\,h^{-1}$ Mpc periodic box and resimulated at higher resolution. Both simulations used the PKDGRAV code developed by Joachim Stadel & Thomas Quinn (see Stadel 2001). The procedure for resimulation is analogous to that described in Eke, Navarro & Steinmetz (2001), and is similar (but not identical) to that used for the SCDM "Local Group" halos (for a detailed description of the procedure used for SCDM halos, see Ghigna et al. (1998) and M99). The mass and force resolution are similar in both SCDM and ΛCDM runs. The number of particles in the "high resolution" region is given in Table 2.1. Using the same procedure and similar initial conditions, the ΛCDM halo has been run with several choices of particle resolution (Power et al. 2003). Table 2.1 shows the parameters of two such runs, labeled as 128^3 ($N_{tot}^{high\,res} = 980,416$) and 64^3 ($N_{tot}^{high\,res} = 262,144$) simulations.

Table 2.2 shows several physical parameters of the halos, measured at the virial radius ($R_{vir} \equiv R_{200}$); the radius where the mean inner density is 200 times the critical density for closure[1]. The two halos in the SCDM model, "M31" and "MWy"[2], have

[1]The virial mass is defined as $M_{vir} \equiv M_{200} = M(R_{200})$, and the circular velocity at R_{200} is $V_{vir} \equiv V_{200} = (R_{200}/\mathrm{kpc})h$ km s^{-1}, at $z = 0$.

[2]Alternatively, we will refer to these two SCM runs as "Halo 1" and "Halo 2", and use these

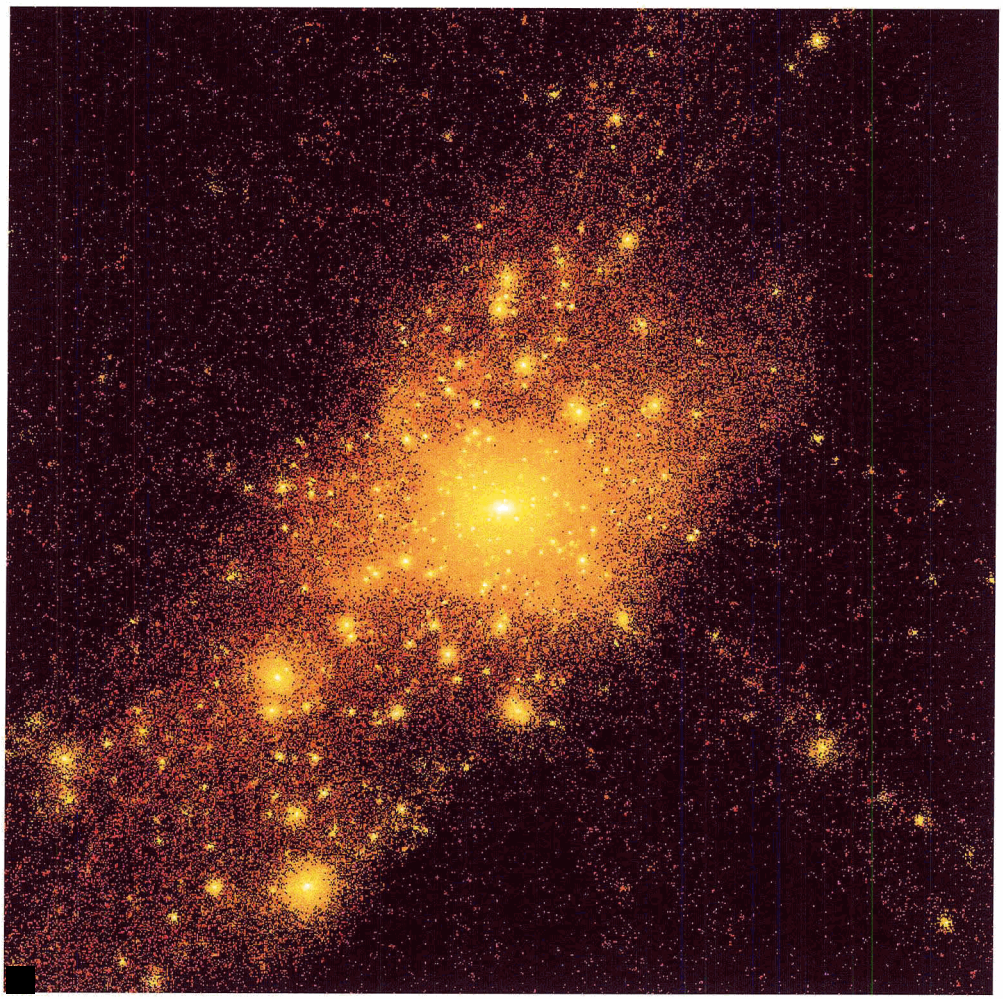

Figure 2.1: The ΛCDM galaxy-sized dark matter halo (high resolution simulation).

V_{vir} = 193 and 172 km s^{-1}, respectively, and each has roughly 10^6 particles within $R_{vir} \sim 350$ kpc. The galactic halo in the ΛCDM has $V_{vir} \sim 202$ km/s, a virial radius of about 290 kpc and a number of particles within R_{vir} of 369,991 for the 128^3 run and 48,724, for the 64^3 run, respectively. We note that the SCDM halos were identified at $z \sim 0.1$, whereas the ΛCDM halo was identified at $z \sim 0$.

notations interchangeably throughout the thesis.

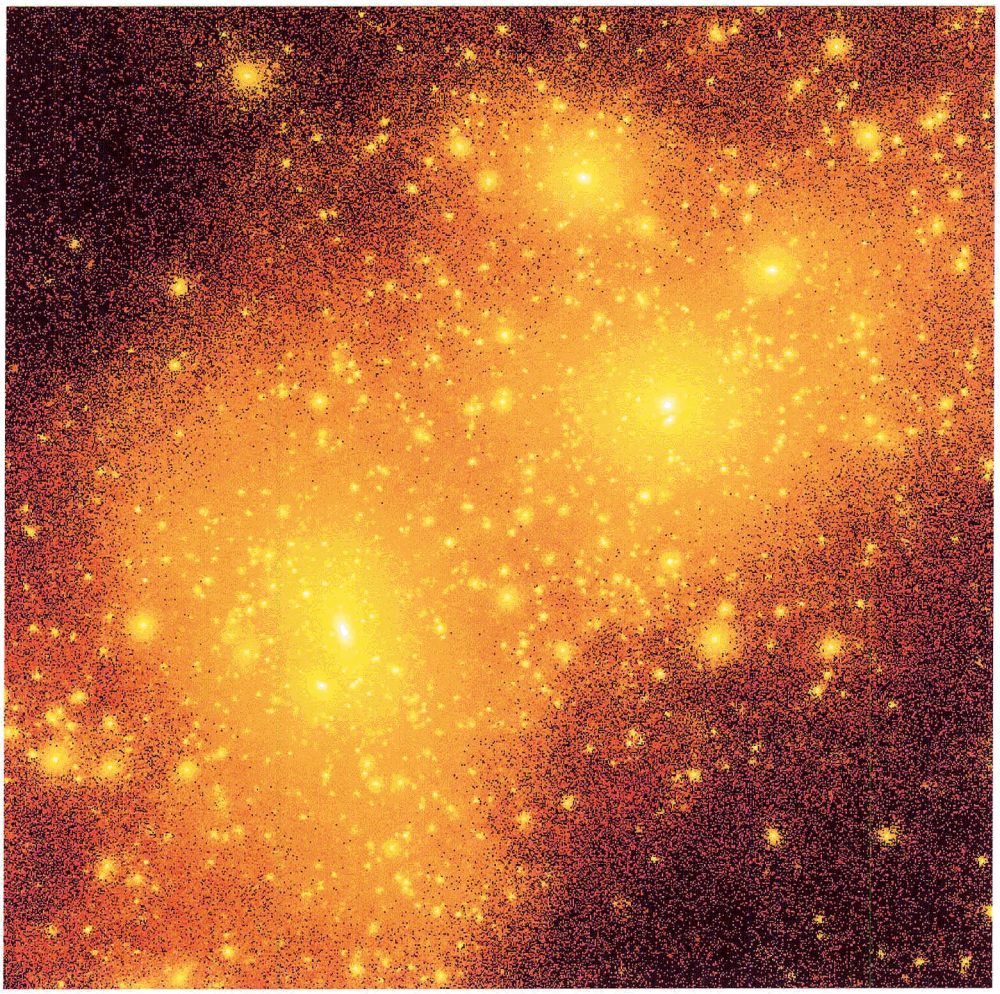

Figure 2.2: The two galaxy-sized dark matter halos in the "Local Group" SCDM simulation (adapted from M99 data).

Table 2.2: Parameters of the galactic halos identified in the cosmological simulations

Model	Halo	R_{vir} (kpc/h)	M_{vir} $(10^{10}\,M_{\odot}/h)$	V_{vir} (km/s)	N_{vir}	M (20 kpc) $(10^{10}\,M_{\odot}/h)$
SCDM	"MWy" ("Halo 1")	163.7	102.1	172	1,225,172	10.7
SCDM	"M31" ("Halo 2")	183.8	144.5	193	1,707,785	11.9
ΛCDM	128^3	202.6	193.489	202.6	369,991	14.3
ΛCDM	64^3	205.8	202.785	205.8	48,724	13.3

2.3 The Number of Substructure in Galactic Halos

2.3.1 Identifying Substructure

Identifying substructure in dark matter halos is crucial to determine the mass function of substructure halos, their internal structure, and their orbital parameters. This is not a straightforward task since, in general, there is no clear-cut separation between the particles in the sub-halos and those in the background. In some cases, the sub-halos may be tidally distorted and part of their material may be unbound. These and other considerations have to be taken into account in a physically motivated algorithm. The group finder SKID[3] is an example of such algorithm, and numerous tests have shown that it outperforms the previous generation of group finding algorithms (such as "friends-of-friends" or DENMAX[4]) in determining the number of substructure and its physical properties (Governato et al. 1997; Ghigna et al. 1998).

SKID is an improved version of the DENMAX algorithm and its main task is to find positions of local maxima in the density field, smoothed on a certain scale. Once the density peaks are found, the code then tests whether the locations correspond to self-bound halos. In doing so, it groups particles which converge towards a common high density peak, as they are allowed to flow along the lines of the density gradient (Stadel 2001). All particles which are unbound to the group (i.e which do not have negative binding energy in the reference frame of the center of mass of the halo) are afterwards discarded. At the end of this procedure, the code outputs the main properties of the identified halos (for example, the mass, distance, peak circular velocity, outer velocity) in a catalog in an electronic format.

Since SKID has been designed for the general purpose of identifying sub-halos on a wide range of spatial scales, ranging from clusters of galaxies down to galaxy-size halos, it has a number of parameters which need to be adjusted to the particular problem under scrutiny. SKID's main input parameter is a characteristic length

[3] `http://www-hpcc.astro.washington.edu/tools/skid.html`

[4] `http://www-hpcc.astro.washington.edu/tools/fof.html`

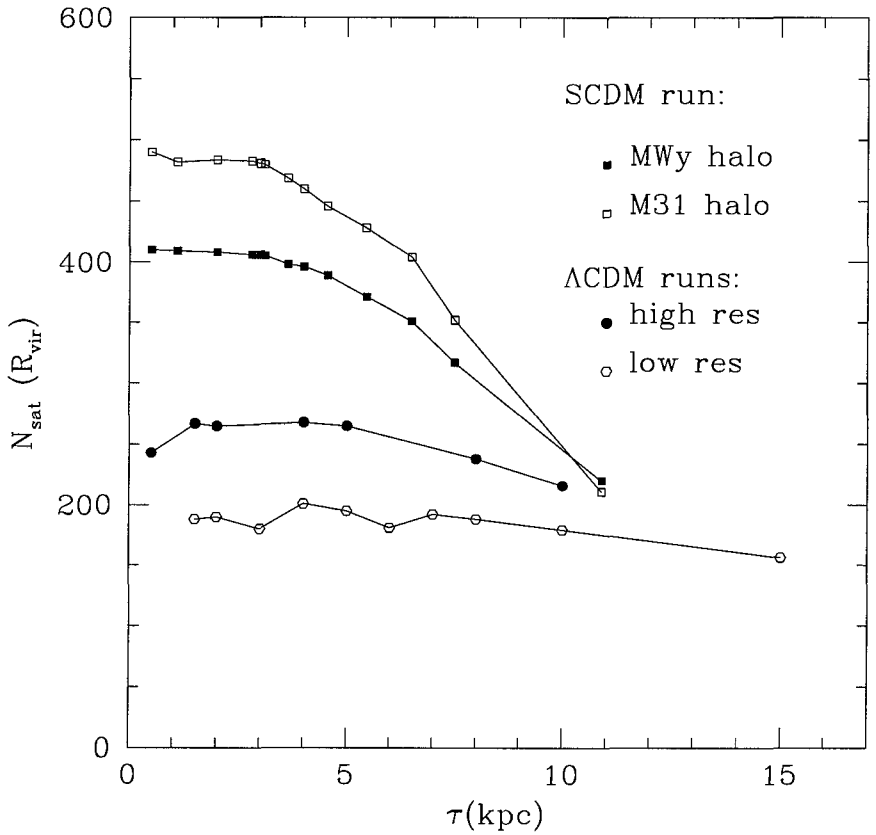

Figure 2.3: The effect of varying the linking length, τ, on the group finding. $N_{sat}(R_{vir})$ represents the number of sub-halos identified in the SCDM ("MWy" and "M31" halos) and ΛCDM runs (both high and low resolution) within the virial radius ($n_{min} = 16$ for SCDM halos and $n_{min} = 8$ for the ΛCDM halo).

scale over which the converged particles are initially linked together (i.e., the "linking length"). The linking length determines the characteristic size of the smallest groups and, indirectly, the minimum separation between clumps. The other free parameter is the minimum number of particles, n_{min}, below which the algorithm discards the collection of particles in the density peak as being not part of a group. The choice of these parameters is important, because an inappropriate value can alter the information about substructure (the code may not record all the groups or, on the contrary,

merge spurious ones). It is therefore important to know to what extent the value of these parameters affects the identification of substructure.

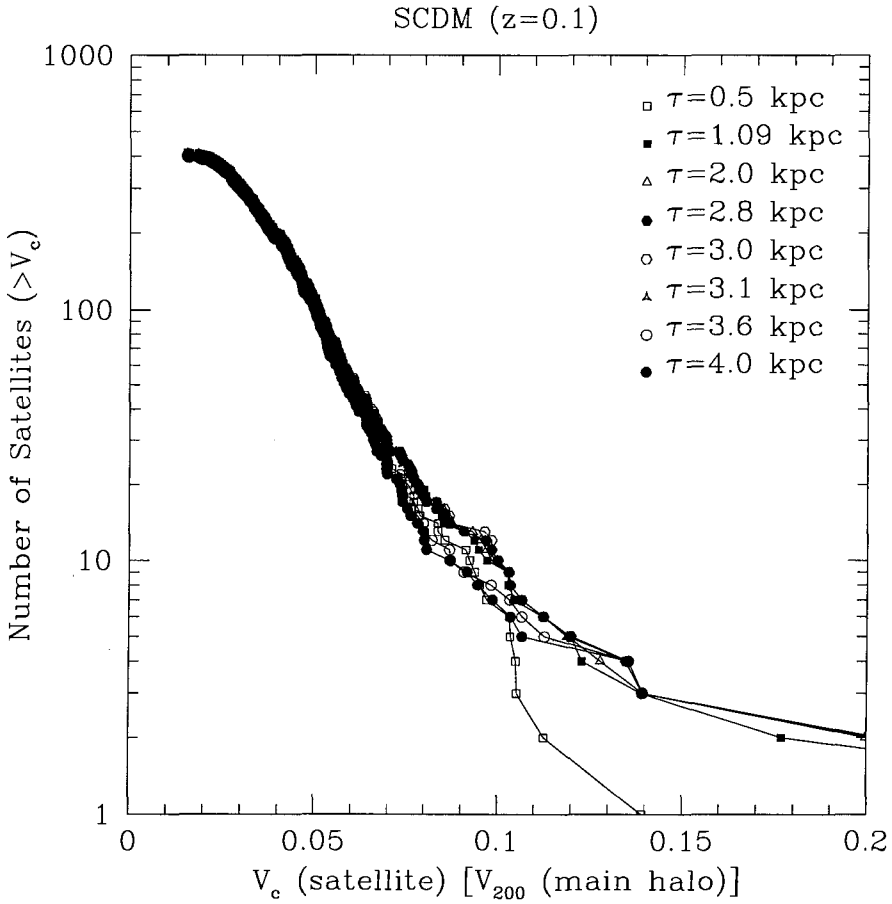

Figure 2.4: Cumulative number of sub-halos within R_{vir} in the SCDM ("MWy") halo, obtained with various SKID linking length parameters. V_c represents the maximum circular velocity of the sub-halos.

For this purpose, we performed several tests by varying these two free parameters in SKID. Figure 2.3 shows the role that the linking length, τ, plays in determining the total number of sub-halos within the virial radius of the main halo, $N_{sat}(R_{vir})$. A number of $n_{min} = 16$ particles were used for determining the substructure in both SCDM halos and $n_{min} = 8$ for the ΛCDM halo, respectively. The figure shows that a large linking length (e.g., $\tau > 5$ kpc) can drastically underestimate the number

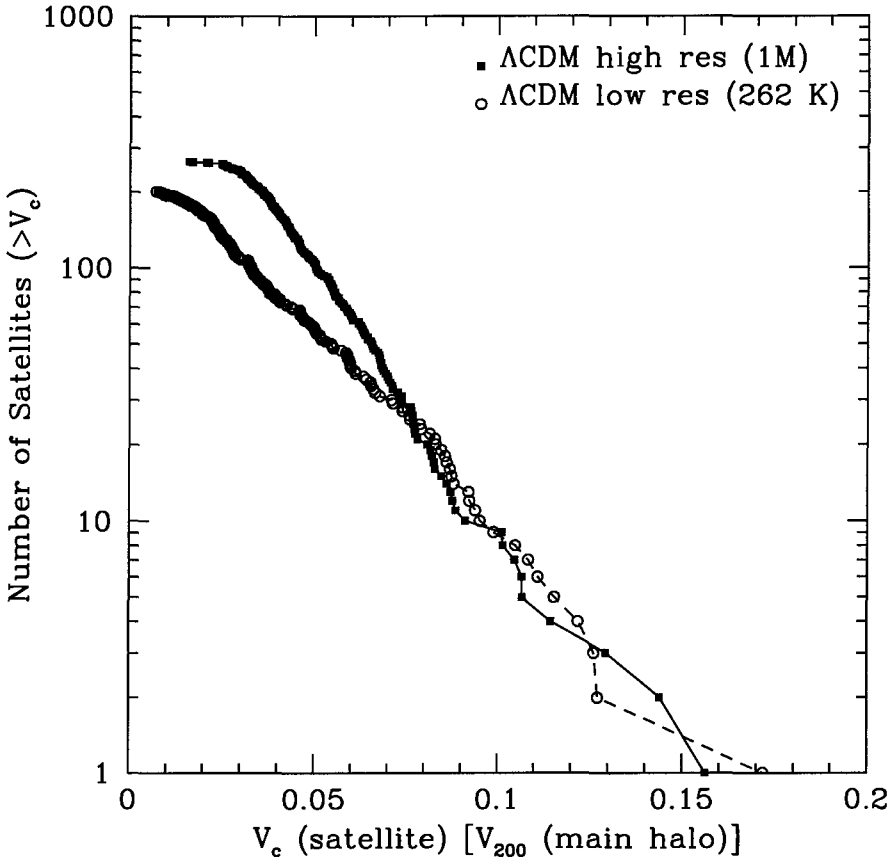

Figure 2.5: Cumulative number of sub-halos within the virial radius of the low and high resolution ΛCDM galactic halos (both SKID runs have $\tau = 4$ kpc, $n_{min} = 8$).

of sub-halos in these systems. For 1 kpc $< \tau <$ 4 kpc the number of sub-halos is relatively independent of τ.

Figure 2.4 shows the circular velocity function, hence CVF, for the substructure within the virial radius of the SCDM ("MWy") parent halo. The CVF represents the cumulative number of substructure halos with normalized circular velocity V_c/V_{200} greater than a certain value. On the horizontal axis, the mass of the satellite increases with its maximum circular velocity, V_c, normalized to the circular velocity of the parent halo measured at the virial radius, V_{200}. The different symbols in Figure 2.4 correspond to several values of τ used in SKID. This plot shows that for a small linking

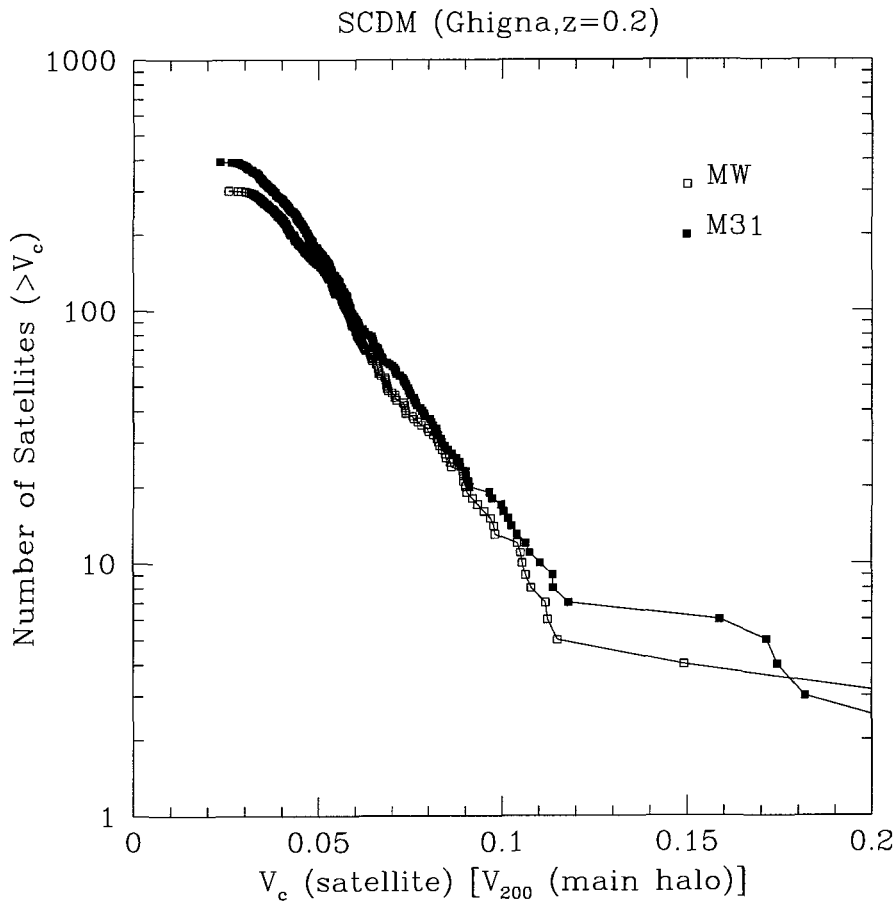

Figure 2.6: Cumulative number of sub-halos within the virial radius of the two parent dark matter halos, similar to Milky Way and M31, found in the SCDM model.

length, such as $\tau = 0.5$ kpc, sub-halos with $V_c/V_{200} > 0.1$ are not properly identified, which (in terms of mass, these correspond to approximately, $M_{sat} > 10^8 - 10^9\, M_\odot$ sub-halos). As the value of the linking length is increased, the algorithm converges, and the substructure mass function becomes relatively independent of τ for $\tau \simeq 1-4$ kpc. Similar tests have been performed for the other halos identified in Table 2.1, and the results are similar to those that presented in Figure 2.4. Also, several tests have been performed by varying the minimum number of particles in a group ($n_{min} = 8, 16, 32$) identified by SKID. We find that the maximum number of sub-halos is obtained with a value of $n_{min} = 16$ for the SCDM run, and $n_{min} = 8$, for the ΛCDM runs.

In addition to testing the input parameters in SKID, it is also useful to investigate how the numerical resolution in the cosmological simulations influences the number of identifiable groups. As discussed by K99, insufficient resolution may artificially erase some of the substructure. Figure 2.5 shows the cumulative number of satellites within the virial radius of the main ΛCDM halo, in the case of low and high resolution simulations (both SKID runs have the input parameters $\tau = 4$ kpc, $n_{min} = 8$). It is clear from this figure that fewer low mass sub-halos are identified in the lower resolution run. Quantitatively, a decrease in the numerical resolution by a factor of ~ 4 in particle number leads to a decrease of a factor of ~ 2 in the number of halos with $V \sim 0.05\, V_{200}$. The two CVFs agree roughly for sub-halos with $V_c > 0.075\, V_{200}$. Below this limit, the velocity function of the low resolution run is not well defined. The typical number of particles for a sub-halo with $V_c \sim 0.075\, V_{200}$ in the low resolution ΛCDM run is about 12, which is very close to the minimum number of particles allowed for this run in SKID ($n_{min} = 8$).

Both simulations agree in the number of high-mass halos, which is expected, since these halos are less affected by numerical resolution. However, the number of large sub-halos ($V_c/V_{200} \simeq 0.15$) seems to be slightly higher, in the case of the SCDM halos (Figure 2.6) than in the ΛCDM halo (Figure 2.5), although this is most likely a result of poor Poisson statistics at the high mass end of the subhalo function.

The final SKID parameters which we choose for identifying substructure are: $\tau = 3$ kpc and $n_{min} = 16$ for the SCDM run, and $\tau = 5$ kpc, $n_{min} = 8$ for the high resolution ΛCDM run. Henceforth, the low resolution ΛCDM run is discarded and we refer to the high resolution ΛCDM run just as "the ΛCDM run". With the above values, SKID yields 685, 1114 and 1091 satellites with masses exceeding $\sim 10^8 M_\odot$ within $2\, R_{vir}$ for the ΛCDM, SCDM ("M31", or "Halo 1") and SCDM ("MWy" or "Halo 2") halos, respectively.

Figures 2.7 and 2.8 show the spatial distribution of the satellites identified in the two galactic halos in the SCDM simulation and in the ΛCDM halos, respectively. The filled squares denote the positions of the centers of mass of the identified sub-

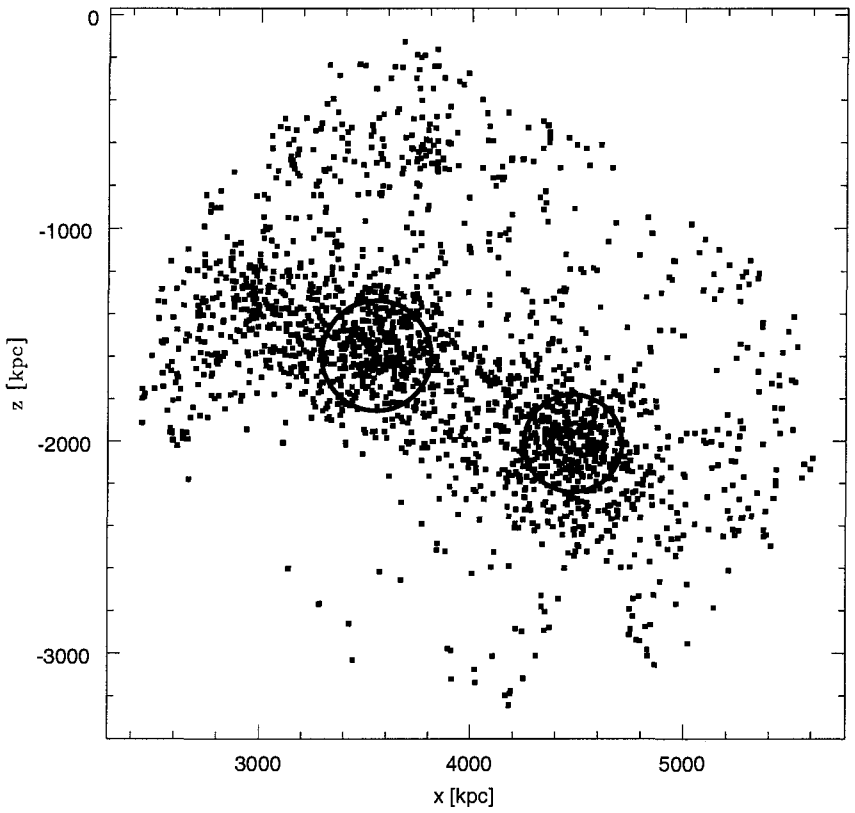

Figure 2.7: Spatial distribution of the sub-halo population identified in the two galactic halos, "MWy" (left) and "M31" (right), in the SCDM run ($z \sim 0$). In this simulation, the MWy halo is located at $x_{MWy} = 3551.78$ kpc, $y_{MWy} = -3059.80$ kpc, $z_{MWy} = -1596.96$ kpc, and the M31 halo is located at $x_{M31} = 4479.25$ kpc, $y_{M31} = -3430.31$ kpc, $z_{M31} = -2011.83$ kpc. The two circles delineate to the virial radii of the two galactic halos.

halos. These plots suggest that there are no strong asymmetries in the distribution of satellites within the respective virial radii. Figure 2.9 shows that dark matter sub-halo distribution over a larger radius, ~ 2 Mpc. An investigation of the orbital properties of the galactic substructure within $2R_{vir}$ is presented in Sections §2.4 and 3.5.

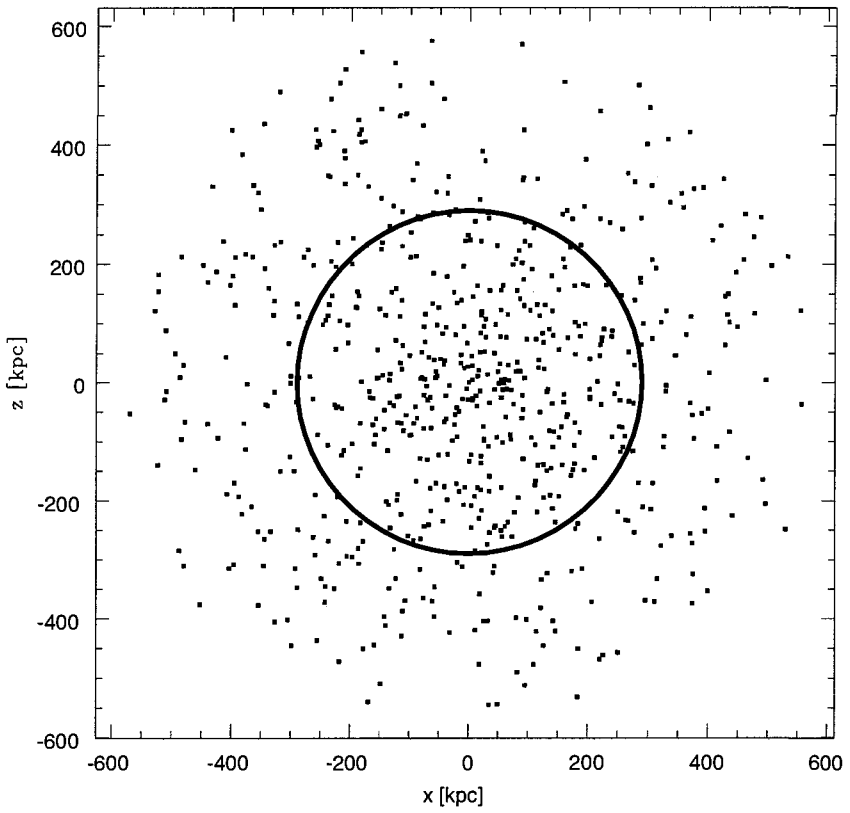

Figure 2.8: Spatial distribution of the sub-halo population within about twice the virial radius ($\sim$ 600 kpc) in the ΛCDM galactic halo. The circle illustrates the virial radius the ΛCDM halo.

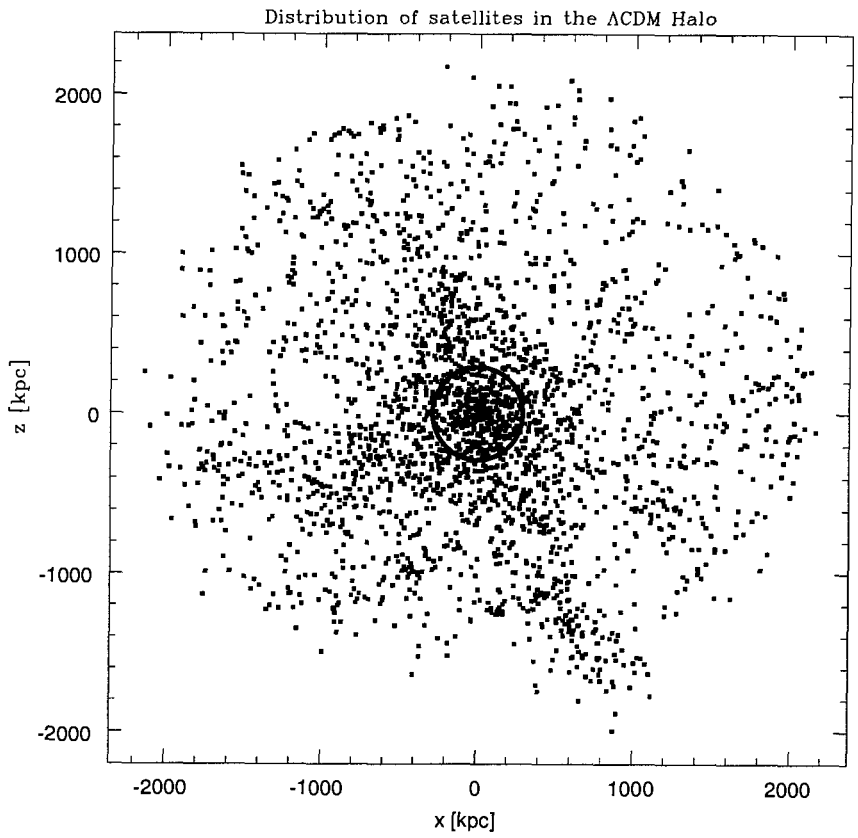

Figure 2.9: Spatial distribution of the sub-halo population within ~ 2 Mpc in the ΛCDM run. The circle corresponds to the virial radius the ΛCDM halo.

2.3.2 Results and Comparison with Previous Studies

Having identified the substructure halos in both SCDM and ΛCDM runs, we proceed to investigate how the current results compare with previous numerical results reported in the literature (K99; M99), as well as with observations (Mateo 1998).

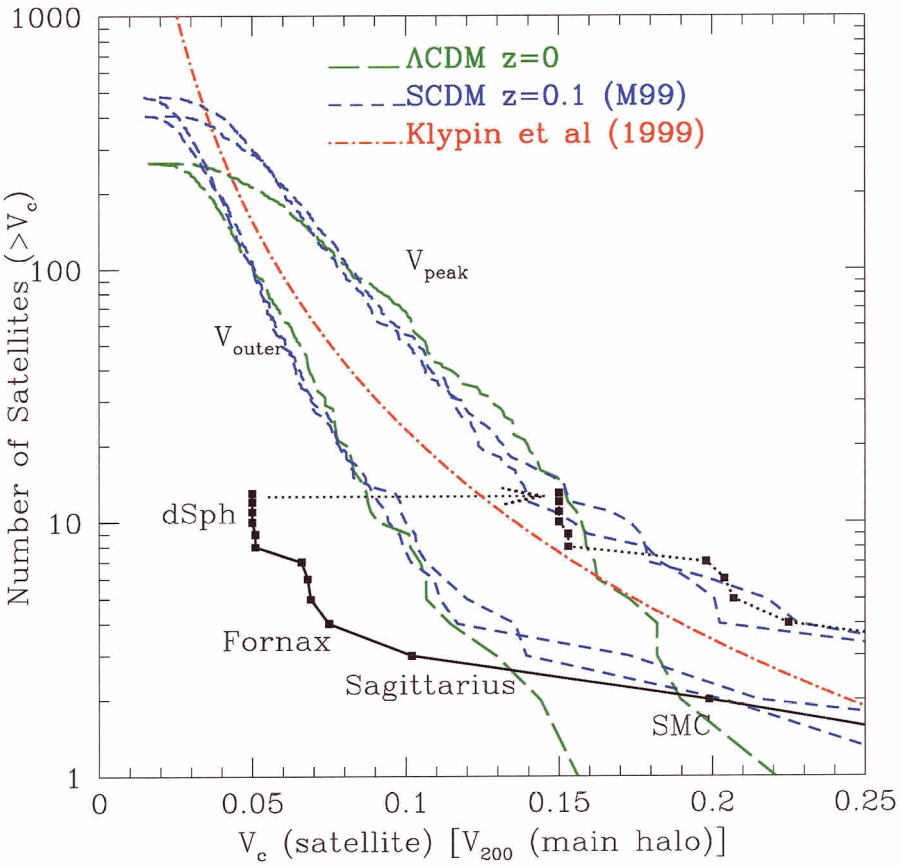

Figure 2.10: Cumulative number of sub-halos in different cosmologies: long-dashed lines represent the ΛCDM cosmology and short-dashed lines the SCDM cosmology. The dot-dashed line shows the analytical fit of Klypin et al. (1999). The arrow shows a correction by a factor of three to reflect the uncertainties in the structure of dark halos surrounding dwarf spheroidals (see text for details).

Figure 2.10 shows the circular velocity function for the SCDM and ΛCDM halos. The two short-dashed lines correspond to the two galaxy-sized halos (denoted as

"M31" and "MWy") in the SCDM model[5] and the long-dashed curves correspond to the high resolution halo formed in the ΛCDM model. Note that for comparison with available observational data we include only sub-halos within the virial radius of the main halo. The two sets of curves, labeled $V_{\rm peak}$ and $V_{\rm outer}$, indicate the result of adopting either the maximum circular speed or the circular velocity at the outermost bound radius of the sub-halo, respectively (these two velocities are returned by SKID). Circular velocity (instead of mass) is typically used to characterize sub-halos because of its weaker dependence on the exact way in which substructure is identified. Still, sub-halo circular velocities do change with radius, and there is no unique way of defining sub-halo circular velocities. $V_{\rm peak}$ and $V_{\rm outer}$ may differ in some cases by up to $\sim 50\%$, implying rather different substructure velocity functions, as can be seen in Figure 2.10. The substructure function presented in M99's Figure 2[6] is consistent with the set of curves labeled $V_{\rm peak}$ in Figure 2.10.

The dot-dashed curve in Figure 2.10 illustrates the power-law fit of slope -2.75 reported by K99, who find that, in a similar ΛCDM simulation ($h = 0.7$), the CVF is well approximated by:

$$n(>V_{circ}) = \begin{cases} 5000(\frac{V_c}{10\,km\,s^{-1}})^{-2.75}\,(h^{-1}Mpc)^{-3}, & \text{if } R < 200h^{-1}\,kpc \\ 1200(\frac{V_c}{10\,km\,s^{-1}})^{-2.75}\,(h^{-1}Mpc)^{-3}, & \text{if } R < 400h^{-1}\,kpc \end{cases} \tag{2.1}$$

K99's results are shown by the dot-dashed line in Figure 2.10 and are in reasonable agreement with the present determination, provided that the circular velocities computed by their substructure-finding algorithm fall, as it seems plausible, between $V_{\rm peak}$ and $V_{\rm outer}$. Our results also agree in the value of the slope with both M99 and K99 studies and it is in rough agreement with the value predicted by the Press-Schechter theory (the CVF scales as $\simeq 3 \times 10^4\,(V_{circ})^{-3.75}\,h^3$ Mpc h^{-3}/km s^{-1}; see K99 for a

[5]We remind the reader that those are the same halos discussed by M99, and therefore can be directly compared with their Figure 2.

[6]Note that the caption to M99's Figure 2 suggests, apparently in error, that $V_{\rm outer}$ is used there rather than $V_{\rm peak}$.

derivation). Also, as pointed out in K99, the scaled CVF is roughly independent of the mass of the parent halo and also rather insensitive to the cosmological model adopted.

How does the subhalo velocity function compare with the observed number of satellites in the vicinity of the Milky Way? The solid squares in Figure 2.10 (joined by a solid line) illustrate the cumulative number of known Milky Way satellites as a function of the circular velocity of their halos, as plotted by M99 with data compiled from Mateo (1998). The discrepancy between the number of dark matter satellites and that of the luminous satellites is evident (up to 500 satellites with circular velocities exceeding $\sim$ 10 km/s may are identified within $\sim$ 300 kpc from the center of the Galaxy, whereas only a dozen of dwarf galaxies are present).

As discussed in Chapter 1, several explanations have been advanced in order to explain this "dwarf satellite problem". Here we mention only one suggestion, which argues that a direct comparison between the peak circular velocity of the dark matter halo and the velocity dispersion of the dwarf galaxy may not be warranted (White 2000). Indeed, in M99 circular velocities are derived for the halos of dwarf spheroidals assuming that stars in these systems are on isotropic orbits in isothermal potentials. This is a plausible, but nevertheless questionable, assumption. Dark halos differ significantly from simple isothermal potentials and numerical simulations indicate that circular velocities decrease substantially near the center. If stars populate the innermost regions of subhalos their velocity dispersions may substantially underestimate the subhalo peak circular velocities. In fact, dwarf spheroidals may plausibly inhabit potential wells with circular velocities up to a factor of 3 times larger than inferred under the isothermal assumption. Following White (2000), the velocity dispersion of the stars, σ_*, relates to the circular velocity of the halos in which they are embedded (V_c), as :

$$\sigma_*{}^2 = \frac{1}{3}\frac{1}{M_\star}\int V_c{}^2(r)\frac{dM_\star}{dr}dr \simeq \frac{1}{3}V_c{}^2(r_h), \tag{2.2}$$

where $V_c^2 = GM(r)/r$, $M_\star(r)$ is the stellar mass profile and r_h is the halo tidal radius.

For dwarf spheroidal galaxies neighboring the Milky Way (eg. Carina, Sculptor, Sextans, Leo II), velocity dispersions and tidal radii are approximately, $\sigma_\star \sim 6.5$ km s^{-1} and $r_h \sim 300$ pc, respectively. Assuming that their halo is characterized by an NFW density profile (Navarro, Frenk & White 1996; 1997, hence NFW), White (2000) derives the maximum circular velocity as being of the order of $V_{peak} \sim 22$ km s^{-1} (see also Stoehr et al. 2002, for a similar derivation). The dotted line in Figure 2.10 corresponds to the observational data, where circular velocities have been increased by a factor of three to reflect the above uncertainty in the structure of dark halos surrounding dwarf spheroidals.

Such a correction may reconcile, at the high mass end, the Milky Way satellite velocity function with the subhalo function, as shown by the dotted line in Figure 2.10. Therefore, it is possible that the number of *massive* satellites expected in the CDM scenario may not be in gross conflict with observation. It should be emphasized, however, that the size of the correction depends sensitively on the inner structure of subhalos, which is poorly resolved even in the best simulations available at present.

2.4 Prograde versus Retrograde Satellites

What about the direction of rotation of the dark matter satellites? We choose to analyze the motion of the satellites within a given radius R with respect to the direction of the total angular momentum of the (halo + satellites) system at that distance,

$$\vec{L}_{tot,\,R} \equiv \vec{L}_{halo,\,R} + \Sigma \vec{L}_{sat,\,R}\,,$$

because any net differential rotation of the satellites will become apparent relative to this axis. In the following, we choose to analyze the rotation of satellites orbiting within the virial radius, $R = R_{vir}$, and within twice the virial radius, $R = 2R_{vir}$,

respectively. In each case, we determine the net rotation of satellites for the three halo distributions, "Halo 1" and "Halo 2" (SCDM) and the ΛCDM halo, respectively.

Let the new axis defined by the total angular momentum $\vec{L}_{tot,\,R}$ be denoted by the generic parameter z'. Note that the direction of the z' axis will be different in the two $R = \{R_{vir},\, 2R_{rvir}\}$ cases, as well as for each satellite population in the three halos considered in each case. However, as a common characteristic of all cases, we mention that the angular momentum of the halo dominates that of the satellites by a few orders of magnitude. Specifically, for $R = R_{vir}$, $|\vec{L}_{halo,\,R}| = 6.3 \times 10^{19}$ $M_\odot$ kpc km $\cdot$ s^{-1}, whereas $|\Sigma\vec{L}_{sat,\,R}|$ equals 5.5×10^{15} $M_\odot$ kpc km $\cdot$ s^{-1} for "Halo 1" (SCDM), 2.3×10^{15} $M_\odot$ kpc km $\cdot$ s^{-1} for "Halo 2" (SCDM) and 7.2×10^{13} $M_\odot$ kpc km $\cdot$ s^{-1} for the ΛCDM halo, respectively. Similarly, for the $R = 2R_{vir}$ case, $|\vec{L}_{halo,\,R}| = 8.24 \times 10^{19}$ $M_\odot$ kpc km $\cdot$ s^{-1}, whereas $|\Sigma\vec{L}_{sat,\,R}|$ equals 1.73×10^{16} $M_\odot$ kpc km $\cdot$ s^{-1} for "Halo 1" (SCDM), 8.48×10^{15} $M_\odot$ kpc km $\cdot$ s^{-1} for "Halo 2" (SCDM) and 8.42×10^{14} $M_\odot$ kpc km $\cdot$ s^{-1} for the ΛCDM halo. This result implies that, for a given radius R, the differences between the z' axes corresponding to the three halo cases are negligible.

Also, let x' and y' be the other two axes contained in the plane perpendicular to the vector $\vec{L}_{tot,\,R}$. Together, they form the new system of coordinates (x', y', z') in which we calculate the new angular momenta of satellites. In the following, we concentrate only on the z' components of the satellites, $L_{z',\,sat}$, and, in particular, to their corresponding specific angular momenta, $l_{z'}$.

Figure 2.11 shows the specific angular momenta ($l_{z'}$) of the satellites orbiting within approximately twice the virial radius ($\sim$ 600 kpc). The top panel corresponds to the ΛCDM halo (full squares), the middle panel to the "Halo 1" (SCDM run) (filled triangles) and the bottom panel to the "Halo 2" (SCDM run) (empty hexagons), respectively. From the even distributions of positive and negative orientations of $\vec{l_{z'}}$, we may infer that in a typical CDM halo there are, on average, as many satellites on prograde orbits as there are on retrograde ones. The specific total angular momenta of the satellite populations in the three simulations, $| < l_{tot,sat} > | \equiv \frac{|\Sigma\vec{L}_{sat}|}{\Sigma M_{sat}}$, have absolute magnitudes much smaller than the specific angular momenta of individual

satellites. For example, $| < l_{tot,sat} > | \simeq 1.3 \times 10^5$ kpc km s^{-1} for the "Halo 1" (SCDM run), 1×10^5 kpc km s^{-1} for the "Halo 2" (SCDM run) and 5.8×10^3 kpc km s^{-1} for the ΛCDM run. These values are consistent with only a a small excess of net angular momentum, given that the range of angular momenta of satellites is in all three cases -10^5 kpc km s^{-1} $\leq l_{tot,sat} \leq 10^5$ kpc km s^{-1}.

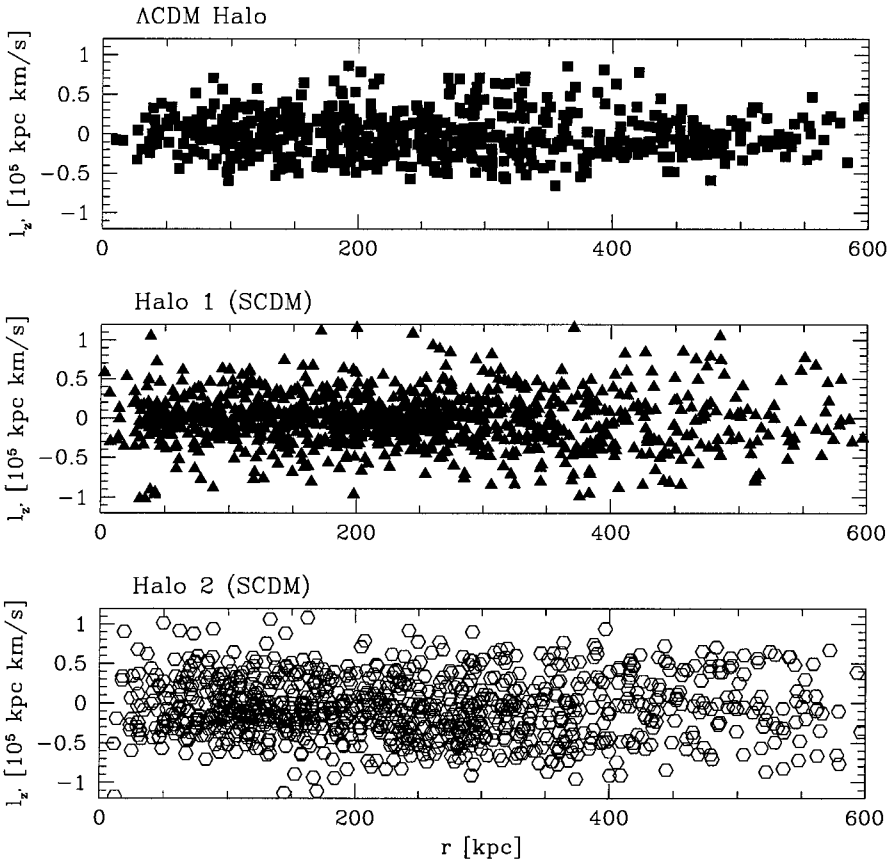

Figure 2.11: Specific angular momenta on the z' -axis for the satellites selected within $2R_{vir}$ of the primary halos, in SCDM and ΛCDM simulations. The top panel corresponds to the ΛCDM halo (full squares), the middle panel to the "Halo 1" (SCDM run) (filled triangles) and the bottom panel to the "Halo 2" (SCDM run) (empty hexagons), respectively.

Figure 2.12 investigates this further by showing the distribution of satellite rotational velocities relative to the z' axis within twice the virial radius of a galactic halo

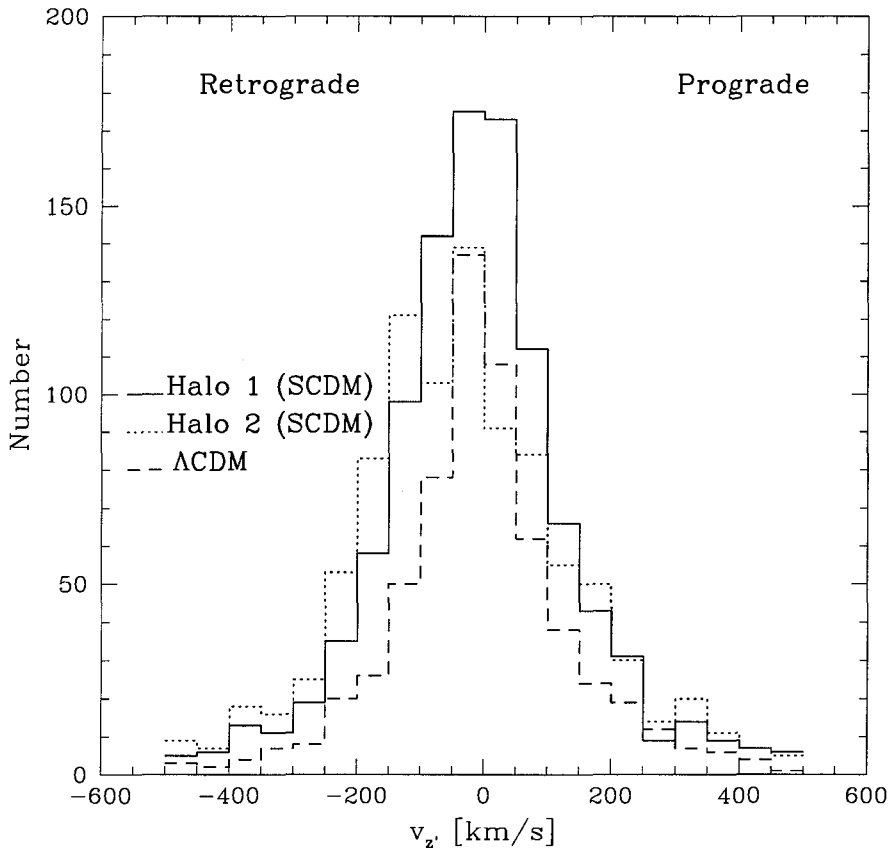

Figure 2.12: Distribution of satellite rotational velocities relative to the disk within twice the virial radius of a galactic halo in the SCDM and ΛCDM runs. On average, there are as many satellites on prograde as on retrograde orbits.

in the SCDM and ΛCDM runs. Here

$$v_{z'} \equiv v_{rot} = (x'v_{y'} - y'v_{x'})/r' = l_{z'}/r',$$

where $r' = \sqrt{x'^2 + y'^2}$. A satellite with a positive rotational velocity $v_{z'}$ (hence $l_{z'}$) component is moving on a prograde orbit and one with a negative $v_{z'}$ is moving on a retrograde one. This figure shows that, on average, there are as many satellites on prograde as on retrograde orbits within $2R_{vir}$.

A similar analysis has been done for the $R = R_{vir}$ case. Figure 2.13 shows the

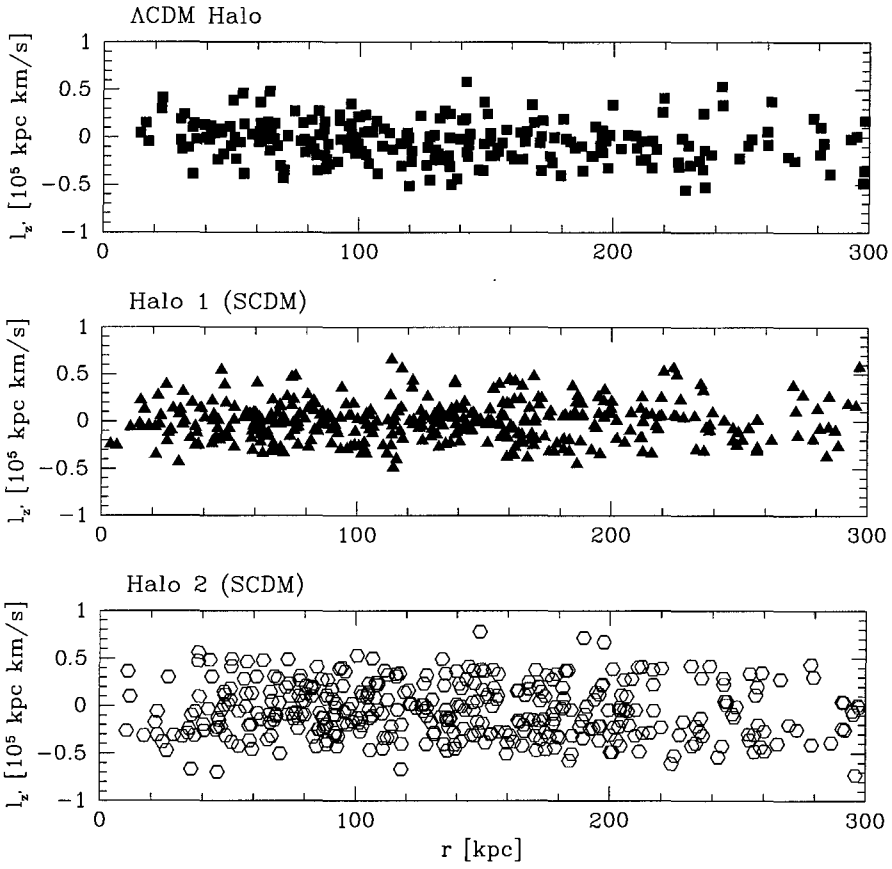

Figure 2.13: Specific angular momenta on the z' -axis for the satellites selected within R_{vir} of the primary halos, in SCDM and ΛCDM simulations. The top panel corresponds to the ΛCDM halo (full squares), the middle panel to the "Halo 1" (SCDM run) (filled triangles) and the bottom panel to the "Halo 2" (SCDM run) (empty hexagons), respectively.

specific angular momenta ($l_{z'}$) of the satellites orbiting within $\sim$ 300 kpc. The panels and symbols are the same as in Figure 2.11. Again, the satellite specific angular momenta appear to be evenly distributed between prograde and retrograde orbits. Top panel in Figure 2.14 shows the distribution of rotational velocities of the galactic halos in the SCDM and ΛCDM runs (again, included within the virial radius). This figure, however, reveals small differences between the three halo distributions, "Halo 1" displaying a small net distribution of prograde orbits, "Halo 2" having a

slighly net distribution of retrograde orbits, whereas the ΛCDM halo seeming to be having roughly an even distribution between prograde and retrograde orbits. Overall, our data sets do not show any clear tendency for either a prograde or retrograde distribution of satellite orbits.

In this latter case (i.e. $R = R_{vir}$), it is instructive to compare the distribution of rotation velocities of satellites with respect to the z' axis and compare it with the distribution of observed rotation velocities of satellites around the rotation axis of the Milky Way disk. The comparison between our data sets and observations is possible under the assumption that the rotation axis of the disk[7] is roughly aligned with that of the z' axis. This is a reasonable assumption since, to reach equilibrium, the rotation axis of the disk will tend to align with one of the principal axes of the halo. In the particular case of our simulations this means that it is also fair to assume that the rotation axis of the disk is aligned with z' (i.e. with $\vec{L}_{tot,\,R}$), because the angular momemntum of the halo dominates $\vec{L}_{tot,\,R}$.

The bottom panel of Figure 2.14 shows, for comparison, similar data compiled from observational studies. The solid line in the bottom panel shows the distribution of satellite radial velocities relative to the disk, as reported by Zaritsky et al. (1997), whereas the dashed line corresponds to the more recent study of Azzaro et al. (2003). Both of these studies use a "statistical" approach, forced by the practical limitation of detecting only a few satellites around any given system. The method used scales all observed satellites to a "fictitious" primary. The sample of Zaritsky et al. (1997) comprises 57 primary galaxies and 95 satellites. Azzaro et al. (2003) include a sample of 141 large spiral galaxies and 200 satellites, selected from the Sloan Digital Sky Survey data release. Figure 2.14 shows that there is no significant nor a consistent preference for either prograde or retrograde orbits.

Overall, we conclude that our numerical result agrees well with the observed properties of satellites by Zaritsky et al. (1997), in the sense that there is no preference for

[7]Note that the rotation axis of the disk is roughly the same as the axis of the total angular momentum of the disk.

either prograde or retrograde orbits. On the other hand, Azzaro et al. (2003) claim that their data show a small excess ($\sim$ 56%) of prograde satellites. However, one should also bear in mind that the observational studies can be affected by selection effects, such as interlopers or in completeness, and therefore, a few percent difference may turn out not to be significant. A more detailed study of the observational effects, as well as a statistical sample of simulated dark matter halos, are needed to clarify these issues.

As a final note, we mention that the distribution of prograde versus retrograde satellites may influence the heating of galactic disks. As shown in previous numerical experiments (eg., VW99), the sense of rotation of the galactic satellites around the spin axis of the stellar disk is an important factor in the evolution of the disk: prograde satellites are more efficient in heating the disk, whereas retrograde satellites have the main effect of tilting the disk. A significant number of either prograde or retrograde orbits, if they exist, should therefore be taken into account in analytical modeling of the heating of galactic disks.

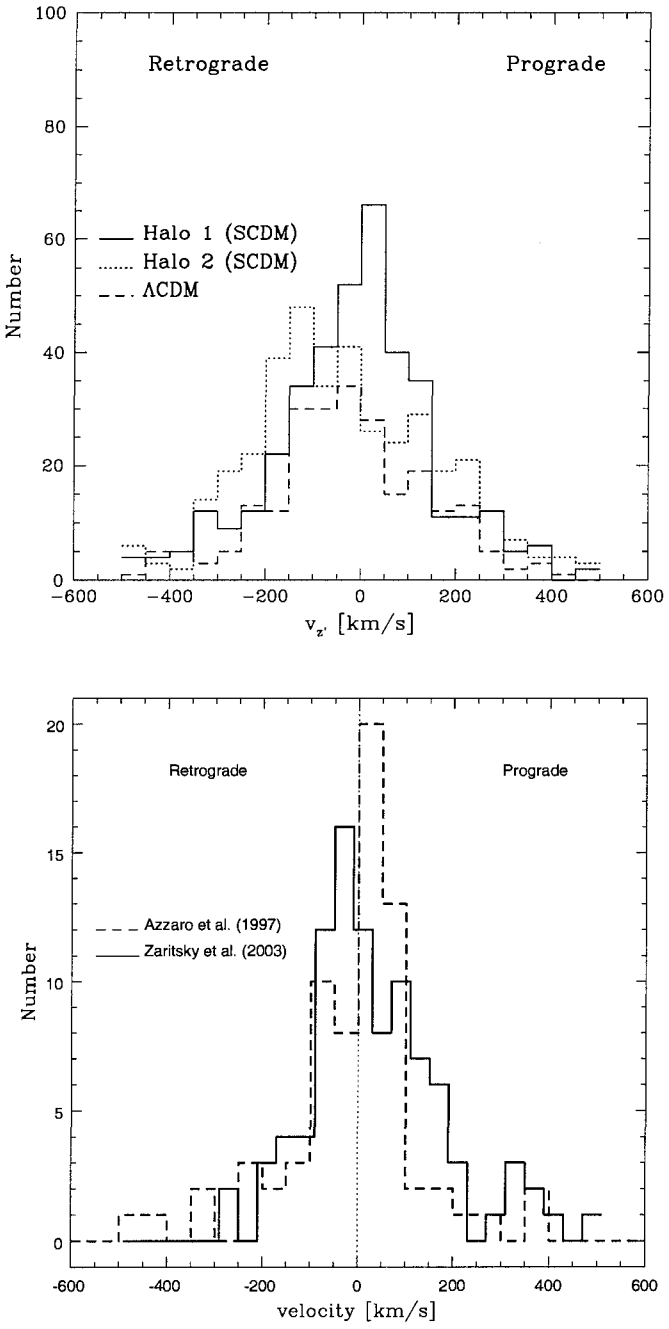

Figure 2.14: Top panel: Distribution of satellite radial velocities relative to the z' axis (assumed to be roughly the same as the rotation axis of the disk) within the virial radius of a galactic halo in the SCDM and ΛCDM runs. On average, there are as many satellites on prograde as on retrograde orbits. Bottom panel: A similar result is obtained in the observational study of Zaritsky et al (1997), whereas Azzaro et al. (2003) find a small excess of prograde satellites.

Summary:

This chapter describes of the process of identification of dark matter substructure in galactic halos. Several tests have been performed in order to study how substructure depends on the numerical resolution of the cosmological simulations; the linking length or number of neighbors in the group-finding algorithm; and the cosmological model. This study re-affirms the discrepancy between the abundance of dark matter satellites and that of corresponding dwarf satellite galaxies in the Milky Way. Several arguments are also discussed, which may help alleviate the problem.

We also investigate the orbital parameters of the dark matter satellites and argue that the realizations of the satellite population presented in this thesis are representative for the CDM models. This suggests that the effects these satellites may have on a stellar disk may be typical in the context of the CDM paradigm as a whole. Finally, we show that the population of dark matter satellites is evenly distributed between prograde and retrograde orbits, in rough agreement with observational studies.

Chapter 3

Multicomponent Galaxy Models Including Substructure

Abstract

This chapter describes the building of a self-consistent equilibrium galaxy model of the Milky Way, consisting of a disk, bulge and halo components, including its dark matter substructure. First, we perform a series of numerical simulations in order to investigate the effects of particle resolution on the secular evolution of isolated[1] galaxy models. We show that departures from the initial equilibrium state of an isolated galaxy model, as measured by the secular increase of the disk thickness or of the Toomre parameter, scale as the product between the number of particles in the disk (N_d) and the number of particles in the halo (N_h), suggesting that the evolution is driven mainly by collisions between disk and halo particles. We find that a stable model of Milky Way galaxy requires $N_d N_h$ of order $\sim 10^{10}$. We conclude that simulations designed to test whether disks as cold and thin as the Milky Way can survive within the lumpy potential well of a Cold Dark Matter halo should be within present computational resources and we present the set-up for a set of these simulations.

[1]Henceforth, the term "isolated" will be used to describe galaxy models which do not include halo substructure (being, therefore, composed of only the disk, bulge and smooth halo)

3.1 Motivation

As mentioned in Chapter 1, one of the shortcomings of present cosmological simulations of galaxy formation is that they are unable to assemble disks of sizes comparable to those observed in the local Universe. The cause of this problem is the severe angular momentum loss of the accreting gas, which translates in to a reduced scale length of the rotationally supported disk –also known as "the angular momentum catastrophe" (Navarro & White 1994). Several explanations for this problem have been advanced. For example, low numerical resolution may result in an unreliable treatment of the gas physics, such as an overestimation of the effects of shocks or of viscosity (Springel & Hernquist 2002). This will cause excessive gas cooling in the simulation and, indirectly, a loss of angular momentum through excessive "clumping" of gas at the center of progenitor dark halos. In addition, present codes do not incorporate accurate recipes for either star formation, or supernovae feedback, and include the UV external background produced by quasars and massive stars only in approximative ways. Increasing the numerical resolution of the simulations, as well as including the effects of feedback suggests that the angular momentum problem is about to be solved (eg., Thacker & Couchman 2000).

Until these complications are fully resolved, one needs to resort to building realistic galaxy disks by other means, rather than forming them *in situ*, in order to study the evolution of stellar disks. This is particularly so for the problem investigated in this thesis, where one needs to construct a spiral disk similar to the present-day Mily Way disk, where observational constraints are strongest. The procedure we adopt here (as outlined in the following sections), consists of building a spiral galaxy in isolation and adding to it the substructure extracted from a dark matter halo formed in realistic cosmological simulations.

3.2 The Hernquist Method

An isolated N-body model of the Milky Way may be constructed following the procedure outlined by Hernquist (1993). This is an algorithm which builds equilibrium models of galaxies by modeling self-consistently all three galactic components (disk, bulge and dark halo).

The major difficulty encountered in the construction of galaxy models is that we do not know the distribution function $f(\mathbf{x}, \mathbf{v}, t)$ of the matter (either baryonic or dark) in the galaxy. If the distribution function were known, the coordinates $(\mathbf{x}, \mathbf{v})$ of a set of particles may be determined in a straightforward way by sampling $f(\mathbf{x}, \mathbf{v}, t)$ in a Monte Carlo fashion. Hernquist (1993) solves this problem by reconstructing the positions of the particles from analytic density distributions for each galactic component and velocities from the moments of the collisionless Boltzmann equation (CBE): $df/dt = 0$. This, of course, is not an exact solution to the problem, since the velocities derived from the moments of the CBE are close, but not identical with the ones derived from a proper $f(\mathbf{x}, \mathbf{v}, t)$. Nevertheless, the method is a reliable way of constructing galaxy models, which are in near-equilibrium and stable for a long period of time.

The density distribution of each galactic component is chosen to fit observations. The disk is modeled with an exponential density profile:

$$\rho_d(R, z) = \frac{M_d}{4\pi {R_d}^2 z_0} e^{(-R/R_d)} \, sech^2(z/z_0), \tag{3.1}$$

where M_d is the mass of the disk, and R_d and z_0 are the radial and vertical scale lengths of the disk. The bulge is assumed to follow the density law (the "Hernquist profile"):

$$\rho_b(r) = \frac{M_b}{2\pi} \frac{a}{r(a+r)^3}, \tag{3.2}$$

where M_b is the bulge mass and a is the scale length of the bulge. Finally, the dark matter halo is approximated by the distribution of an isothermal sphere, truncated

at a certain radius:

$$\rho_h(r) = \frac{M_h\,\alpha_h}{2\pi^{3/2}\,r_{cut}}\,\frac{e^{(-r^2/r_{cut}^2)}}{r^2+\gamma^2}, \tag{3.3}$$

where M_h is the halo mass, γ and r_{cut} are the core and cut-off radii of the halo, and α_h is a normalization constant:

$$\alpha_h = \frac{1}{1-\sqrt{\pi}q\exp(q^2)[1-\mathrm{erf}(q)]}, \tag{3.4}$$

where $q = \gamma/r_{cut}$.

As mentioned before, the velocity distribution cannot be directly computed. However, since the density distributions represent the lowest order moment of the CBE, the velocity distribution can be approximated from the second moments of the CBE. A complete description and implementation of the algorithm is provided by Hernquist (1993).

A final important step in the algorithm consists in constraining the velocity ellipsoid of the stars in the disk, in order to obtain a thin and stable disk. A physical parameter that acts as a basic thermometer for the disk is the Toomre parameter (Toomre 1963):

$$Q = \frac{\sigma_r\kappa}{3.36G\Sigma_r}, \tag{3.5}$$

where σ_r is the radial velocity dispersion of the disk stars, Σ_r is the radial surface density, and κ is the epicyclic frequency, $\kappa = \frac{3}{R}\frac{\partial\Phi}{\partial R} + \frac{\partial\Phi^2}{\partial R^2}$. Cold disks have low velocity dispersion and therefore a low Toomre parameter, $Q \simeq 0$, whereas hot disks have high dispersion and high Q. A value of $Q > 1$ is necessary for the stellar disk to be stable against axi-symmetric perturbations (see BT87, §6.3 and references therein). The Hernquist (1993) method fixes the value of the Toomre parameter to 1.5 at a certain location which, for convenience, is chosen to be at the solar radius ($\sim$ 8.5 kpc from the Galactic center).

3.3 Noise - Induced Evolution in Isolated Galaxy Models

Present-day numerical simulations of disk galaxies are still affected by limited computational power. This is particularly true for fully self-consistent particle methods, such as tree codes, that do not impose symmetry constraints on the problem nor make use of expansion methods to speed up the computation of the gravitational potential. Computational power limits the number of particles that may be used, compromising the reliability of the integration of collisionless systems over many dynamical times. Although particle-particle interactions can be reduced by "softening" the gravitational potential, still two-body collisions are the main factor limiting the applicability of N-body methods to collisionless systems. These problems are exacerbated for cold, centrifugally-supported, thin stellar disks. Collisions between disk particles, as well as between disk particles and those used to model the dark halo where the disk is embedded, lead disk stars to deviate systematically from their nearly circular orbits and to increase the amplitude of their excursions outside the plane. The effect is to "heat" the disk away from its initial state, transforming it progressively into an ever thicker component where rotation plays a monotonically decreasing role in supporting the disk.

This section presents a study of the secular evolution of isolated galaxy models constructed with the Hernquist method. Our aim is to identify the optimal conditions for minimizing the effect of random 2-body interactions between disk - disk particles and disk - halo particles. The assessment of the role of this numerical noise is an important first step in the study of disk heating because it allows us to distinguish between heating caused by physical processes and spurious heating introduced by limited numerical resolution.

The mass resolution (measured by the mass per particle, $m \equiv M/N$) is an important factor in the outcome of particle-particle collisions. Most numerical simulations of galaxies which employ a "live" dark matter halo adopt halo particles that are sev-

eral times more massive than disk particles, $m_h > m_d$ (for example, VW99 use a ratio $m_h/m_d = 3$). This is chosen so as to save computational time by reducing the number of halo particles in the system (the dark halo is usually much more massive than the disk). However, this results in increased heating of the disk and, therefore, in a spurious increase in the disk thickness. It is also important to have a large number of particles in the disk so as to diminish the scattering of disk particles. Therefore, increasing either N_h or N_d will have a general effect of decreasing the noise in the disk. We explore below a suite of numerical experiments designed to establish empirically the dependence of numerical disk heating on N_h and N_d.

3.3.1 The Test Grid

Table 3.1 presents the various steps we have explored in order to survey the parameter space (N_d, N_h, N_b). In all runs the halo, disk and bulge masses are kept constant and the bulge particle mass is chosen to be the same as the disk's. Thus the only free parameters may be taken to be N_d and m_h/m_d.

As shown in Table 3.1, N_d and N_h have been varied so that $N_d \simeq 5,000$, $10,000$, $20,000$, $30,000$, $40,000$, and $50,000$ and the ratio of halo to disk particle mass surveyed is $m_h/m_d = 1$, 2, 3. The code units are chosen so that $M_d = 1$, $R_d = 1$, $G = 1$. Scaled to the Milky Way, with $M_d = 5.6 \times 10^{10} M_\odot$ and $R_d = 3.5$ kpc, the code units for time and velocity are 1.3×10^7 years and 262 km s^{-1}, respectively (an explanation for the choice of the physical parameters for the Galactic components is given in Section §3.4). All models have a halo cut-off radius of $r_{cut} = 24$ [code u.] = 84 kpc, a core radius $\gamma = 1$ [code u.] = 3.5 kpc, a bulge scale length $a = 0.15$ [code u.] = 525 pc and spline softening lengths for disk, bulge and halo of $\epsilon_D = \epsilon_B = \epsilon_H = 0.05$ [code u.] = 175 pc (see details in Table 3.3). Also, the models have fixed bulge/disk and halo/disk mass ratios, $M_b/M_d = 1/3$, $M_h/M_d = 14$, as well as fixed mass per particle in the bulge as in the disk, $m_b = M_b/N_b = m_d = M_d/N_d$. Masses in Table 3.1 are given in code units and $N_{tot} = N_d + N_b + N_h$.

Table 3.1: Parameters of the galaxy models with small halos (r_{cut} = 84 kpc)

Model	N_d	N_b	N_h	$\frac{m_b}{m_d}$	$m_d = m_b$ ($\times 10^{-4}$)	m_h ($\times 10^{-4}$)	$\frac{m_h}{m_d}$	N_{tot}
A	4,998	1,666	69,972	1	2	2	1	76,636
B	4,998	1,666	34,986	1	2	4	2	41,650
C	4,998	1,666	23,324	1	2	6	3	29,988
D	9,999	3,333	139,986	1	1	1	1	153,318
E	9,999	3,333	69,993	1	1	2	2	83,325
F	9,999	3,333	46,662	1	1	3	3	59,994
G	19,998	6,666	279,972	1	0.5	0.5	1	306,636
H	19,998	6,666	139,986	1	0.5	1	2	166,650
I	19,998	6,666	93,324	1	0.5	1.5	3	119,988
J	30,000	10,000	420,000	1	0.33	0.33	1	460,000
K	30,000	10,000	210,000	1	0.33	0.66	2	250,000
L	30,000	10,000	140,000	1	0.33	1	3	180,000
M	39,998	13,333	559,986	1	0.25	0.25	1	613,318
N	39,999	13,333	279,993	1	0.25	0.5	2	333,325
O	39,999	13,333	186,662	1	0.25	0.75	3	239,994
R	50,001	16,667	350,007	1	0.2	0.4	2	416,675
S	50,001	16,667	233,338	1	0.2	0.6	3	300,006

Table 3.2: Parameters of the galaxy models with large halos (r_{cut} = 300 kpc)

Model	N_d	N_b	N_h	$\frac{m_b}{m_d}$	$m_d = m_b$ ($\times 10^{-4}$)	m_h ($\times 10^{-4}$)	$\frac{m_h}{m_d}$	N_{tot}
M1	40,000	13,333	2,207,200	1	0.25	0.25	1	2,260,533
S1	50,001	16,667	920,000	1	0.2	0.6	3	986,668

In addition, we extended models M and S to the case where the radial extent of the halo (controlled by r_{cut} in equation (3.3)) is increased to about 300 kpc. These halos enclose the same mass within 84 kpc as the "small halo" models listed in Table 3.1. These new models are labeled M1 and S1 and their characteristic parameters are listed in Tables 3.2 and 3.3. The motivation for increasing the halo size is twofold: firstly, a cut-off halo radius of $\sim$ 300 kpc matches better the halo sizes expected in CDM models (Navarro, Frenk & White 1996; see also §3.4.1); secondly, by increasing the number of halo particles well beyond the location of disk, it becomes possible to

Table 3.3: Parameters of the test grid galaxy models (in code units)

	(Small halos)	(Large halos)
	Disk:	
M_D	1	1
R_D	1	1
z_0	0.2	0.2
Q_0	1.5	1.5
$R_\odot$	2.43	2.43
ϵ_D	0.05	0.05
	Bulge:	
M_B	0.333	0.333
a_B	0.15	0.15
ϵ_B	0.05	0.05
	Halo:	
M_H	14	55.18
γ	1	1
r_{cut}	24	86
ϵ_H	0.05	0.05

study whether heating of the disk is a local effect created by halo particles close to the disk or, on the contrary, if it is a global one.

All test grid models were evolved in isolation for about 2.6 Gyr (equivalent to $\sim$ 10 full rotations at the solar circle in the disk), using PKDGRAV.

3.3.2 Measurements of Disk Heating

We use two physical parameters to characterize the global heating of the disk: *i)* z_{median}, the median of the absolute vertical distances $|z|$ of all disk particles as a measure of vertical heating; and *ii)* the Toomre parameter evaluated at the solar circle (i.e., at $R_\odot = 8.5$ kpc from the center of the Galaxy), $Q_\odot$, which monitors directly the radial heating of disk particle orbits.

Figures 3.1 and 3.2 show how z_{median} and $Q_\odot$ evolve with time in models A-S. All

models display a monotonic increase of both z_{median} and $Q_{\odot}$ with time; however, the magnitude of the effect generally decreases with increasing number of disk particles.

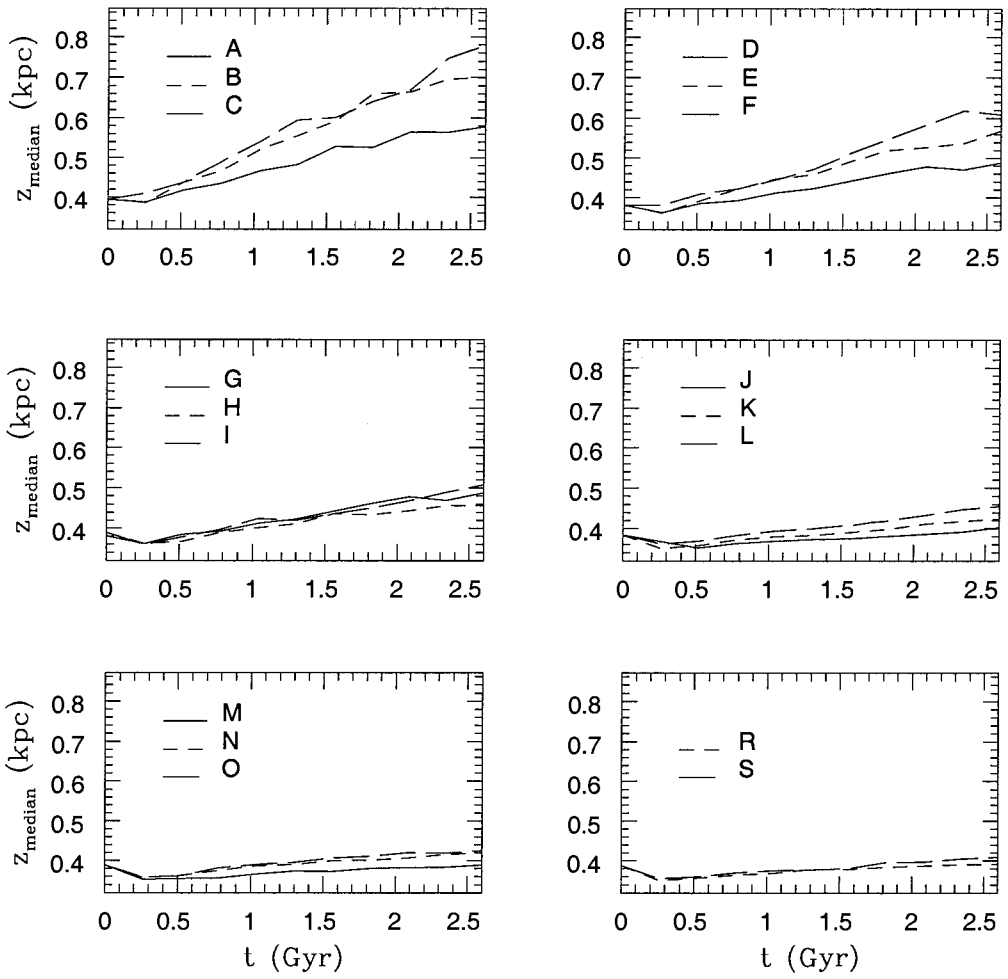

Figure 3.1: Evolution of z_{median} for models $A - S$.

Figure 3.3 shows a snapshot of the mean thickness of the disk, z_{median}, measured for models $A - S$ after 2.6 Gyrs of evolution. Note that ratios $m_h/m_d = 1, 2, 3$ correspond to values of $N_h \sim 14$, 7 and 4.67 times larger than N_d, respectively. Solid lines correspond to models with the same number of disk particles. This figure clearly shows that, as the number of particles in the disk or in the halo increases, the disk

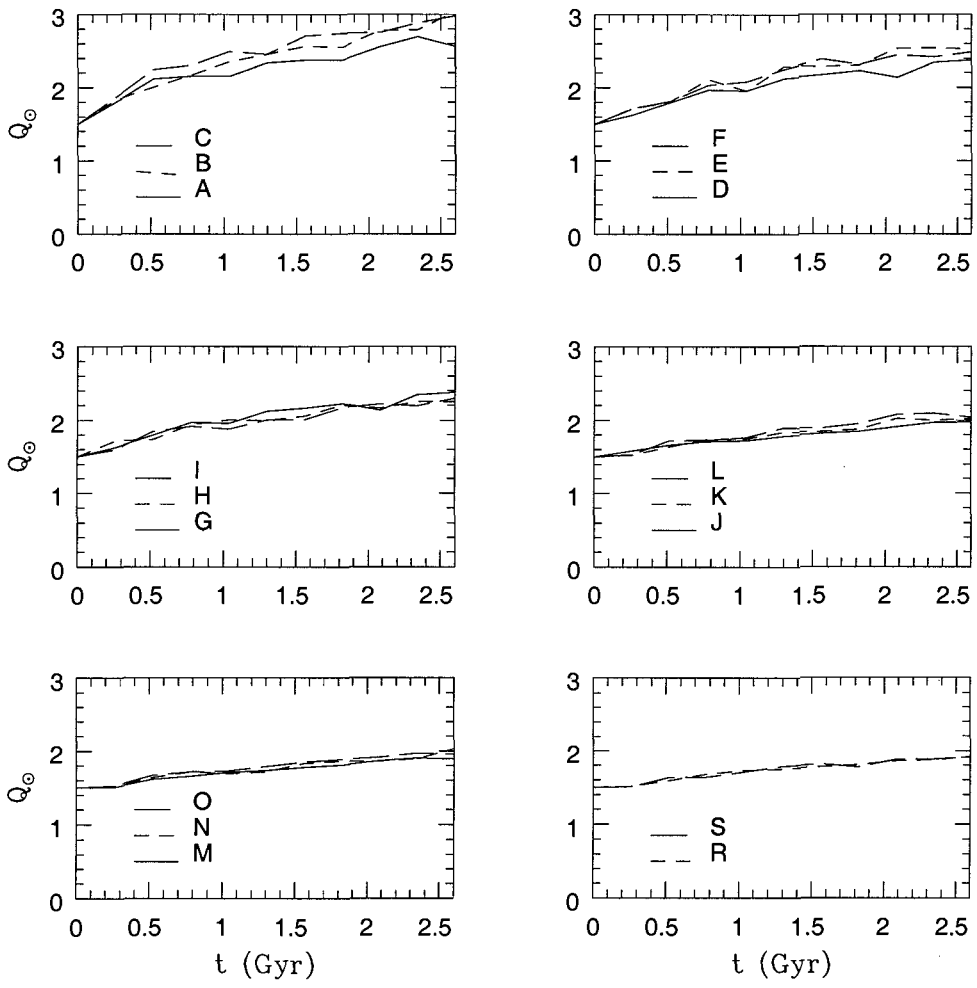

Figure 3.2: Evolution of $Q_\odot$ for models $A - S$.

thickens less (at $t = 0$ the average z_{median} for all models is 0.11 in code units = 0.385 kpc). This result is not unexpected, since the potential becomes smoother as the number of particles increase, leading to smaller deviations from the initial equilibrium state. Figure 3.4 shows that a similar conclusion also applies to the evolution of $Q_\odot \equiv Q(R_\odot)$, shown here also after $t = 2.6$ Gyr of evolution.

It is clear from these plots, however, that the total number of particles is not the main parameter determining the effects of disk heating. For example, models M and

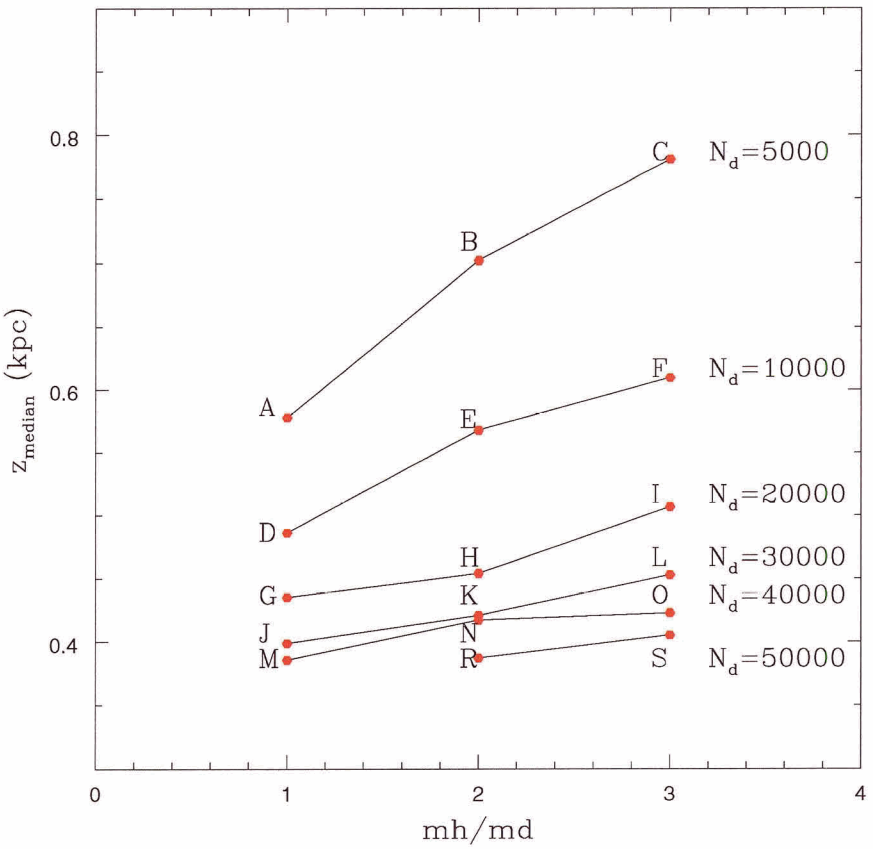

Figure 3.3: z_{median} after 2.6 Gyr versus m_h/m_d for models A – S.

R have the same final z_{median} although model M has 50% more particles. A similar comment applies to model pairs J and S, or G and L. Also, it appears that the number of particles in the halo alone is an important factor in the final z_{median} and $Q_{\odot}$, particularly for small number of disk particles. That is, increasing N_{halo} decreases both z_{median} and $Q_{\odot}$. However, once N_{halo} is relatively large, the differences between the effects of different halos become negligible and the main factor in determining the heating of the disk is N_{disk} (see, for example, the pair of models D and H, or the pair K and G). A similar behaviour seems to occur for N_{disk}, although given the small range of N_{disk} investigated in our study, we cannot draw a definitive conclusion in this

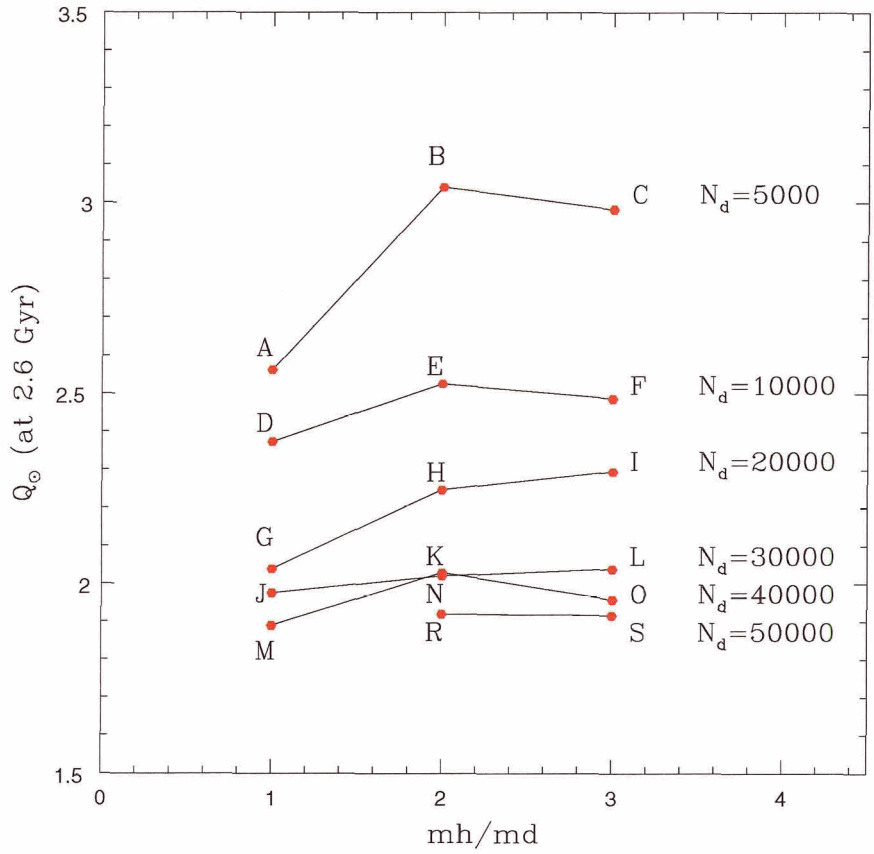

Figure 3.4: $Q_\odot$ after 2.6 Gyr. Models A $-$ S. $Q_\odot$ at $t = 0$ is 1.5

regard. Figures 3.5 illustrates the effects of the disk and halo particles separately, by tracking z_{median} at 2.6 Gyr function of N_{disk} (top panel) and of N_{halo} (bottom panel), respectively. A similar behavior for $Q_\odot$ versus N_{disk} and N_{halo} is illustrated in Figure 3.6 (top and bottom panels, respectively).

What combination of N_h and N_d determines the final disk thickness? This is investigated in Figure 3.7, which shows the root-mean-square (rms) deviation for models $A - S$, from simple power-law fits,

$$z_{median}\,(t = 2.6\,\mathrm{Gyr}) \propto N_h^{\alpha} N_d^{\beta}, \tag{3.6}$$

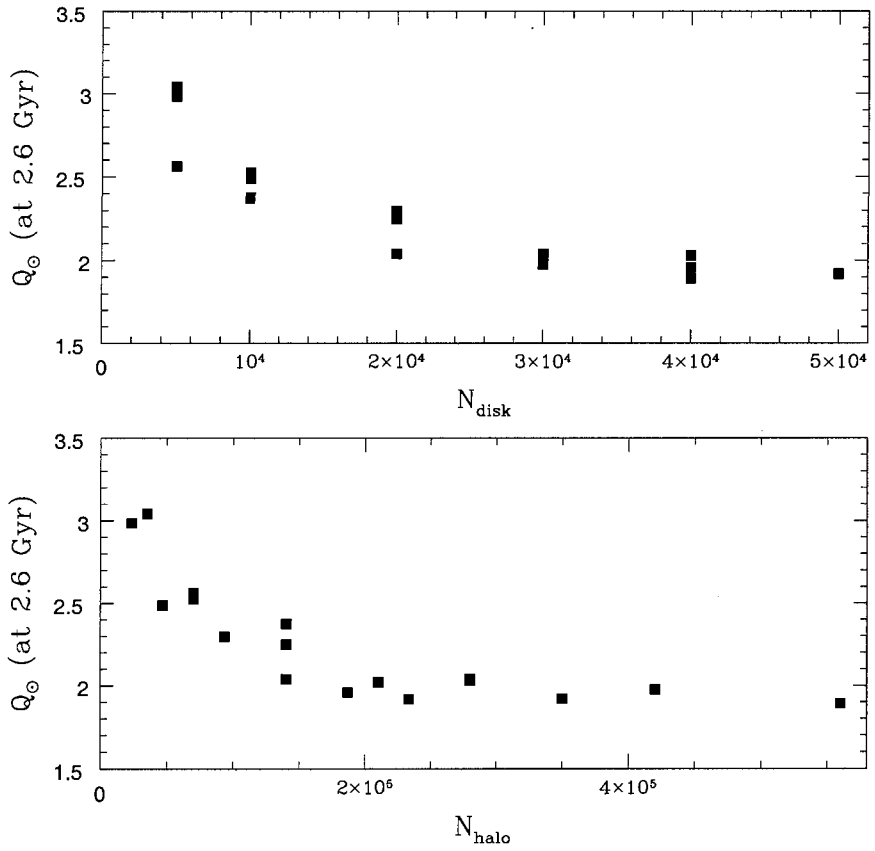

Figure 3.5: z_{median} after 2.6 Gyr versus N_{halo} and N_{disk} for models A – S.

and

$$Q(R_\odot)\,(t = 2.6\,\mathrm{Gyr}) \propto N_h^\alpha N_d^\beta. \tag{3.7}$$

Since it is only the ratio α/β that is important in the scaling, Figure 3.7 shows the rms deviation from the best fit as a function of α, keeping $\beta = 1$. From Figure 3.7, it is clear that z_{median} and $Q(R_\odot)$ are best predicted by equations (3.6) and (3.7) when α is of order unity. In other words, the data suggest that the numerical heating is mainly a function of the product $N_d N_h$. This result suggests that, in a first approximation, it is the direct collisions between the disk and halo particles which contribute the most to the heating rate.

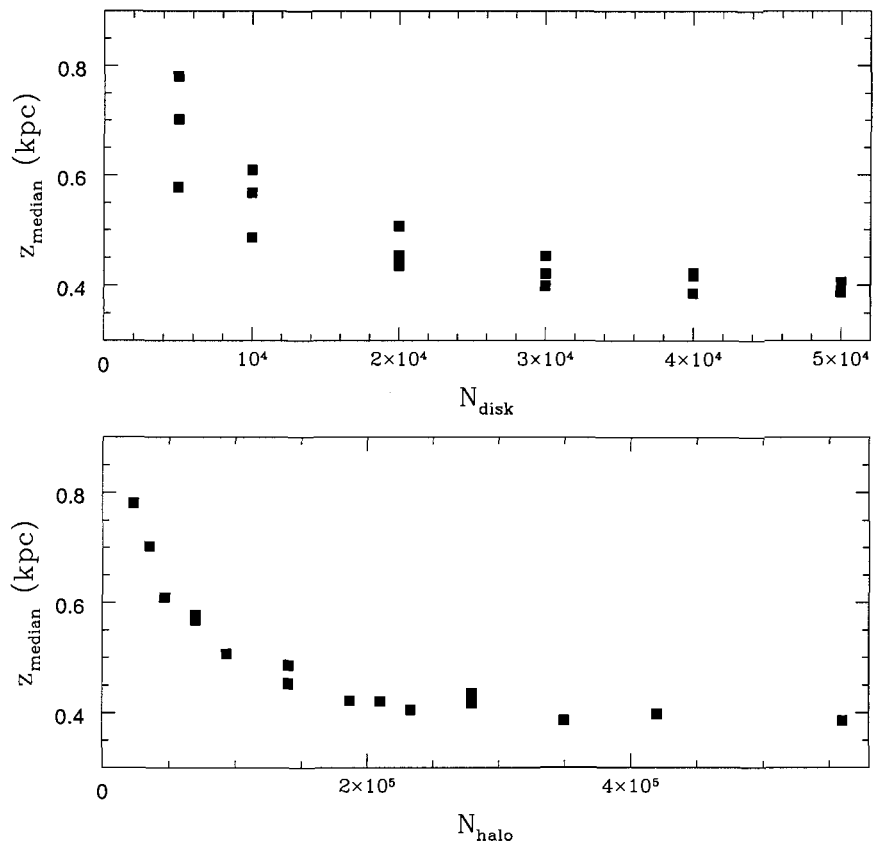

Figure 3.6: $Q_\odot$ after 2.6 Gyr versus N_{halo} and N_{disk} for models A – S.

Another type of collisions which can potentially contribute to the heating rate in the disk are the collisions between the disk particles themselves. We can gauge this effect by estimating the collisional relaxation time in the disk (see, for example, Shu 1982),

$$t_{relax} = \frac{V_d^3}{4\pi G^2 m_d^2 n_d \ln(\frac{RV_d^2}{Gm_d})}, \tag{3.8}$$

where m_d is the mass of a disk particle, n_d is the number density of disk particles, V_d is the total velocity of a "typical" disk particle and R is the radius in the disk where the measurement is made. We further the nominal values for a typical disk particle:

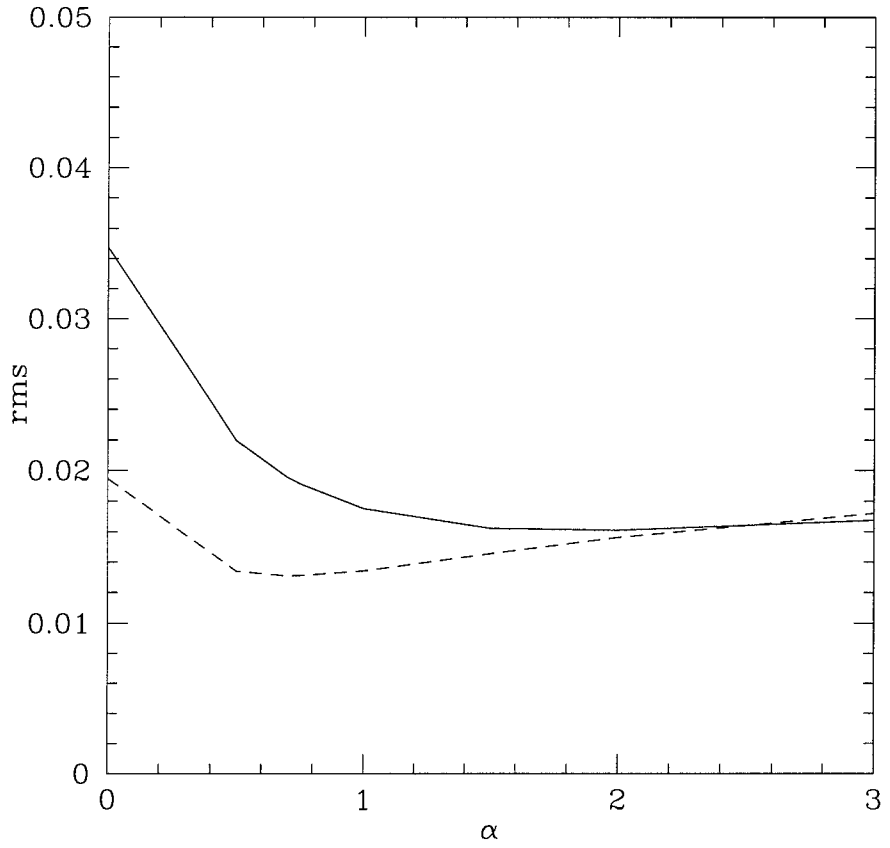

Figure 3.7: The rms of the best fits for the equations (3.6) and (3.7) versus α (β is kept constant and equal to 1). The dashed line gives the rms values for z_{med} function and the full line for $Q_{\odot}$.

$m_d = 5.6 \times 10^5 M_{\odot}$, $V_d \simeq 40$ km/s and $R \sim 10 - 20$ kpc. If we take model M as an example, the total number of disk particles is $N_d = 40,000$. The number denisty of disk particles, n_d can be easily calculated by dividing the number of disk particles over the volume (note that the disk height can be approximated as twice the vertical scale height of the disk, $2z_0 \simeq 1.4$ kpc). With these numbers we estimate the relaxation time to be $t_{relax} \simeq 4.8 \times 10^{13}$ s. In comparison, the orbital period of the stars in the disk at the solar radius is $T_{orb} \simeq 2.34 \times 10^8$ yr $\simeq 7.39 \times 10^{15}$ s. The relaxation time is thus much shorter than the orbital period in the disk and therefore, collisions

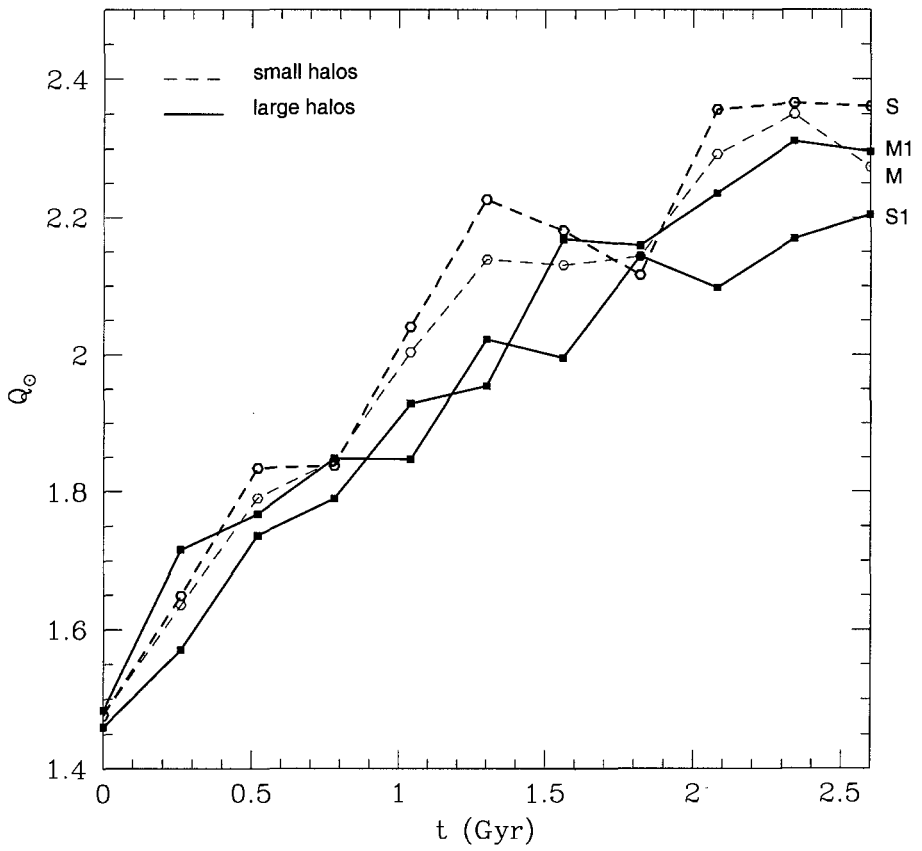

Figure 3.8: The evolution of $Q_\odot$ for a sample of small and large halo models.

between disk particles must influence to some degree the heating of the disk[2].

Is the heating due to local collisions (i.e. involving only disk particles) or is it caused by some global instability process? The latter hypothesis may be tested by studying the effect of collisions between halo particles. For this, we have adopted a different halo model, as described in Table 3.2. Models $M1$ and $S1$ are identical

[2]Note that the relaxation time calculated here refers only to the *disk particles* in our models, which are about 6 orders of magnitude more massive and about 7 orders of magnitude less numerous than the average disk stars (there are about 3×10^{11} stars in the disk, typically weighting $\sim 0.5\ M_\odot$; see for example, Shu 1982, p265). In the Milky Way disk, the relaxation time is $\simeq 5 \times 10^{21}$ s $\gg T_{orb}$ and, therefore, collisions between *disk stars* do not contribute to the heating rate in the disk.

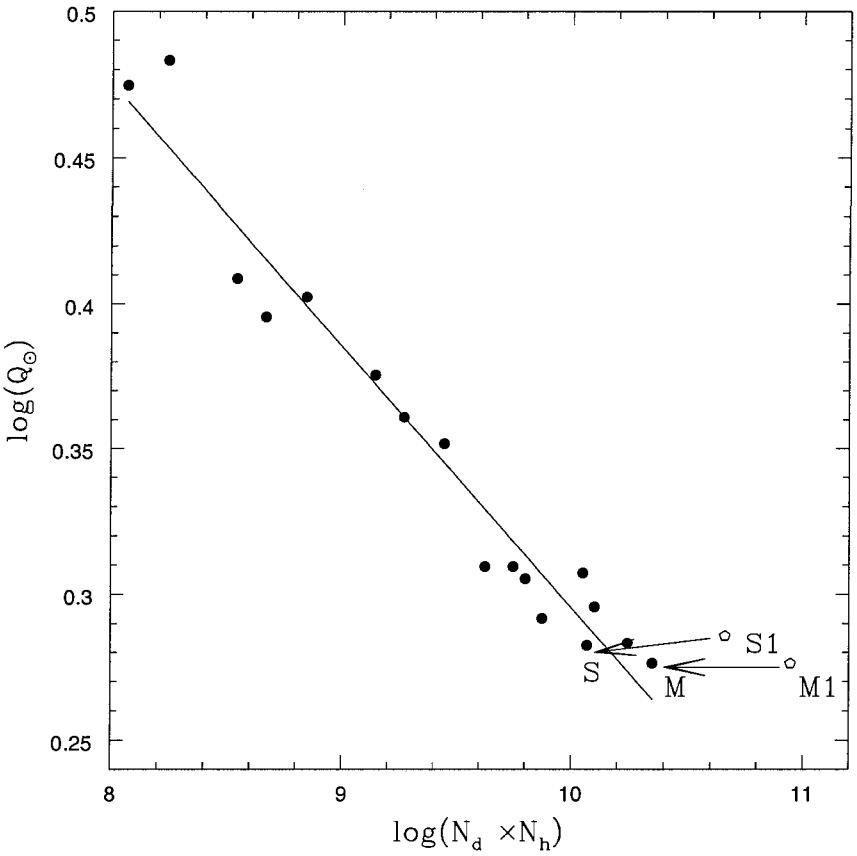

Figure 3.9: Best fit line is given by $log(Q_\odot) = 1.194 - 0.089 \cdot log(N_h N_d)$. The full circles represent the small halo models, and the empty circles comparatively, the four large halo size models.

to models M and S inside 84 kpc, but the cut-off radius of the halo is now $\sim$ 300 kpc (and the corresponding total mass of the halo is $M_h = 55.18$). Since most halo particles in models with large halo cut-off radius are added well beyond the confines of the disk, one may expect the heating rates of models with large halo sizes to be similar to those of models with small halo sizes with the same m_h and m_d, even though the former have more than *twice* as many particles. This is shown in Figure 3.8 which follows the evolution of $Q_\odot$ for two small (M, S) and large halo ($M1$, $S1$) models. Figure 3.9 shows the $Q_\odot$ measured at $t = 2.6$ Gyr as a function of the product $N_d N_h$

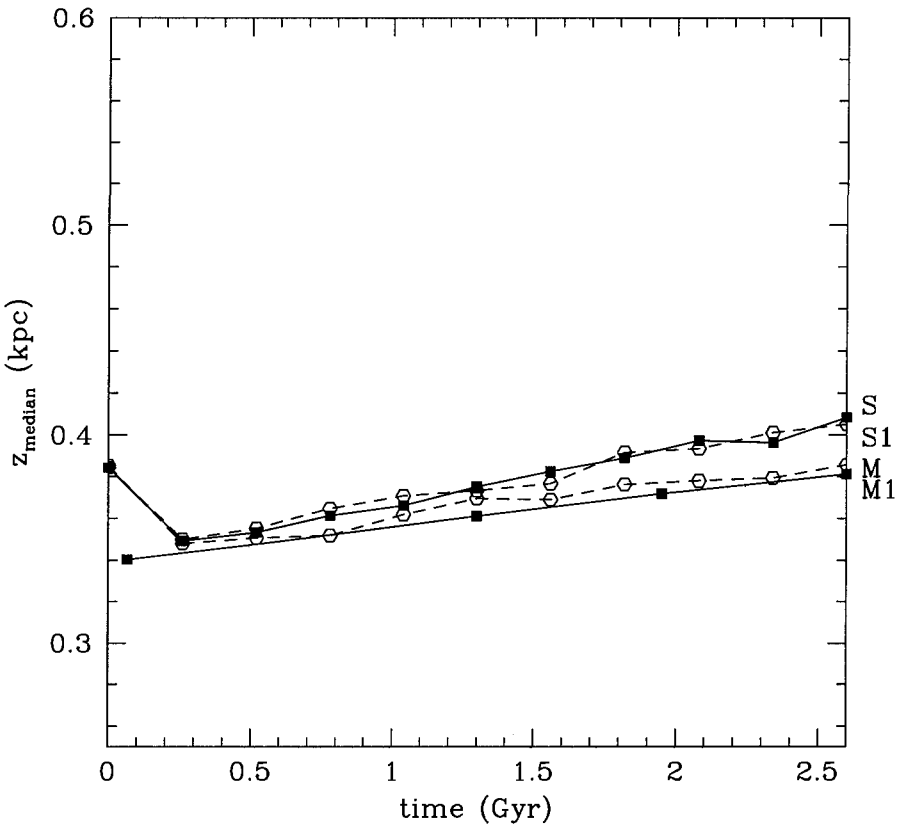

Figure 3.10: The evolution of z_{median} for a sample of small and large halo models.

for all small and large models. Filled circles correspond to models A-S, open circles to the large halo models, $M1$ and $S1$. These plots suggest that the larger halo models experience roughly the same radial heating as small halo ones.

Figure 3.10 compares the vertical heating for the same set of two small (M, S) and large halo models ($M1$, $S1$); and Figure 3.11 shows z_{median} at $t = 2.6$ Gyr as a function of the product $N_d N_h$ for all small and large halo models. Symbols are the same as in Figure 3.9. The arrows connect the corresponding large and small models. These plots suggest that the disks in large halo models experience similar vertical heating as small halo models. Clearly, the halo particles which are able to influence

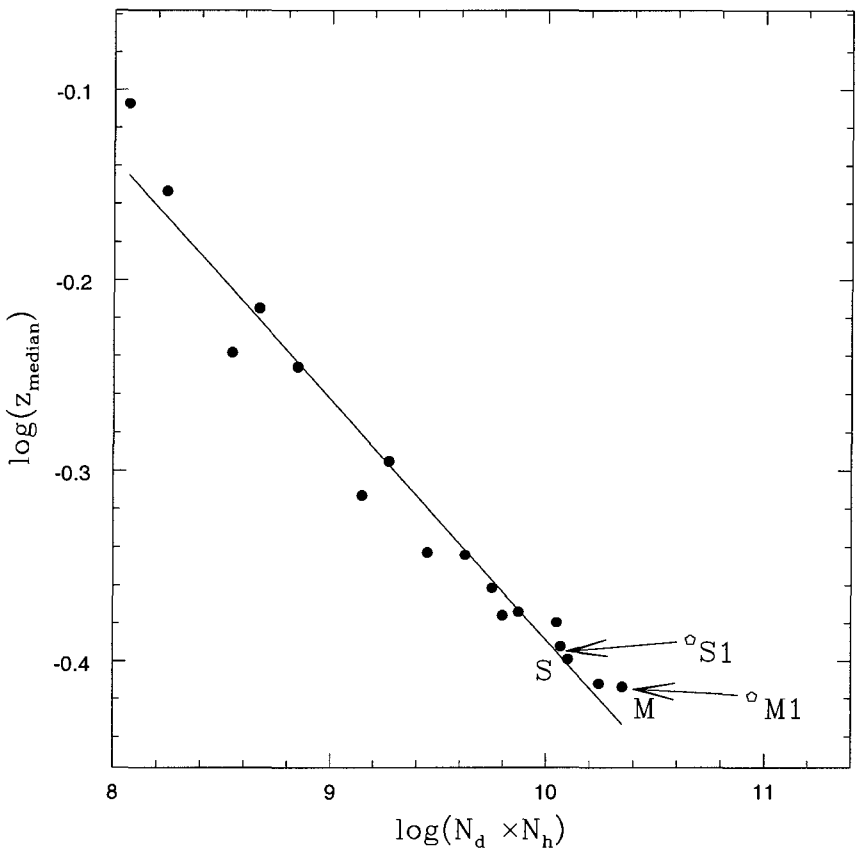

Figure 3.11: Best fit line is given by $log(z_{med}) = 0.874 - 0.126 \cdot log(N_h N_d)$. The full circles represent the small halo models, and the empty circles comparatively, the four large halo size models.

significantly the disk heating rate are not those at large radii, but those in the inner region (implicitly, halo particles which directly hit the disk are expected to have the maximum impact on the heating rate).

We conclude that, in a first approximation, the two body heating mechanism is influenced primarily by direct disk - halo particle collisions, i.e. the heating process is mainly a local one. Secondary (although not negligible) contributions to the heating rate are also provided by disk - disk particle collisions. Global halo particle collisions are not found to contribute significantly to the disk heating rate.

We have found that numerical heating of the disk depends roughly on the number of particles in the simulation as $\frac{1}{N_h^{\alpha} \cdot N_d^{\beta}}$, where $\alpha \sim \beta \sim 1$. This result suggests that for a certain value of the product $N_h N_d$, increasing either of the parameters N_h or N_d will have approximately the same effect in reducing the numerical heating of the disk. These results are helpful for improving the efficiency of numerical models that attempt to follow the evolution of galaxies like the Milky Way over a Hubble time.

Also, from the evolution of z_{median} and $Q_{\odot}$ in Figures 3.3 – 3.4, we conclude that the number of particles required to bring the galaxy model down to an acceptable level of numerical heating is about $N_d N_h \sim 10^{10}$ within a radius of about 100 kpc, which is affordable with present computational resources.

3.4 The "Milky Way" Galaxy Models

3.4.1 The Isolated (Disk/Bulge/Halo) Model

The final galaxy model we have adopted is a version of Model M where the physical parameters are chosen so as to reproduce the observed properties of the Galaxy (with the exception of some of the halo parameters, this set-up is similar to that of VW99). Table 3.4 lists the physical parameters adopted for the Galaxy model. The density profiles of the galactic components are given by equations (3.1), (3.2) and (3.3). The disk is assumed to have a total mass of $5.6 \times 10^{10} M_{\odot}$ (similar to the value of $6 \times 10^{10} M_{\odot}$ compiled by BT87, p13), an exponential scale length of 3.5 kpc (the observed value is 3.5 ± 0.5 kpc; de Vaucouleurs & Pence 1978; Bahcall, Schmidt & Soneira 1982), and a vertical scale height of 700 pc (corresponding to the disk thickness measured for stars similar to the Sun; Bahcall & Soneira 1980). The bulge is modeled with a Hernquist model of total mass $1.87 \times 10^{10} M_{\odot}$ and 525 pc scale radius.

The large radial extent of the dark halo (which extends out to 300 kpc) is more realistic than in previous numerical studies. For example, VW99 use a halo which extends to about 84 kpc, and Huang & Carlberg (1997)'s halo model has a half-mass

Table 3.4: Parameters of the Galaxy model

	Symbol	Value	Value in code units
Disc:			
	M_D	$5.6 \times 10^{10} M_\odot$	1
	R_D	3.5 kpc	1
	z_0	700 pc	0.2
	Q_0	1.5	
	$R_\odot$	8.5 kpc	2.42857
	ϵ_D	175 pc	0.05
Bulge:			
	M_B	$1.87 \times 10^{10} M_\odot$	0.33333
	a_B	525 pc	0.15
	ϵ_B	175 pc	0.05
Halo:			
	M_H	$3.09 \times 10^{12} M_\odot$	55.18
	γ	3.5 kpc	1
	r_{cut}	300 kpc	85.71
	ϵ_H	175 pc	0.05

radius of 16 kpc. Although these values are reasonable approximations to the mass enclosed within the radius considered in their analysis, the true extent and mass of the halo of the Milky Way are likely to be much larger. Cosmological simulations of the formation of cold dark matter halos show that halos with circular velocities of order 200 km/s have masses $\sim 10^{12}$ $M_\odot$ and virial radii of $\sim 200 - 300$ kpc (eg. NFW). Unfortunately, the methods employed in estimating the mass of the dark halo directly from observations have not been able so far to pin down the exact value, mainly because few observational probes extend beyond ~ 100 kpc. However, methods which can probe large distances tend to support the cosmological claim. For example, estimates of the halo mass from the dynamics the ensemble of satellite galaxies in the Local Group (Einasto & Lynden-Bell 1982; Lynden-Bell, Cannon & Godwin 1983; Zaritsky et al 1989) typically give values in the range $1.5 - 2.6 \times 10^{12}$ $M_\odot$ for the dark mass within ~ 200 kpc. A somewhat lower value is obtained by Dubinski, Mihos & Hernquist (1996), who attempt to set an upper limit on the halo

mass and size by examining the length of tails created by spiral-spiral mergers. They suggest that, in order to form lengthy tails, such as those observed in two interacting galaxies NGC 4038/39 (also known as "the Antennae"), the mass of the halo cannot exceed $6 \times 10^{11}\ M_{\odot}$, within a radius of < 100 kpc (although see Springel & White 1999 for a different view, which argues that the most important factor in determining the tail lengths is not the halo to disk mass ratio, but the way matter is distributed in these components). From the orbital dynamics of the Large Magellanic Cloud (LMC), Lin, Jones & Kremola (1995) estimated that the mass of the Galactic halo is of the order of $5.5 \pm 1 \times 10^{11}\ M_{\odot}$ up to a distance of ~ 100 kpc.

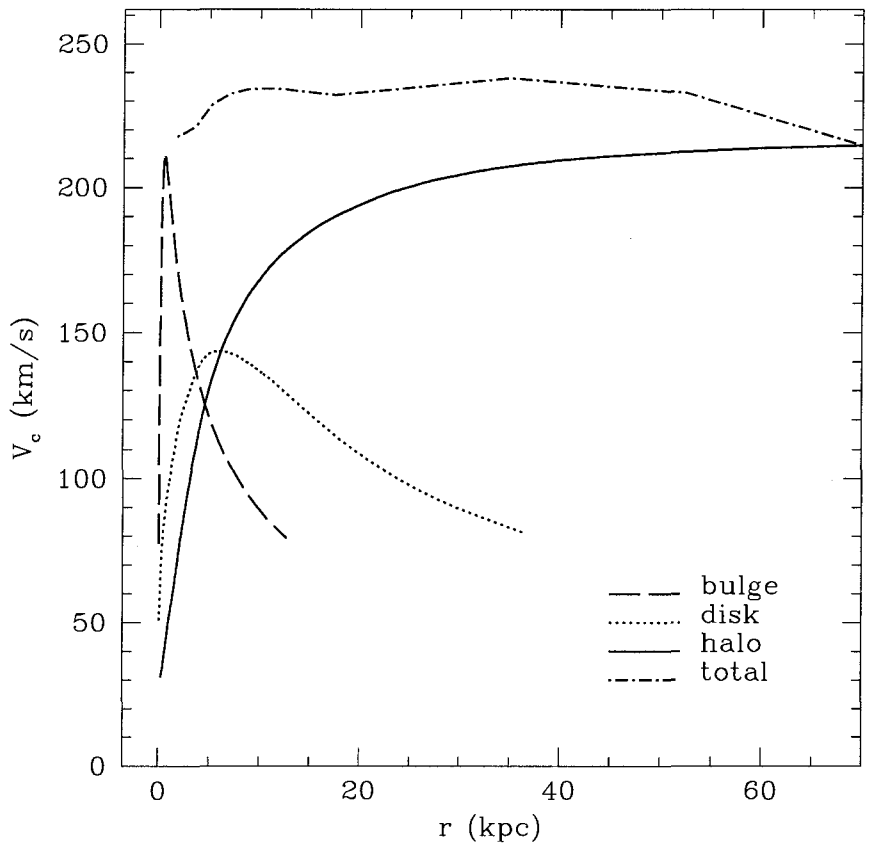

Figure 3.12: Contribution to the circular velocity profile of the disk, bulge and halo components in the Milky Way model.

Figure 3.12 shows the circular velocities for the three galactic components. The circular velocity for the disk (calculated as $V_c = Rd\Phi/dR$, where R is the radial distance in the disk) peaks at about $2.2R_d = 7.7$ kpc. For bulge and halo, the circular velocities are calculated as $V_c = GM(r)/r$ where $M(r)$ is the mass corresponding to each component within the radius r. The total circular velocity attains $\sim$ 220 km s^{-1} at about 8.5 kpc (i.e. solar circle), similar to the one observed in the Milky Way (the measured value is 220 $\pm$ 15 km s^{-1} at 8.5 $\pm$ 1 kpc; see BT87 p.17; Mihalas & Binney 1981, §8.3 and references therein).

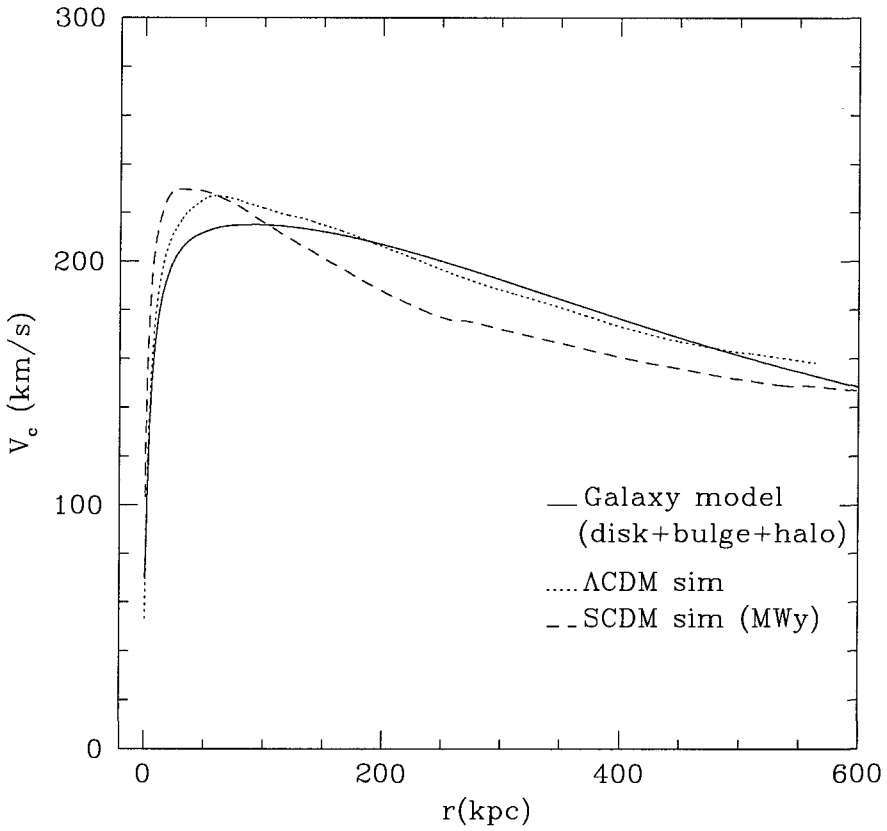

Figure 3.13: Comparison between the circular velocity curve of the dark halo in the Milky Way model and those of galactic halos in the SCDM and ΛCDM cosmological simulations.

The dark matter halo built in the present simulations weighs $3.09 \times 10^{12} M_\odot$ (ap-

proximately 60 times more massive than the disk), and has been modeled with a non-singular isothermal density profile, with a core radius $r_c = 3.5$ kpc and exponentially truncated at a distance of 300 kpc from the center. The circular velocity profile of the galaxy model closely resembles that of the ΛCDM and SCDM halos out to ~ 600 kpc, about twice its virial radius, as illustrated in Figure 3.13.

The galaxy model was first evolved in isolation, in order to assess the deviations from equilibrium induced by numerical noise in the particle distribution. The simulation uses $40,000$ particles in the disk, $13,333$ in the bulge, and 2.2×10^6 in the halo, all of equal mass, $m_p = 1.4 \times 10^6 M_\odot$ and has been run with PKDGRAV, using a spline softening length of 175 pc and a hierarchy of timesteps, the smallest of which is typically less than 10^5 yrs.

3.4.2 Including Dark Matter Halo Substructure

Substructure halos, despite their large numbers, make up a small fraction of the total mass of the system (approximately 10% of the mass of the dark halo), implying that it is possible to add subhalos to the disk/bulge/halo galaxy model without seriously compromising the global equilibrium of the system. Here we present the procedure for including subhalos in the isolated galaxy model.

As discussed in Chapter 2, the sub-halo parameters (positions, velocities and structural parameters) are provided directly by the group-finding algorithm, SKID. We construct three separate galaxy models in which we include satellites of MWy (or "Halo 1", SCDM) and M31 ("Halo 2", SCDM) and the high resolution ΛCDM halo. We selected all satellites within twice the virial radius because most of them have time to reach the pericenter of their orbits (near the disk) during the timespan of the simulations.

Since the dark matter halos in the cosmological models differ from the galactic halo constructed with the Hernquist method, the parameters of the satellites need to be rescaled. The positions and velocities of each satellite are scaled so that they

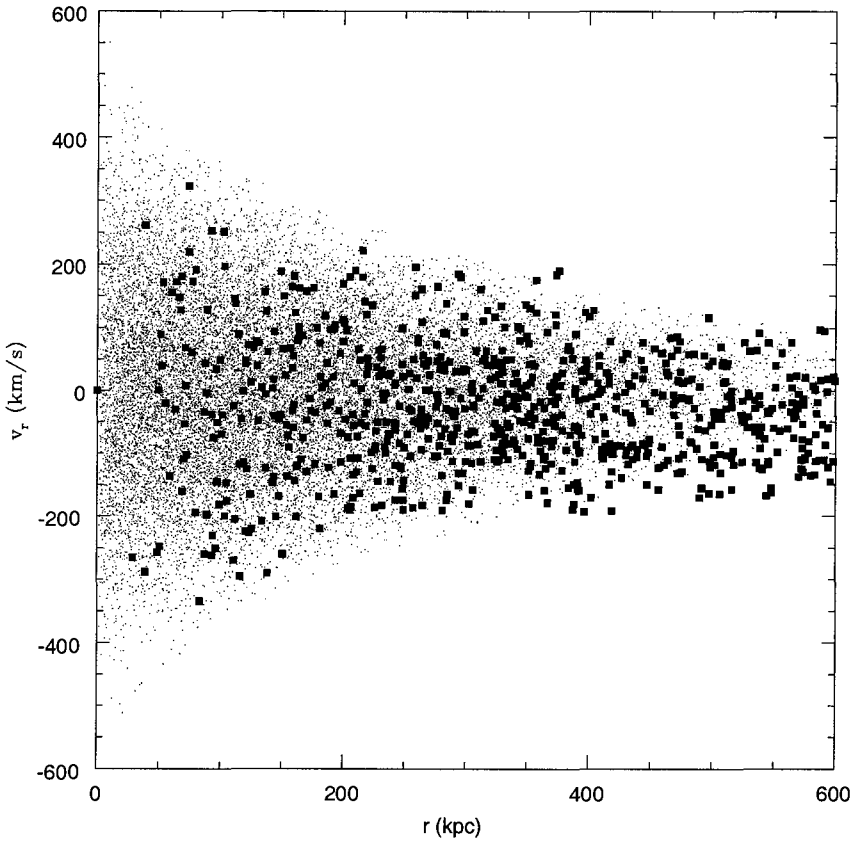

Figure 3.14: The radii and radial velocities ($v_r = \vec{v} \cdot \frac{\vec{r}}{|\vec{r}|}$) of the original dark matter satellites (solid squares) are rescaled to the Milky Way model (model shown corresponds to the MWy halo).

correspond to the same fractions of $R_{vir}(= R_{200})$ and $V_{vir}(= V_{200})$, respectively, in both the parent CDM halos and in the galaxy model (see Figure 3.14).

We model the internal structure of dark matter satellites by the "universal" density profile of Navarro, Frenk & White (NFW 1996; 1997). This profile has the general form:

$$\frac{\rho(r)}{\rho_{crit}} = \frac{\delta_c}{(\frac{r}{r_s})(1+\frac{r}{r_s})^2}, \tag{3.9}$$

where $\rho_{crit} = 3H^2/8\pi G$ is the critical density of the universe, $\delta_c = \frac{200}{3} c^3/[ln(1+c) - c/(1+c)]$ is a characteristic overdensity, $r_s = R_{200}/c$ is a scale radius, and c is the concentration parameter. For given halo mass, the NFW model has a single free parameter, the concentration c.

We can further recast equation (3.9) in the form:

$$\rho(r) = \frac{\rho_0 {r^3}_s}{r(r+r_s)^2}, \tag{3.10}$$

where r_s is the scale radius and $\rho_0 = 4\rho(r_s)$.

The NFW satellite model parameters are chosen to match the values of $r_{\rm peak} \equiv 2r_s$ and $V_{\rm peak}$ (the maximum circular velocity) returned by SKID. These two parameters, $r_{\rm peak}$ and $V_{\rm peak}$, fully specify each NFW halo model. In order to ensure a finite mass we truncate each NFW model at $5\,r_{\rm peak}$. The masses of the satellite particles are set to be equal to those of the other components of the galaxy model, i.e. $m_{sat} = m_d = m_b = m_h$.

An example of the final galactic model is shown in Figure 3.15, which shows the edge-on view of the disk/bulge/halo Milky Way galaxy model including the substructure corresponding to the ΛCDM simulation.

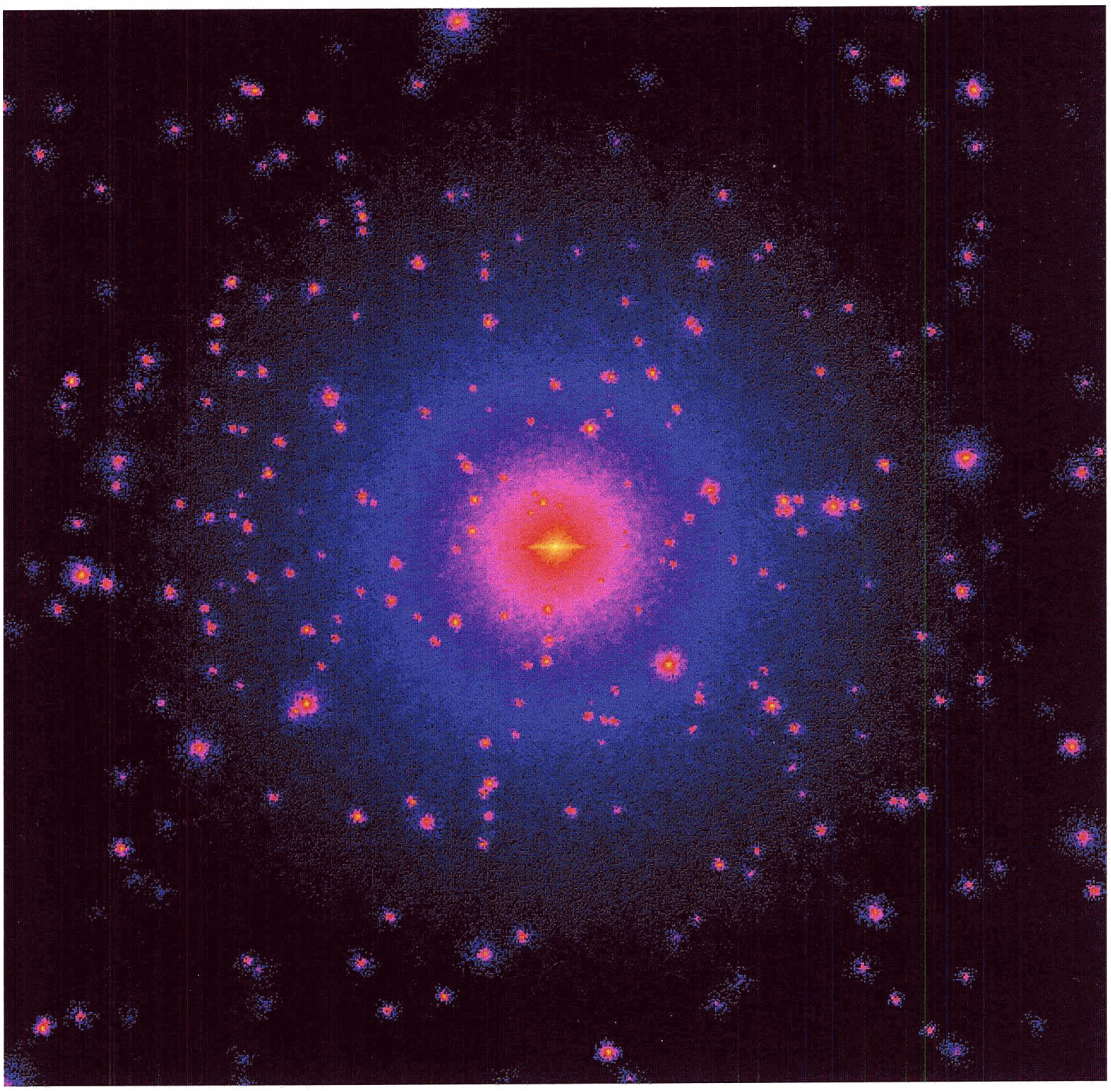

Figure 3.15: Edge-on view of the disk/bulge/halo Milky Way galaxy model including substructure halos. The box is 600 kpc on the side (about twice the virial radius) and shows the disk and bulge at the center (in yellow), surrounded by a dark halo populated with several hundred dark matter satellites. The halo and its substructure extend well beyond the size of the luminous disk and therefore collisions between substructure and the stellar disk occur quite infrequently.

3.5 Orbital Parameters of Dark Matter Sub-halos

In addition to the physical parameters that characterize the structure of the satellite (such as their bound mass, tidal radius, internal velocity dispersion or peak circular velocity), it is also important to know their orbital parameters. This is important because it is mostly the satellites that approach the disk the ones that may cause significant damage. The pericenter can normally be identified directly from the simulations. Alternatively, the pericentric radius of the sub-halos may be derived entirely analytically (under some simplifying assumptions). For example, as a first approximation, the satellite can be regarded as a point mass evolving in a static spherical potential. In this case its orbit can be easily integrated. Because of the spherical symmetry of the problem, it is appropriate to use the polar coordinates (r, ψ), where r is the distance to the center of the orbit and ψ is the azimuthal angle in the orbital plane. With an additional coordinate transformation, $u \equiv \frac{1}{r}$, one can write the "radial energy" equation (BT87, p. 105):

$$(\frac{du}{d\psi})^2 + \frac{2\Phi}{L^2} + u^2 = ct = \frac{2E}{L^2}, \tag{3.11}$$

where $\Phi = \Phi(u)$ is the potential energy, L is the angular momentum (which is assumed to be a conserved quantity) and E is the total energy. With the condition at turning points (pericenter or apocenter) $du/d\psi = 0$, the above equation becomes:

$$u^2 + \frac{2(\Phi(1/u) - E)}{L^2} = 0 \tag{3.12}$$

The equation (3.12) admits two real roots, u_1 and u_2, between which the satellite oscillates radially as it simultaneously revolves in the azimuthal direction ψ. The inner radius, $r_1 = {u_1}^{-1}$, corresponds to the pericenter and the outer radius, $r_2 = {u_2}^{-1}$, to the apocenter distance.

This analytical treatment is likely to be accurate only for the first few pericentric passages, since the tidal field of the galaxy continuously strips material from the

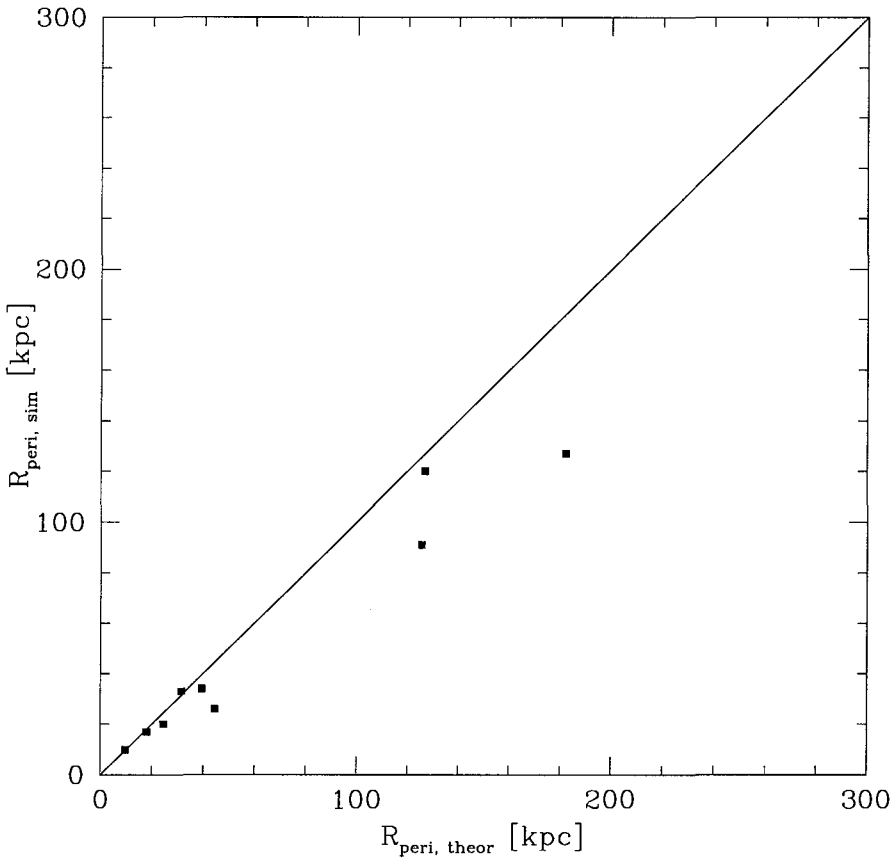

Figure 3.16: Comparison between the pericenter values obtained with the analytical formula (3.12) and by measuring them directly in the simulation. This is shown only 9 of the most massive satellites in one of the galactic halos ("MWy") from the SCDM model. The full line corresponds to a 1:1 relationship.

satellite and dynamical friction forces the orbit to decay. Figure 3.16 shows good agreement between the values of the pericenter of the first orbit obtained analytically and measured directly from the simulation. The result gives us confidence that, at least for the case of the first few orbits the analytical treatment is a satisfactory approximation. Pericenters of all dark matter satellites located within $2R_{vir}$ distance from the centers of the galactic halos have been derived in this way.

Figure 3.17 shows, for example, the masses and first pericentric radii for the sub-

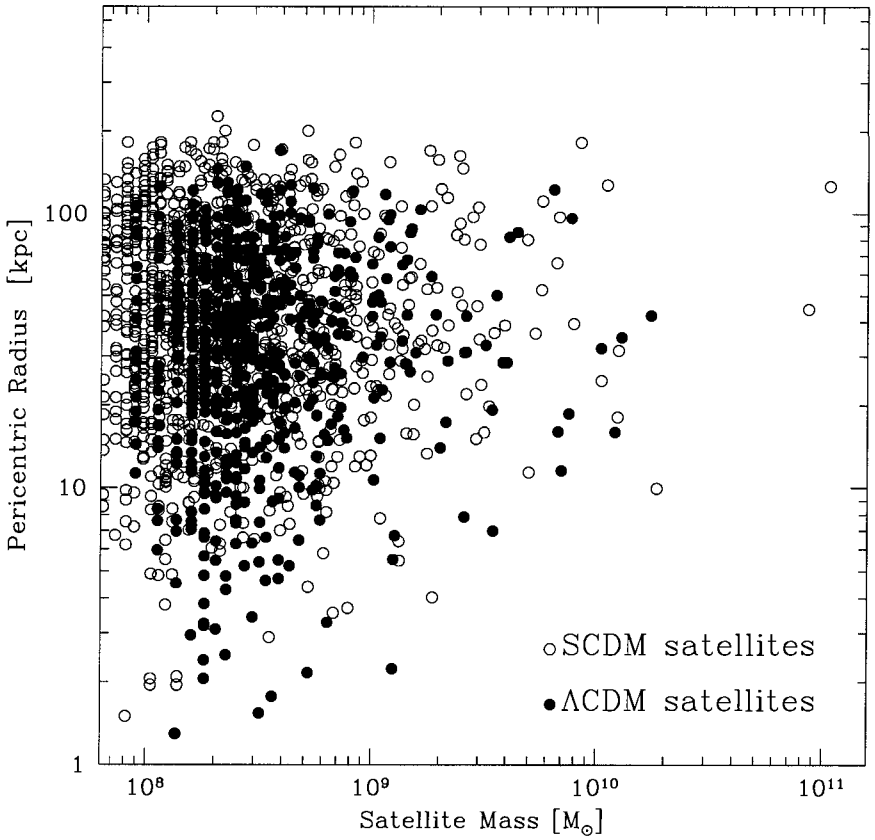

Figure 3.17: Masses and pericentric radii for subhalos identified in "Halo 1", i.e. "MWy", in the SCDM run and ΛCDM halo, after scaling to the virial radius and circular velocity of the Milky Way disk model. (Note that satellites within $2R_{vir}$ of the primary halo are selected).

halos identified in two of these simulations: the "Halo 1" (SCDM) and the high resolution (ΛCDM) halo. Several interpretations may be drawn from this figure. Firstly, it is clear that the two cosmological models do not differ significantly from each other in terms of the orbital information, suggesting that this orbital distribution may be fairly typical and representative for the CDM cosmology. Indeed, Ardi, Tsuchiya & Burkert (2003) show that the distribution presented in Figure 3.17 is consistent with the distribution of field particles in an equilibrium halo, with the exception of the

inner region, where satellites are conspicuously missing. This is not at all surprising, since the mechanism of tidal disruption operates very efficiently at small distances and, thus, eliminates most satellites near the center. This aspect is also evident in Figure 3.17, which shows that pericentric values are, on average, well beyond the confines of the disk ($\sim$ 20 kpc); only about 3% of the total population of satellites is able to reach within the solar circle. Of these, according to Figure 3.17, the majority have masses less than $\sim 10^9 M_\odot$.

Summary:

We describe the construction of a realistic equilibrium model of the Milky Way galaxy. As a first step, we study the secular evolution of the galaxy disk, in order to quantify the effect the artificial heating caused by two-body interactions. We show that heating in the disk is mainly a local process, where disk - inner halo and disk - disk particle collisions play the major role. We find that the heating of the disk increases directly proportional with the number of disk and halo particles, $\sim N_d \times N_h$. Moreover, we show that an accurate model of Milky Way galaxy requires of order $N_d N_h \sim 10^{10}$ particles, which is affordable with the present computational resources. In the second part, we describe the method used for incorporating realistic dark matter satellites in our Galaxy model. These are chosen to match the substructure realizations in two representative halos in the SCDM cosmology and in one ΛCDM halo.

Chapter 4

Disk Heating by Cold Dark Matter Substructure

Abstract

We examine recent suggestions that substructure in cold dark matter (CDM) halos may be in conflict with the presence of thin, dynamically fragile stellar disks. N-body simulations of an isolated disk/bulge/halo model of the Milky Way including substructure halos with masses, densities and orbits derived from high-resolution cosmological CDM simulations, indicate that the disk is able to survive relatively unscathed for as long as a Hubble time. Our simulations also suggest that heating by the CDM substructure is not the dominant heating mechanism in the solar neighborhood. However, at large disk radii, the heating by dark matter substructure can be substantial and may explain the increase in disk scale height with radial distance observed in spiral galaxies.

4.1 Measuring the Disk Heating

In this Section, we analyze the evolution of different parameters that reflect the heating of the disk: the disk thickness (measured by either the median disk height z_{median} or the disk scale height z_0 (equation (3.1)); the velocity dispersions in the

three Galactic coordinates, σ_R, σ_ϕ and σ_z, and the total velocity dispersion, $\sigma_{tot} \equiv \sqrt{({\sigma_R}^2 + {\sigma_\phi}^2 + {\sigma_z}^2)}$; and the Toomre parameter Q (as defined in the equation (3.5)). This section is structured as follows: §4.1.1 defines the system of coordinates in which these quantities are measured; and §4.1.2 presents the results for the three simulations with substructure (with initial conditions taken from the ΛCDM Halo and the two SCDM halos, "Halo 1" , i.e. "MWy" and "Halo 2", i.e. "M31") and, as a control, a simulation without substructure.

4.1.1 The Coordinate System

Tilting and precessing of the disk are physical mechanisms by which the disk re-adjusts the orientation of its angular momentum to account for the effect of orbiting satellites. As a satellite's orbit decays towards the disk, some of its angular momentum can be lost to the disk, inducing changes in its orientation. Thus, the plane of the disk may tilt and precess as the result of tidal interactions[1], a process that is indeed seen to occur in our simulations.

The physical parameters that characterize the heating of the disk have to be calculated with respect to the current plane of the disk. Neglecting the re-orientation of the disk plane may result in projection effects which introduce a spurious increase in the quantities we are trying to determine. This means that we have to define a new system of coordinates coincident with the instantaneous principal axes of the disk.

As is well known, if we start with an initial cartesian coordinate system (x, y, z), we can arrive to any given final system (x', y', z'), through two successive rotations of the axes:

$$(x, y, z) \rightarrow (x_\star, y_\star, z_\star) \rightarrow (x', y', z'),$$

[1]The implications of the tilting of the disk will be analyzed in more detail in §5.2. However, for the specific problems addressed in this chapter, the tilting of the disk is considered only in order to setup a proper coordinate system where disk heating may be measured.

where the two transformations denote the two successive rotations: a rotation of angle ϕ around the initial z axis (note that, in this case, $z_\star \equiv z$), and a subsequent rotation with the angle θ around the $y_\star$ axis (where the final axis y' is identical with $y_\star$).

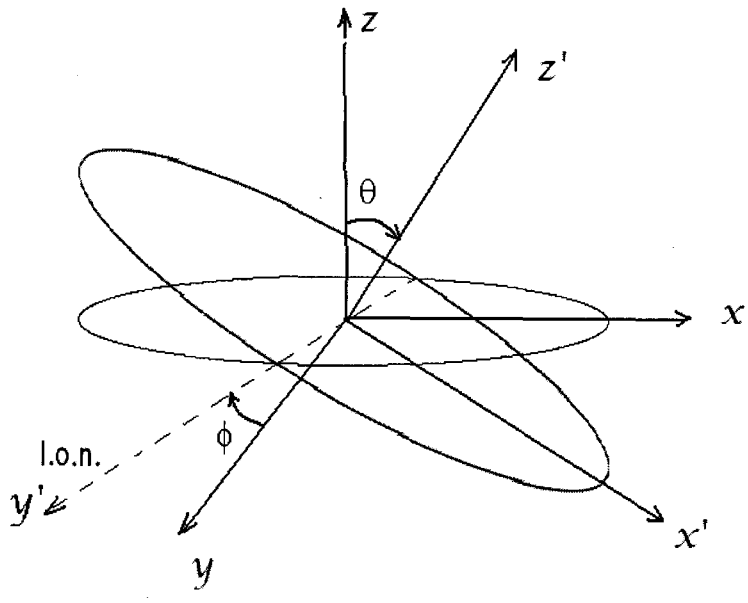

Figure 4.1: The coordinate transformation from the initial system (x, y, z) to the (x', y', z') system can be done through two successive rotations.

In mathematical form, the two rotations are defined by the two matrices:

$$R_1 = \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and

$$R_2 = \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix}$$

whereas, the entire transformation is the multiplication of the two matrices: $R = R_1 \times R_2$, or, in other words, $(x', y', z') = R \times (x, y, z)$.

We choose the axis z^{prime} to coincide with the instantaneous direction of the angular momentum of the disk, $\vec{L_{tot}}$.

For completeness, we mention that we also include another correction which accounts for the shift in the center of mass of the disk which may occur during the interaction with the ambient satellites. This effect is generally negligible because of the much larger mass of the disk in comparison with that of an average sub-halo (we noticed a shift at the most of a few kpc, occurring for the most massive, $\sim 10^9 - 10^{10}$ $M_\odot$, satellites). This correction implies recentering all the disk particle positions and velocities relative to the position of the center of mass of the system. All results presented from now on include, implicitly, the above mentioned corrections.

4.1.2 Disk Secular Evolution

In the following, we will present results regarding changes in the vertical structure of the disk, as well as in the velocity dispersion of disk stars, for the case of the galaxy model that includes substructure in the ΛCDM and SCDM models. For comparison, the same parameters will be monitored in the case of the isolated galaxy model. The two simulated Galaxy models with dark matter substructure drawn from two representative present-day halos in the SCDM cosmology will be referred to in short as the "Halo 1" and "Halo 2" simulations. We follow the evolution of several disk parameters in order to determine the effects of the interaction with the substructure. The parameters analyzed are: z_{median}, z_0, several σ components, and the Toomre parameter Q.

i) Disk thickness:

Figure 4.2 shows the median thickness of the disk (previously introduced in §3.3.2) as a function of time, for the model with ΛCDM substructure (dashed line) and the model without substructure (full line). In both cases, the increase in the disk thickness is slow and monotonic: over a period of 3.5 Gyr, z_{median} rises from $\sim$ 340 pc to $\sim$ 390

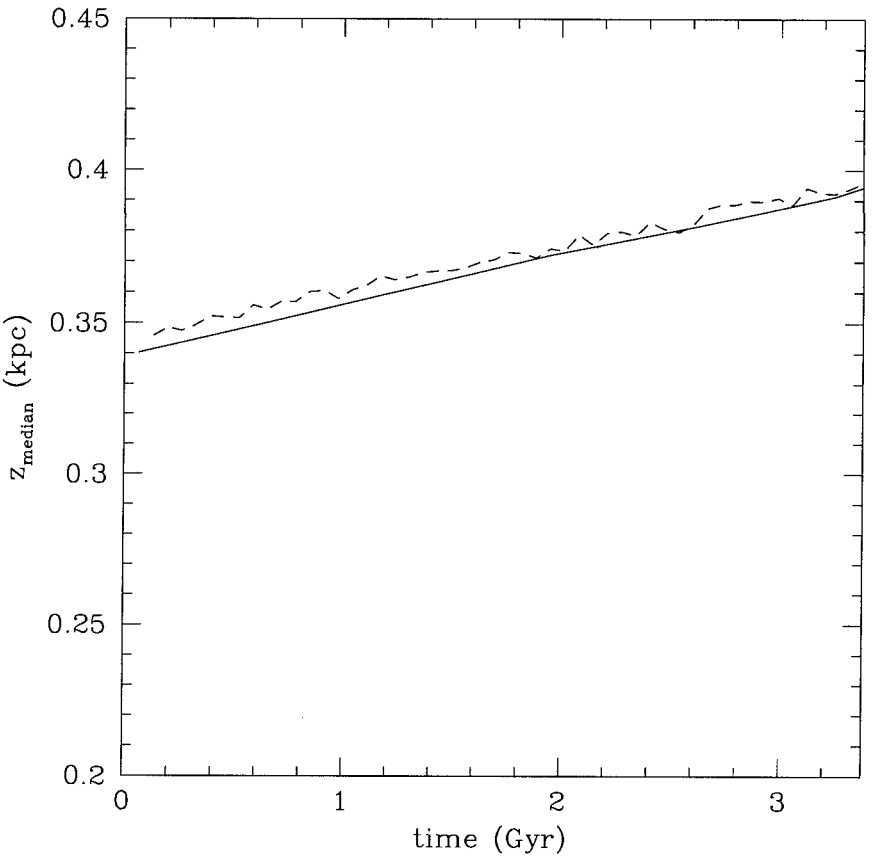

Figure 4.2: Evolution of z_{median} for the ΛCDM simulation. The full line corresponds to the measurement in the case without substructure, and the dashed line, to the case with substructure.

pc (an increase of about 15% or $\sim$ 15 pc/Gyr). The fact that the thickness of the disk evolves in a similar fashion in the case with substructure as well as in the one without substructure, suggests that the disk evolution is driven by inhomogeneities in the potential associated with the particle realizations of the disk and halo components and not by the presence of substructure.

It is possible that substructure affects mainly the outer parts of the disk, and therefore, we investigate the variation of the disk scale height with radial distance. This analysis has the advantage of allowing a direct comparison with observations

(the disk scale height can be obtained by fitting the light distribution of the disk component in spiral galaxies; eg. Schwarzkopf & Dettmar 2000). For this purpose, the disk in our simulations is divided into cylindrical bins of width 3.5 kpc ($\simeq R_d$). The vertical density distribution in these bins is then fitted with a function of the form:

$$f(z) = A \cdot sech(\frac{z}{z_0})^2, \tag{4.1}$$

where z_0 denotes the scale height at the particular radius R and A is a constant (note that this relation is similar to the equation (3.1)). The best fit is obtained with the nonlinear least-squares Marquardt-Levenberg algorithm built in the GNUPLOT software. The procedure returns the best fit values for both A and z_0. Figure 4.3 shows the best fit scale height z_0 values obtained versus the radial distance R for both the isolated model (empty squares) and the model with ΛCDM substructure (filled squares), at a snapshot after ~ 4 Gyr. For comparison, we also present the same data analysis for the isolated model at time $t \simeq 0$, that is before evolution (shown with hexagon symbols).

The $z_0(R)$ dependencies in the isolated model and the model with ΛCDM substructure are very similar, and consistent with a shallow rise of the scale height with distance. From the comparison with the $z_0(R)$ data for the isolated model at $t \simeq 0$, we infer that the changes in the two models are caused mainly by numerical relaxation of the disk and that substructure plays only a negligible role. The only hint of difference in the $z_0(R)$ behavior between the two models is in the outskirts of the disk ($R > 20$ kpc), however, the large error bars in this region prevent us from drawing a conclusion in this regard. Note that the uncertainty in the z_0 data points (which is inversely proportional with the square root of the number of particles in the bin, $1/\sqrt{N_{bin}}$) increases with the radial distance R. This mainly reflects our choice of dividing the surface of the disk in equal radial bins which, at large distances, means

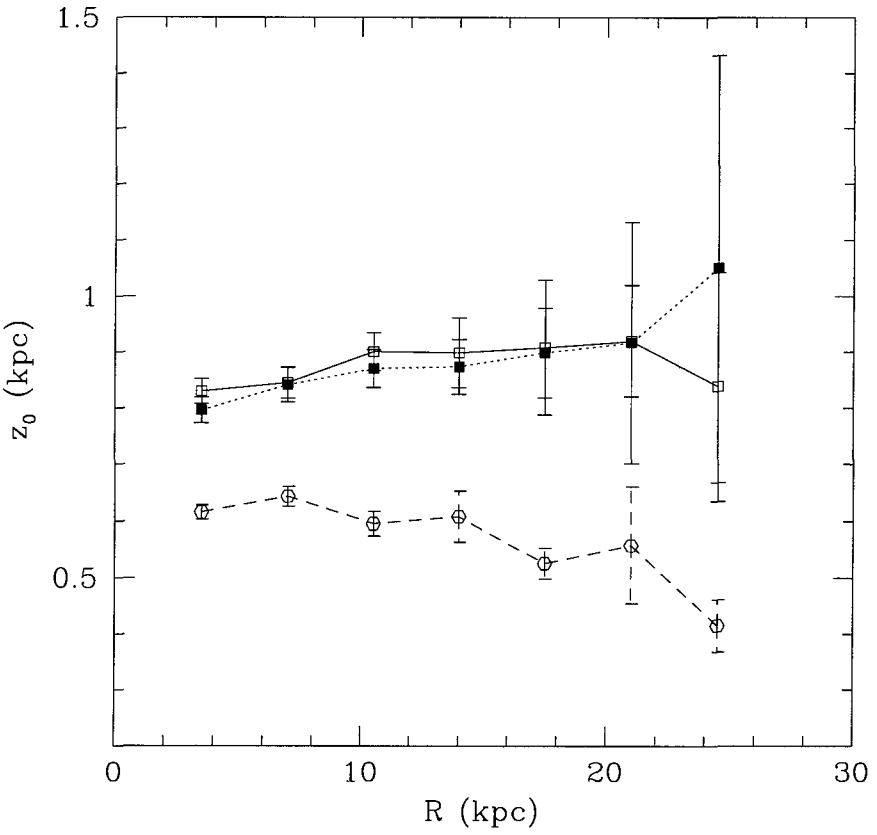

Figure 4.3: Scale height z_0 of the vertical density distribution versus radial distance R in the plane of the disk. The empty squares correspond to the isolated model and the full squares, to the model with ΛCDM substructure, measured after $\sim$ 4 Gyr of evolution. For comparison, we show with hexagon symbols the results of the same data analysis for the isolated model at time $t \simeq 0$.

that there will be fewer particles in the bins to fit equation (4.1) to[2] (in the plane of the disk, the density distribution of stars decreases exponentially with distance).

The result in Figure 4.3 is consistent with the conclusions drawn from Figure 4.2. We therefore conclude that there are no significant differences in the vertical

[2]Note, however, that all the bins with less than 50 particles have been excluded from the $z_0(R)$ plots.

structure of the disk between the simulation with ΛCDM substructure and that with no substructure.

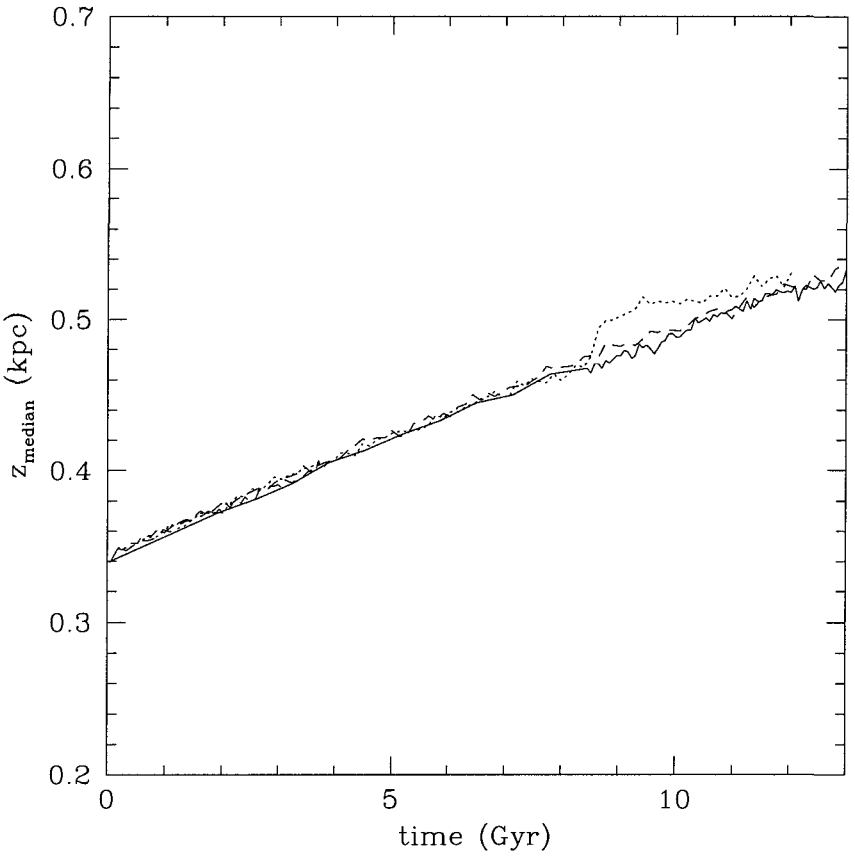

Figure 4.4: The evolution of z_{median} for the SCDM simulations. The solid line corresponds to the case without substructure, whereas the dotted (Halo 1) and the dashed (Halo 2) lines correspond to disks evolved in the presence of substructure.

What about the models with SCDM substructure? Figure 4.4 shows the evolution of the median disk thickness, z_{median}, for the isolated model (plotted with the full line) and for the model with SCDM substructure, Halo 1 (dotted line) and Halo 2 (dashed line), respectively. All models have been allowed to evolve for about a Hubble time[3] (i.e. $>$ 10 Gyr). The slow increase in the disk thickness in the case with SCDM

[3]Note that since the model with ΛCDM substructure does not display significant effects on

substructure is similar to that in the isolated model (and, implicitly, to the model with ΛCDM substructure).

As in the case of the run with ΛCDM substructure, we fit the vertical density distribution of the disk stars with the relation (4.1). Figure 4.5 shows the best fit values of the disk scale height z_0 as a function of distance R in the plane of the disk, and measured at a snapshot of $\sim$ 10 Gyr. The isolated model at this time is represented by empty squares and the two models with SCDM substructure are plotted with filled squares (Halo 1) and triangles (Halo 2), respectively. For comparison, we also show with hexagon symbols the results of the same data analysis for the isolated model at time $t \simeq 0$. Although the secular evolution of the disk is again evident for all models evolved for $\sim$ 10 Gyr, this time the effect of substructure stands out, particularly for the model with Halo 1 substructure. The figure also shows that, for the latter run, there is a clear increase of z_0 with the radial distance: For example, if we compare the disk scale height of the last radial bin $z_0(R \simeq 25\,\mathrm{kpc})$ measured after $\sim$ 10 Gyr in the model with Halo 1 substructure and the isolated model, we obtain a difference by a factor of $\sim$ 1.8. We note that the changes in scale height are mainly noticeable in the outer parts of the disk and that they are minor at $R \simeq R_\odot$. In the case of the run with Halo 2 substructure, the effect is less pronounced. The last radial bin seems to have a larger scale height, however, the errorbars are large (due to very few disk particles in the bin). Excluding the last data point, the scale height at the outskirts of the disk (i.e $z_0(R) = 21\,\mathrm{kpc}$) increases by a factor of at least 1.2 over the same period of 10 Gyr.

We therefore conclude that cold dark matter substructure is capable in certain

the disk structure, we have stopped the simulation after $\sim$ 4 Gyr. The simulations with SCDM substructure, however, have been run for a longer time ($\geq$ 10 Gyr). In that sense, a direct comparison between the runs with ΛCDM and SCDM substructure beyond $\sim$ 4 Gyr of evolution is not possible. However, this does not affect the interpretation of our results, since in this Section we are mainly interested in comparing the models with substructure with the isolated model, and not among themselves.

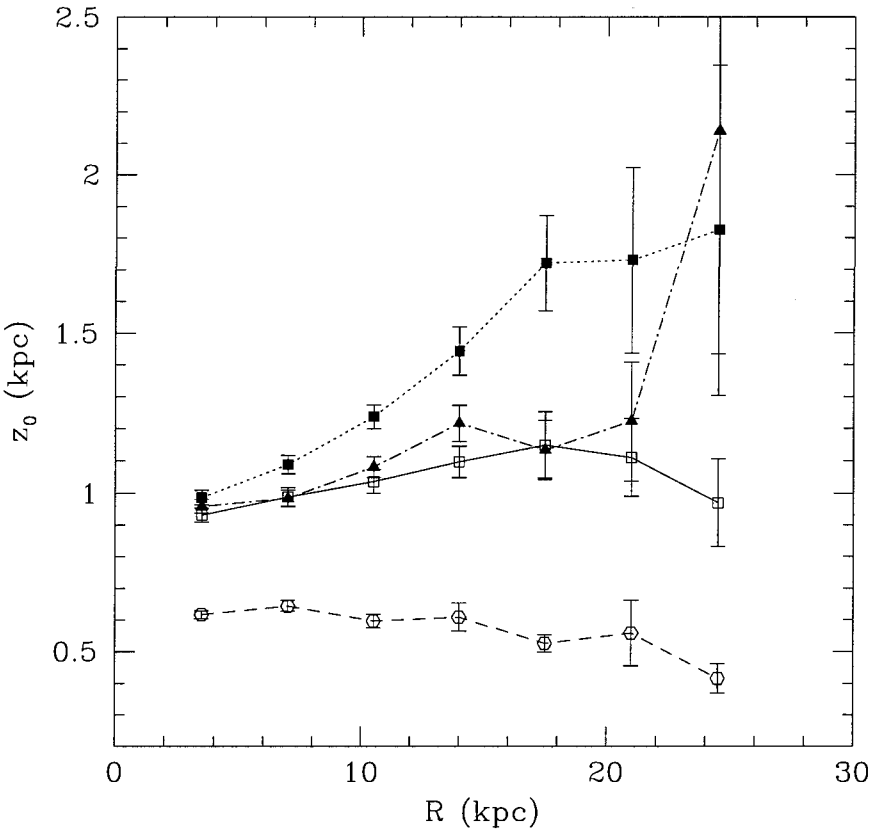

Figure 4.5: Scale height z_0 as a function of radial distance R in the plane of the disk. The empty squares denote the result for the isolated run, whereas the full squares and filled triangles denote the runs with SCDM substructure (Halo 1 and Halo 2, respectively), all data measured after $\sim$ 10 Gyr of evolution. For comparison, we show with hexagon symbols the results of the same data analysis for the isolated model at time $t \simeq 0$.

cases (eg. Halo 1 run) to induce significant changes in the outer vertical structure of the disk. A comparison with typical values for disk scale heights measured in the observations will be presented §4.3.

ii) The Velocity Dispersion:

Figures 4.6 and 4.7 show the changes in the velocity dispersion of disk stars mea-

sured in radial annuli of width $\simeq 0.74$ kpc, centered at various multiples of the solar radius: $R_\odot$, $2R_\odot$ and $3R_\odot$, respectively. In each panel, we follow the evolution of the corresponding vertical and radial components of the velocity dispersion, σ_z and σ_R, as well as of the total velocity dispersion, σ_{tot}, of the disk stars. The full line corresponds to the model evolved without substructure and the dashed line to the model with ΛCDM substructure. The figure shows that the presence of substructure does not influence any of the velocity dispersion components in the time span of ~ 4 Gyr monitored here, beyond the secular evolution induced by limited numerical resolution. The result is consistent with the results of previous analyses in the case of the ΛCDM run (Figures 4.2 - 4.3) and reflects the overall limited interaction between substructure and the disk in this particular case.

Figure 4.8 shows the three galactic components of the velocity dispersion, $(\sigma_R, \sigma_\phi, \sigma_z)$, measured at $R_\odot$, for the two halos in the SCDM simulation (Halo 1 - top panel; Halo 2 - bottom panel). The evolution of the velocity dispersion components with time is approximately the same with and without substructure. The (R, ϕ, z) velocity dispersion of disk particles at the solar circle grow[4], in all cases, from (31, 26, 27) km s^{-1} to approximately, (48, 37, 34) km s^{-1}, during a time interval of 10 Gyr.

Figure 4.9 shows the $(\sigma_R, \sigma_\phi, \sigma_z)$ components for the two simulations with SCDM substructure (Halo 1 - dotted line, top panel; Halo 2 - dashed line, bottom panel) and the isolated model (full line, both panels), calculated at a distance of $2R_\odot$ in the plane of the disk. Figure 4.10 shows similar information at $3R_\odot$. Over a period of ~ 10 Gyr, the (R, ϕ, z) velocity dispersions at twice the solar radius grow from (12.1, 8.6, 9.3) km s^{-1} to (28.3, 20.4, 20.8) km s^{-1} in the run with Halo 1 substructure, and to (26.5, 19.3, 16.8) km s^{-1} in the run with Halo 2 substructure. Over the same period of time, the (R, ϕ, z) components of $\sigma(3R_\odot)$ grow from (3.2, 2.3, 6) km s^{-1} to (23.2, 18.7, 12.7) km s^{-1} in the run with Halo 1 substructure, and to (14.8, 10.2, 14) km s^{-1} in the run with Halo 2 substructure.

[4]For comparison, the velocity ellipsoid of the thick disk is $(\sigma_R, \sigma_\phi, \sigma_z) = (46 \pm 4, 50 \pm 4, 35 \pm 3)$ km s^{-1} (Chiba & Beers 2000).

From the increase in the velocity dispersions we can infer two trends: *i)* at fixed radius, the variations are larger in the simulation with Halo 1 substructure in comparison with the simulation with Halo 2 substructure; *ii)* for both simulations with substructure, the increase is substantially larger further from the center of the disk.

It is also instructive to analyze the ratio σ_z/σ_R. The reason is that, if the CDM substructure is the mechanism responsible for the heating levels detected observationally in the disk, then it should be able not only to increase the velocity dispersion in all three components, but also to reproduce the measured relative ratios. For example, the observations collected in the solar neighborhood in our Galaxy suggest that the velocity dispersion does not increase the same in all directions. The velocity dispersion is seen to be larger in the Galactic plane than in the vertical direction. The measured ratio σ_z/σ_R in the solar neighborhood is also very well constrained: 0.5 ± 0.1 (Dehnen & Binney 1998; Hänninen & Flynn 2002).

What do our simulations show? At the solar radius, in both models with substructure, the ratio σ_z/σ_R decreases from the initial value of 0.87 to about 0.71, over a period of 10 Gyr. However, this change is identical with that obtained in the simulation without substructure; therefore the change is due to the effect of the small inhomogeneities in the galactic potential (*i.e* 2-body interactions) and not caused by substructure.

The situation changes at larger radii. Over the same period of $\sim$ 10 Gyr, the ratio σ_z/σ_R at $2R_\odot$ decreases from 0.77 to 0.73 in the simulation with Halo 1 substructure, and from 0.77 to 0.63 in the simulation with Halo 2 substructure. For reference, in the isolated model, the σ_z/σ_R ratio remains approximately constant, at 0.77. Similarly, at $3R_\odot$, the ratio σ_z/σ_R decreases from 1.87 to 0.55 in the simulation with Halo 1 substructure, and from 1.87 to 0.94 in the simulation with Halo 2 substructure. In the model without substructure, the σ_z/σ_R ratio drops to 0.71.

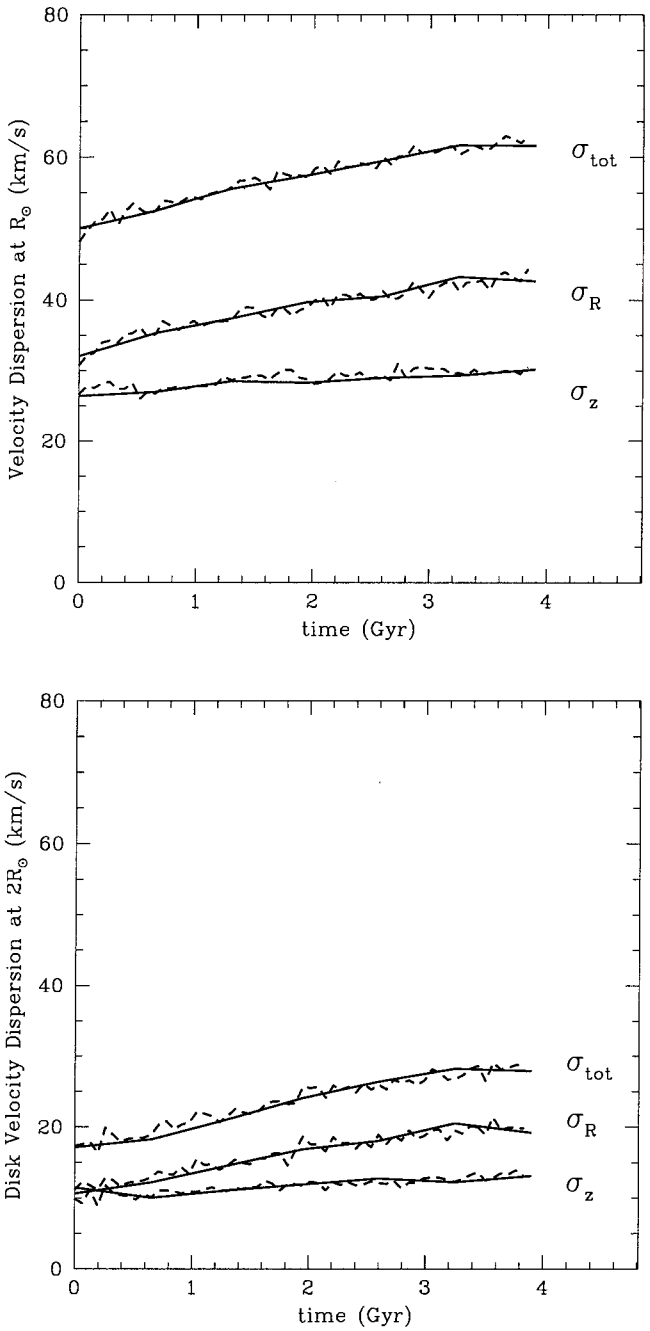

Figure 4.6: The evolution of the velocity dispersions measured at $R_\odot$ (top panel) and $2R_\odot$ (bottom panel), respectively. The full line corresponds to the model without substructure, and the dashed line to the model with ΛCDM substructure. Note that, for the time interval analyzed here, the number of snapshots for the model without substructure is smaller than that available for the model with ΛCDM substructure (hence the apparent "smoothness" of filled lines).

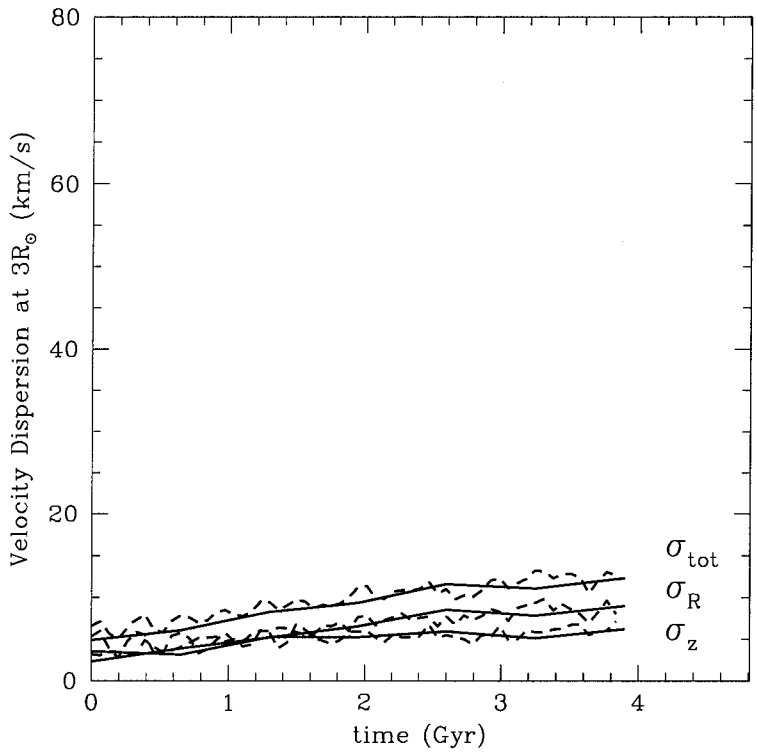

Figure 4.7: The evolution of the velocity dispersions measured at $3R_\odot$. The full line corresponds to the model without substructure, and the dashed line to the model with ΛCDM substructure.

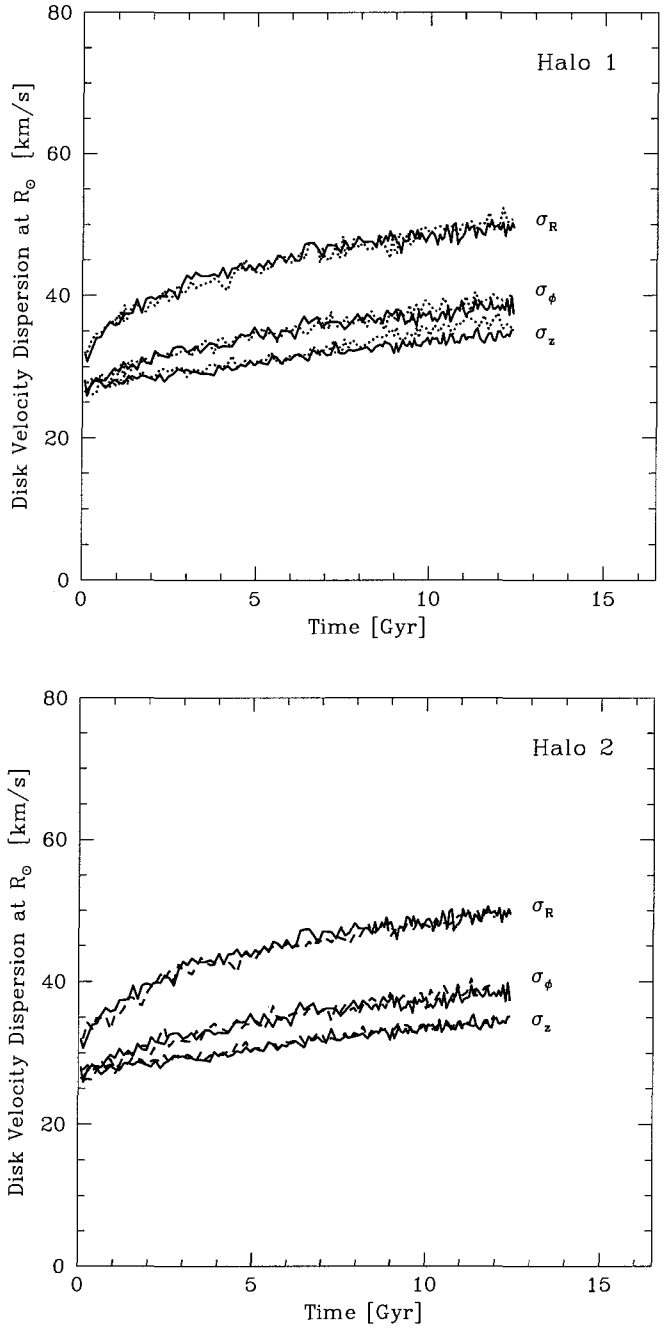

Figure 4.8: The three galactic components of the velocity dispersion, $(\sigma_R, \sigma_\phi, \sigma_z)$, measured at $R_\odot$, for the two runs with SCDM substructure (Halo 1 - top panel; Halo 2 - bottom panel). Full lines correspond to the isolated galaxy model, in both panels; dotted lines to Halo 1; and dashed lines in bottom panel, to Halo 2 substructure model.

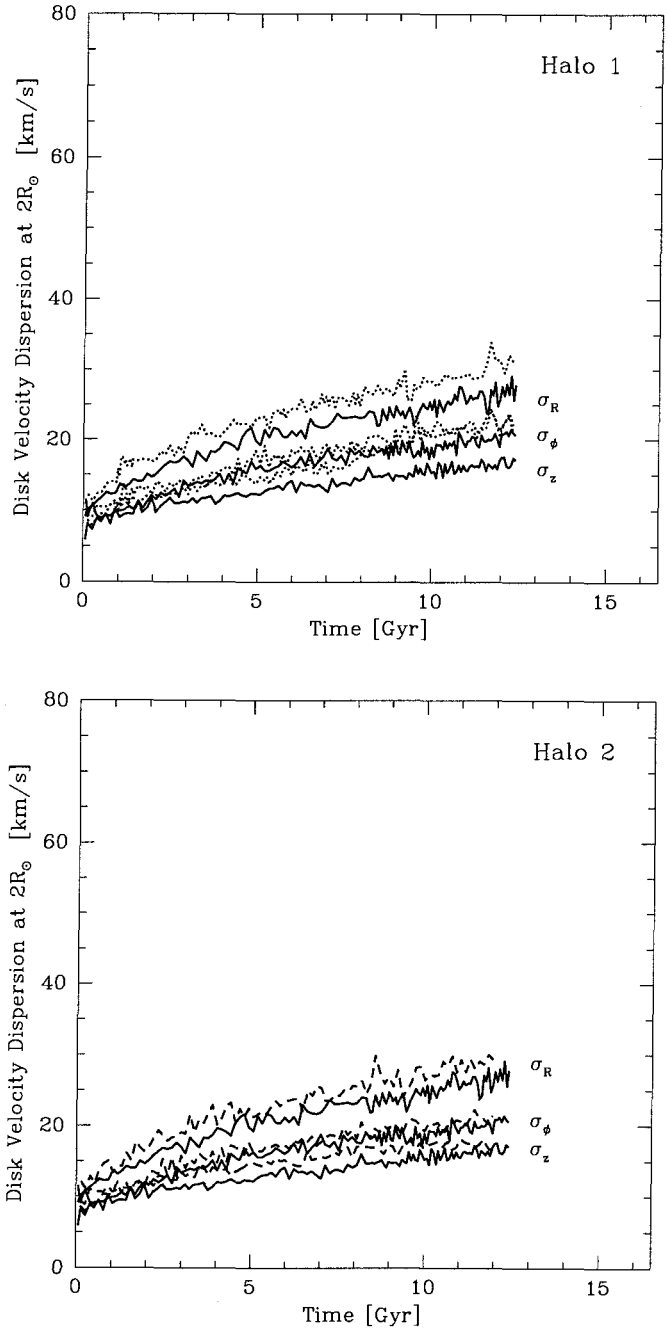

Figure 4.9: Velocity dispersion $(\sigma_R, \sigma_\phi, \sigma_z)$ at $2R_\odot$ for the two runs with SCDM substructure (Halo 1 – top panel; Halo 2 – bottom panel). The full, dotted and dashed lines are as in Figure 4.8.

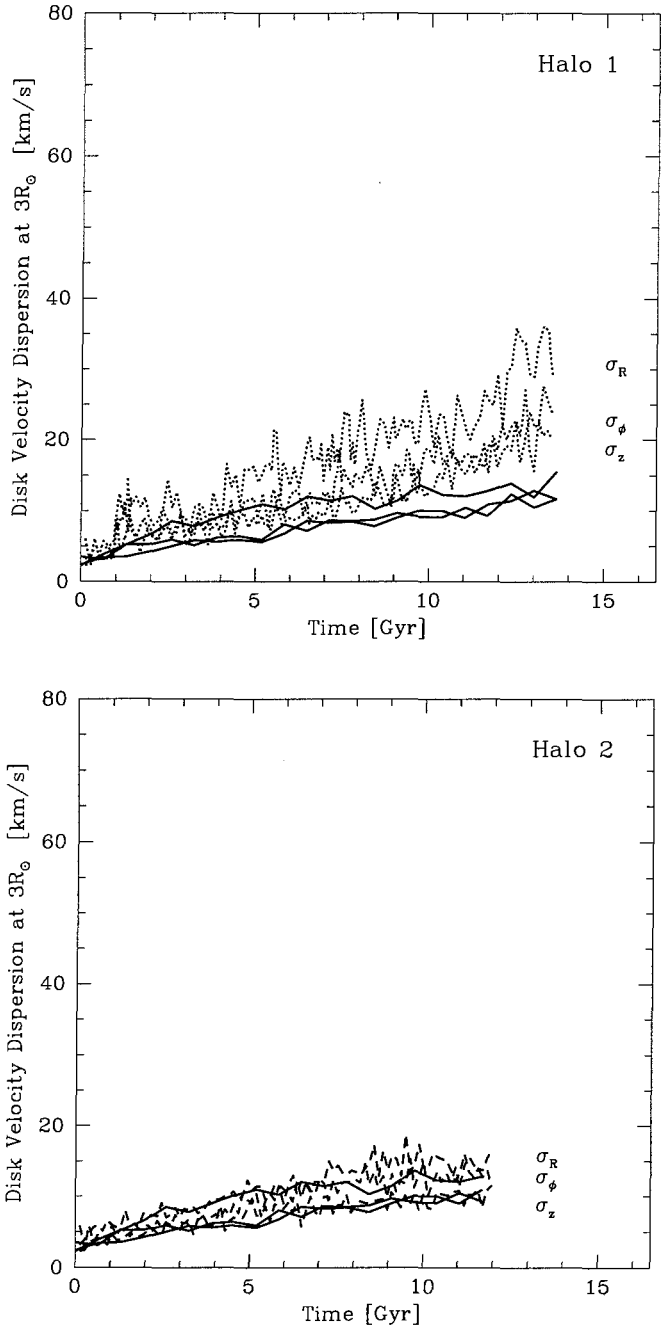

Figure 4.10: Velocity dispersion $(\sigma_R, \sigma_\phi, \sigma_z)$ at $3R_\odot$ for the two runs with SCDM substructure (Halo 1 – top panel; Halo 2 – bottom panel). The full, dotted and dashed lines are used as in Figure 4.8.

Our results predict that the substructure does not heat the disk at the same rate in all galactic components. Specifically, this heating mechanism seems to be more efficiently in the radial direction than in the vertical one (the σ_z/σ_R ratio decreases more sharply when the interaction with the CDM substructure is more pronounced).

iii) The Toomre Parameter, Q

The Toomre parameter, Q, provides additional information about the changes in the structure of the disk (recall equation (3.5)). The evolution of the Q parameter will include the effect of the increase of the radial excursions of the disk stars from their circular orbits (measured by σ_R) with time, and, when applicable, the effect of larger fluctuations in the gravitational field caused by close passages of satellites.

Figure 4.11 shows the Toomre parameters $Q(R_\odot)$ (top panel) and $Q(2R_\odot)$ (bottom panel) for the run with ΛCDM substructure. As with previous parameters (eg. z_{median}, σ), the secular evolution of the disk induces a monotonous increase of Q with time. Substructure in the ΛCDM halo leaves no detectable signatures in the evolution of $Q(R_\odot)$, although it appears to have a minor effect in the evolution of $Q(2R_\odot)$.

Figure 4.12 shows the Toomre parameter $Q(R_\odot)$, for the two SCDM satellite realizations (Halo 1- top panel; and Halo 2 - bottom panel) in comparison with the case of the isolated galaxy model. From this plot we can see that, again, neither substructure distribution has a substantial impact on the radial structure of the disk near the solar radius.

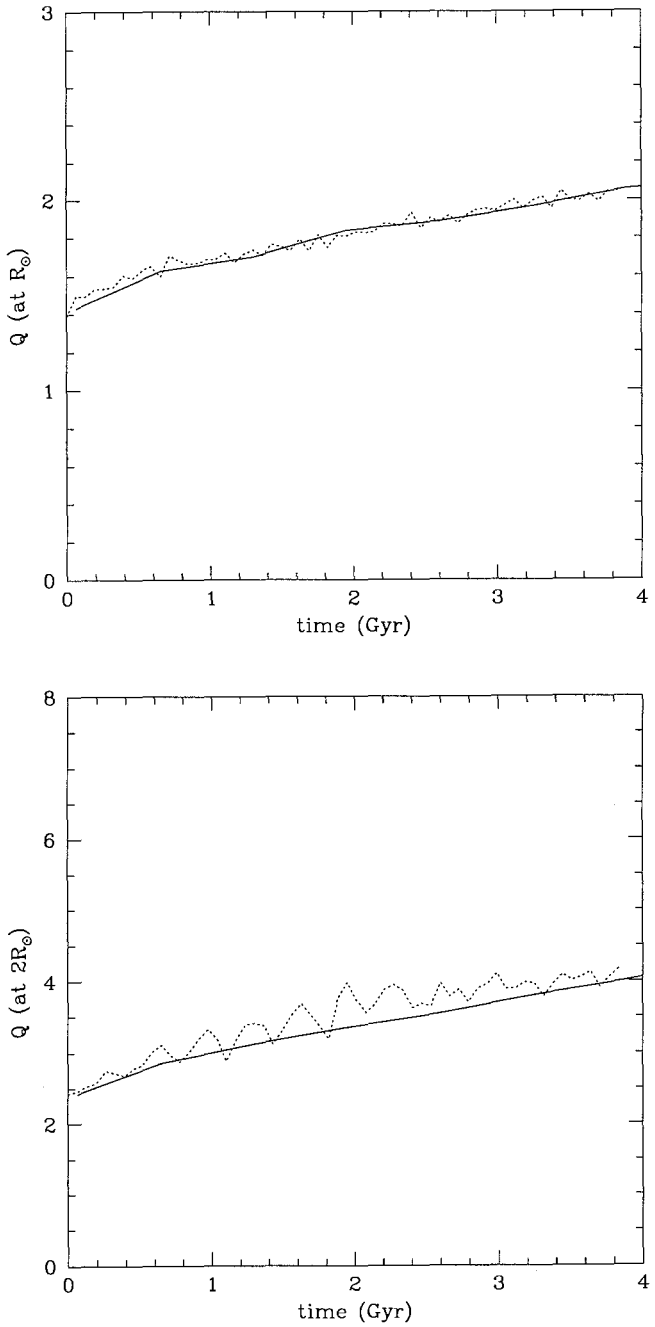

Figure 4.11: Toomre parameters $Q(R_\odot)$ (top panel) and $Q(2R_\odot)$ (bottom panel) for the run with ΛCDM substructure. The full line corresponds to the isolated galaxy model, the dotted line to the model with ΛCDM substructure.

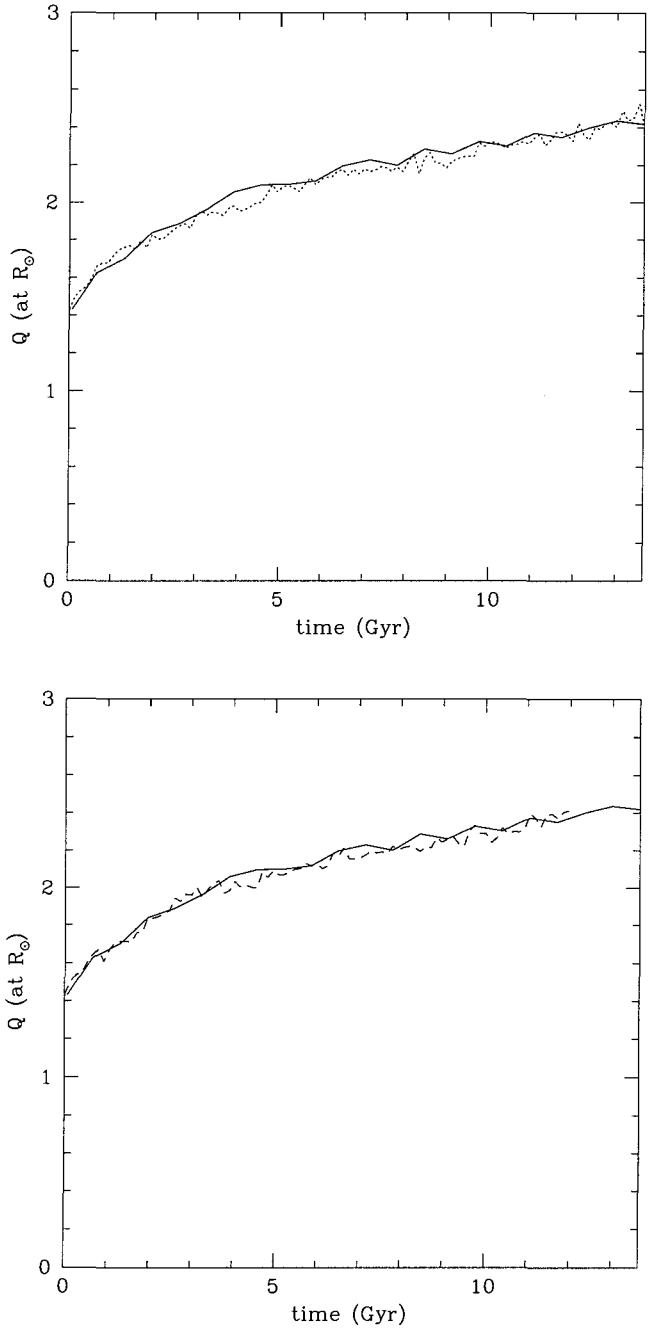

Figure 4.12: Toomre parameter $Q(R_\odot)$ for the two runs with SCDM substructure. The full line corresponds to the isolated galaxy model, the dotted line (top panel) to Halo 1 (SCDM) model, and the dashed line (bottom panel), to Halo 2 (SCDM) model.

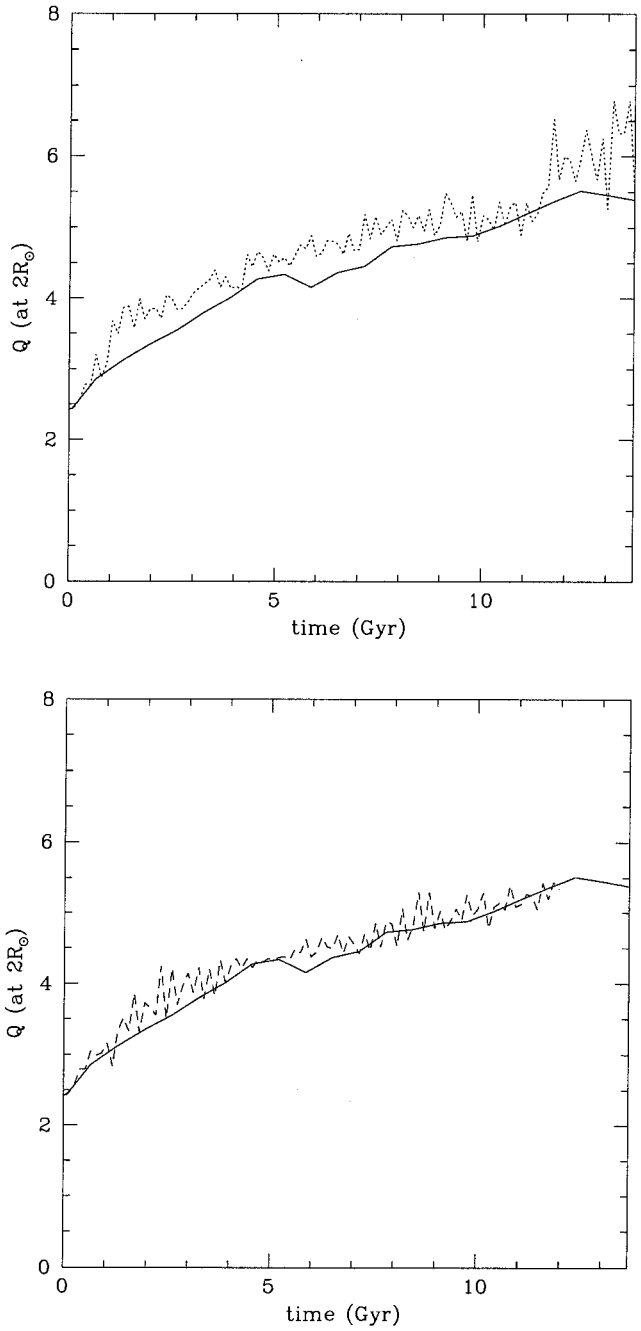

Figure 4.13: Toomre parameter $Q(2R_\odot)$ for the two runs with SCDM substructure (Halo 1 – top panel; Halo 2 – bottom panel). The full, dotted and dashed lines are used as in Figure 4.12.

The evolution of $Q(2R_\odot)$ for the two SCDM substructure realizations is shown in Figure 4.13. Over a period of $\sim$ 10 Gyr, $Q(2R_\odot)$ increases by about 10% in the case of Halo 1 above the corresponding value for the model without substructure. However, the increase is not monotonous, but it happens at early times ($t \simeq 1$ Gyr) as a result of a strong tidal interaction with a satellite[5] of a mass of $\sim 8.6 \times 10^{10} M_\odot$ passing near the disk at a distance of about 30 kpc. The next significant disturbance of the $R \sim 2R_\odot$ region of the disk due to the interaction with Halo 1 substructure occurs around $t \sim 11$ Gyr. The effect of this interaction[6] is also seen in the top panel of Figure 4.13. Over the same period of $\sim$ 10 Gyr, the disk embedded within Halo 2 sees little effect in its $Q(2R_\odot)$ parameter – a fact which can be related, again, to the fact that in this case the orbits of satellites do not come close to the disk.

Figures 4.12 and 4.13 suggest that the Toomre parameter responds more quickly to the tidal interactions than the parameters that measure the vertical changes (even when the vertical changes are minimal, the radial ones can be important).

Conclusions of our analysis:

- In the case of the simulation with ΛCDM substructure, the overall changes in the structure of the stellar disk are negligible. Parameters such as z_{median}, σ or Q show no significant changes in the presence of the substructure. A slight variation may be seen, however, in the evolution of the disk local scale height z_0 at the outskirts of the disk, however this measurement has a large uncertainty.

We caution that the null results obtained in the case with ΛCDM substructure should not necessarily be interpreted as a general prediction for the ΛCDM cosmology, but rather a consequence of the properties of the higher mass end of the CVF (recall Figure 2.10) in our particular simulation. In the case of the substructure realization analyzed here, we can confidently relate the lack of disk heating with the fact that no

[5]The orbital evolution of this satellite (denoted GS2) will be analyzed in detail in §5.1.

[6]Around this time, the disk interacts with another satellite (denoted GS1), whose orbital motion will also be analyzed in detail in §5.1.

massive satellites are orbiting in the vicinity of the disk. Further studies, employing a statistical set of ΛCDM substructure realizations, are required in order to determine whether the orbital parameters of the satellites in our simulation are representative for the ΛCDM cosmology as a whole.

- In the case of the simulations with SCDM substructure, there are a few notable changes in the evolution of several physical parameters of the disk (eg. z_0, σ, Q). These occur mainly in the run with Halo 1 substructure and at large radii (typically, beyond $2R_\odot$). The Halo 2 substructure has less influence on the physical parameters of the disk than the Halo 1 substructure.

In the following, we analyze in more detail the simulation which displays the strongest effect of the dark matter substructure on the disk, i.e the Halo 1 run. Since in the other two runs the effects of substructure are roughly of the same order or even below the numerical noise in the simulations, we do not include these cases in our analysis. However, we caution that we do not seek to describe the case of Halo 1 distribution as being "typical" for the CDM models (as we have cautioned before, a statistical set of simulations is needed to determine whether this is the case). Rather, our goal is that by choosing this particular example to understand the mechanism by which dark matter substructure contributes to the heating of the Galactic disk.

4.2 Comparison with the Heating Rate in the Galaxy

The purpose of this Section is to compare the heating rate caused by CDM substructure with the dynamical heating rate measured in the Milky Way disk. First, we introduce the physical parameters which are generally used to quantify the heating rate in the disk, i.e. the diffusion coefficients. We then review the observational evidence of the heating of the Galactic disk and present empirical fits to the age-velocity dispersion relation. We describe the difficulties encountered in analyzing the observational data in a self-consistent way. A good understanding of the limitations of both the observational results and theoretical models is essential when making compar-

isons between the two. Finally, we assess the diffusion coefficients due to encounters between the disk stars and the CDM substructure in the Halo 1 (SCDM) run and investigate whether they are of the same magnitude as the coefficients inferred from observations.

4.2.1 Diffusion Coefficients

The analytical approach for determining the heating rate due to substructure is similar to that which is generally applied to other stochastic heating mechanisms. Regardless of the nature of the perturber, one can understand the behavior of the system by recalling several basic principles of collision theory (eg. Chandrasekar 1960). The first step is to write the equation which determines the rate of change of the distribution function $f(\mathbf{x}, \mathbf{v}, t)$, also known as the Boltzmann equation. Thus, in the case when collisions are present, the overall change in the distribution function is non-zero:

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla f - \nabla\Phi \cdot \frac{\partial f}{\partial \mathbf{v}} \equiv \Gamma[f] \neq 0, \tag{4.2}$$

where we introduced the new parameter $\Gamma[f]$, also known as the *collision term.*
Let $\Psi(\mathbf{w}, \Delta\mathbf{w})\, d^3\mathbf{w}\, dt$ be the probability that a star located in the six-dimensional phase-space $\mathbf{w} = (\mathbf{x}, \mathbf{v})$ is scattered (or "diffused" in the phase-space) through random collisions, in a volume $d^3\Delta\mathbf{w}$ and time interval dt. We can then measure the change of the quantity $\Delta\mathbf{w}$ with the following integral over the phase-space:

$$D(\Delta w_i) \equiv \int \Delta w_i \Psi(\mathbf{w}, \Delta\mathbf{w}) d^3\mathbf{w}, \tag{4.3}$$

where $D(\Delta w_i)$ is also known as the first order *diffusion coefficient.*
Similarly, one can also derive the second-order diffusion coefficients (which measure the change of both $\Delta\mathbf{w_i}$ and $\Delta\mathbf{w_j}$):

$$D(\Delta w_i \Delta w_j) \equiv \int \Delta w_i \Delta w_j \Psi(\mathbf{w}, \Delta\mathbf{w}) d^3\mathbf{w}, \tag{4.4}$$

The first and second order coefficients are, in general, sufficient to describe the diffusion of stellar orbits in the phase-space, since higher order ones are much smaller and usually negligible.

The collision term $\Gamma[f]$ can be then calculated by expanding it in a Taylor series and truncating the sum after the first and second-order terms (BT87, p507):

$$\Gamma[f] = -\sum_{i=1}^{6} \frac{\partial}{\partial w_i}[f(\mathbf{w})D(\Delta w_i)] + \frac{1}{2}\sum_{i=1}^{6}\sum_{j=1}^{6} \frac{\partial^2}{\partial w_i \partial w_j}[f(\mathbf{w})D(\Delta w_i \Delta w_j)], \quad (4.5)$$

an approximation which is also known as the Fokker-Planck equation (BT87, p506 - 510).

In theory, the value of the diffusion coefficient can be obtained by solving the Fokker-Planck equation (4.5). In practice, this cannot be done, since the distribution function $f(\mathbf{w}, t)$ is difficult to infer from observations, and remains still unknown. Therefore, the Fokker-Planck equation is usually simplified by making some additional assumptions.

For example, since the collisions are assumed to be random, the average of the velocity changes Δv_i over a long time will be zero, $< \sum \Delta v_i >= 0$. However, the sum of the squares of the velocity is non zero and positive, and in more general terms, will depend of time. Therefore, the first order diffusion coefficient $D(\Delta v_i)$ in the Fokker-Plack equation (4.5) vanishes. In the so called "local approximation", we can assume that the encounters change only the velocity of the stars and not its position. Thus, all the terms in the Fokker-Planck equation which include a change in the position $\Delta \mathbf{x}$, such as $D(\Delta x_i)$, $D(\Delta x_i \Delta x_j)$, $D(\Delta x_i \Delta v_j)$, also become zero. In the end, we are left only with the second-order terms of the form $D(\Delta v_i \Delta v_j)$, of which only the non-orthogonal ones are non-zero (*i.e.* $D(\Delta v_i \Delta v_i) \neq 0$).

Furthermore, we can assume for simplicity that the perturbations applied to the star are purely irregular perturbations, *i.e* there are no regular perturbations along the orbit (the so-called "force-free diffusion" approximation). In this case, the equation (4.5) can be separated completely in the three galactic components (Wielen 1977).

With these approximations, the Fokker-Planck equation (4.5), written in the usual galactic coordinates (**U, V, W**), simplifies to:

$$\Gamma[f] = \frac{\partial f}{\partial t} = \frac{1}{2}[\frac{\partial}{\partial U}(D_U \frac{\partial f}{\partial U}) + \frac{\partial}{\partial V}(D_V \frac{\partial f}{\partial V}) + \frac{\partial}{\partial W}(D_W \frac{\partial f}{\partial W})] \tag{4.6}$$

4.2.2 The Age-Velocity Dispersion Relation (AVR)

With the above assumptions, we can concentrate only on the second-order diffusion coefficient and, in particular, on its velocity-only component, $D(\Delta v_i \Delta v_i)$. We can therefore determine $D(\mathbf{v}, t)$ empirically from the velocity information of a number of stars.

Thus, given a sample of n stars, the equation (4.4) can be approximated as:

$$< \sum_{i=1}^{n} (\Delta v_i)^2 >= D \cdot \Delta t, \tag{4.7}$$

where D is the second order diffusion coefficient. In other words, the velocity dispersion of stars varies roughly as:

$$\frac{d(\sigma^2)}{dt} = D(\mathbf{v}, t) \equiv D, \tag{4.8}$$

which means that the velocities of stars scatter with an ever increasing dispersion around the initial value.

From observations one can derive the rate of change of the velocity dispersions in a given sample of stars and, therefore, determine empirically the diffusion coefficient, $D(v, t)$. Additionally, we can guide the search with the help of a few physically motivated arguments.

Thus, we can distinguish three simple cases:

- the diffusion coefficient $D(v,t)$ is constant: $D \equiv D_0$.

  Then, by integrating equation (4.8), we obtain:

$$\sigma^2 = \sigma_0^2 + D_0\tau, \tag{4.9}$$

  where σ_0 is the velocity dispersion with which the star is born and τ represents the time-span between the present and the time of the formation of the star, $\tau = t_p - t_f$.

- the diffusion coefficient $D(v,t)$ depends on velocity only.

  This case is the most physically motivated, being supported by simple arguments from collision theory. The diffusion coefficient is expected to scale proportionally with the length of the encounter, which in turn, is inversely proportional to the speed of the encounter, $\sim {V_{en}}^{-1}$ (Chandrasekar 1960). Thus, the diffusion coefficient is inversely proportional to the relative velocity of the encounter between the star and the perturber, $D = D_0/V_{en}$. If the encounters are fast, it is easy to see that the greater the velocity of the encounter, the smaller the effect the encounter has on the star (this is also known as the "impulse approximation" ; BT87, p. 433).

  The integration of the equation (4.8) in this case, yields:

$$\sigma^3 = \sigma_0^3 + \frac{3}{2} D_0 t \tag{4.10}$$

  consistent with heating by collisions between stars and molecular clouds (Spitzer & Schwarzschild 1951; 1953).

- the diffusion coefficient depends on both velocity and time, $D(\mathbf{v},t)$.

A physically motivated choice in this case, would be $D = \gamma(t)/v$, where $\gamma(t)$ is a parameter that contains the time-dependence of the diffusion. Wielen (1977) assumes that $\gamma(t)$ decreases with time, as a result of the "smoothing" of the gravitational field in the disk with time (assuming a more turbulent environment at earlier ages)[7]. He uses for γ the exponential decay form: $\gamma(t) = \gamma_p \cdot e^{-\frac{(t-t_p)}{T_\gamma}}$, where T_γ is the decay time of γ, and the index "p" corresponds to values at the present time, and derives:

$$\sigma^3 = \sigma_0^3 + \frac{3}{2}\gamma_p T_\gamma [e^{(\tau/T_\gamma)} - 1] \tag{4.11}$$

4.2.3 The AVR in the Galactic Disk

Information about the AVR in the Galactic disk has been accumulating for nearly five decades now, starting with the first studies of kinematics of different populations of stars in the solar neighborhood (eg. Roman 1954). Since then, the data have expanded to include larger samples of stars and its accuracy has improved as stellar ages and distances have become available. In particular, the data collected by the Hipparcos satellite are now sufficiently accurate to constrain the age-velocity dispersion relation (AVR) of stars in the solar neighborhood. Difficulties in measuring the kinematics of stars at large distances prevent us from extending the measurements significantly beyond our immediate location in the Galaxy. Therefore, the reader should bear in mind that the solar neighborhood is the only real constraint we have on the heating processes operating in the disk.

In the following, we will summarize the results derived from the most recent samples of nearby stars and the various studies aimed at finding the AVR that can best fit the data.

[7]This is a reasonable assumption if one considers only the encounters with other objects in the disk, such as other stars or molecular clouds. This assumption becomes questionable, however, if one takes into account the external encounters, which may create large perturbations in the gravitational field over time (eg., with satellite galaxies).

One of the first analyses of the diffusion coefficients and of the age-velocity dispersion relation in the Galaxy was presented by Wielen (1977). The author compiled from the McCormick catalogue a sample of $\sim$ 1000 nearby stars, most of them being K and M dwarfs. This sample has been later extended by Fuchs et al. (2001) to include about 3000 stars. The data shows a general trend in which the total velocity dispersion of stars varies from $\sim$ 20 km s^{-1} for stars of ages $\sim 10^7$ yr, to about $60-80$ km s^{-1}, for stars of ages $10 - 12$ Gyr. As explained in §4.2.2, one can find, either empirically or guided by some physically motivated arguments, an AVR that reproduces the observed increase of the velocity dispersion with stellar age. Wielen (1977) and Fuchs et al. (2001) show that their data can be well fitted with the relation: $\sigma \sim t^{1/2}$ (which is identical to the function (4.9)). Wielen (1977) finds that the best fit of this form yields the values: $\sigma_0 = 10$ km s^{-1} and $D_0 = 600$ km^2 s^{-2} Gyr^{-1}, with a $\sim$ 20% uncertainty in both parameters. The author also fits other functions to the age-velocity dispersion data. The best fit with function (4.10) yields the values: $\sigma_0 = 10$ km s^{-1} and $D_0 = 1.4 \times 10^{-5}$ (km s^{-1})^3/yr; whereas for the best fit with function (4.11), he obtains the set of parameters: $\sigma_0 = 10$ km s^{-1}, $\gamma_p = 1.1 \times 10^{-5}$ (km s^{-1})^3/yr and $T_\gamma = 5 \times 10^9$ yr.

The more recent measurements of stellar parallaxes and kinematics by the Hipparcos satellite have prompted a re-assessment of the AVR. For example, Binney, Dehnen & Bertelli (2000) have analyzed 11865 main sequence and subgiant stars from the Hipparcos data set. Based on this sample of stars, Dehnen & Binney (1998) and Binney, Dehnen & Bertelli (2000) argue that the $\sigma \sim t^{1/3}$ function is a better fit to the data than $\sigma \sim t^{1/2}$. More recently, Nordström et al. (2004) used a larger sample of 16682 F and G dwarf stars, of which 14139 stars have complementary kinematic information. These authors also obtain heating exponents of about 1/3: $\sigma_U \sim t^{0.31}$, $\sigma_V \sim t^{0.34}$, $\sigma_W \sim t^{0.47}$ and $\sigma_{tot} \sim t^{0.34}$.

Hänninen & Flynn (2002) propose a purely empirical power-law function of the general form:

$$\sigma(t) = \sigma_0(1 + \frac{t}{\tau})^{\alpha}, \tag{4.12}$$

where α is the "heating index". Hänninen & Flynn (2002) have compiled several data samples from the literature and fitted them, in a self-consistent manner, with the above formula. We will take advantage of this large set of observational samples in order to compare to our numerical results.

Table 4.1: Measured values of the heating index α in equation (4.12), as compiled from observations by Hänninen & Flynn (2002). The last entry corresponds to the result of the power law fit of Nordström et al. (2004).

α	Data sample
0.61 ± 0.04	Wielen (1977)
0.59 ± 0.04	Fuchs et al. (2001)
0.47 ± 0.002	Edvardsson et al. (1993)
0.26 ± 0.008	H. Rocha Pinto (priv. comm.)
0.45 ± 0.04	Holmberg (2001); all stars
0.33 ± 0.03	Holmberg (2001); thick disk removed
0.33 ± 0.03	Binney et al. (2000)
0.34 ± 0.05	Nordström et al. (2004)

Table 4.1 reproduces the best fit results obtained by Hänninen & Flynn (2002) for different samples of stars. In addition to a re-evaluation of the samples discussed above, the Table includes the sample of Edvardsson et al. (1993) containing 189 F and G stars; the sample of Holmberg (2001) containing 1486 post-Hipparcos F and G stars; and the sample of Rocha-Pinto with 425 stars, mainly of F, G and K spectral type. For reference, in the last entry of the table we have also added the results of a similar power-law fit done by Nordström et al. (2004) using a sample of 14139 F and G stars with kinematic information. The best fits of the form (4.12) for the samples of stars compiled by Hänninen & Flynn (2002) constrain the heating index α in the

range $0.3 - 0.6$. These constraints are useful for identifying the heating mechanism in the disk. A study of how CDM substructure and other heating mechanisms in the disk influence the rate in velocity dispersion of stars will be made in the following sections.

Before that, we make a final note that the above constraints on the AVR are valid only for the solar neighborhood. At present, we do not know if the AVR has the same or different characteristics at other locations in the disk. Future advances in the observations will enable us to probe larger distances in the disk and to answer this question. On the other hand, numerical simulations do not suffer the same limitation as the observations. Therefore, by analyzing the behavior of the velocity dispersion at different radial distances in our simulations, we can make predictions regarding the signature of the CDM substructure on the evolution of the AVR.

4.2.4 The Contribution of CDM Substructure to the Heating Rate

This subsection is concerned with determining the contribution of substructure to the heating rate in the disk. We shall concentrate only on the case in which the heating generated by substructure is clearly seen in the simulation, namely the Halo 1 (SCDM) run. In this case, we want to determine what is the most appropriate functional form for the rate of change of velocity dispersion of disk stars. The numerical results are then compared with information about the heating rate inferred from the observational fits introduced earlier in this Section.

Figure 4.14 shows the evolution of σ_{tot} measured at several locations in the disk ($R_\odot$, $2R_\odot$, $3R_\odot$) in the case of the simulation with Halo 1 substructure. Over a period of about 10 Gyr, the total velocity dispersion at the solar radius grows from $\sim$ 50 km s^{-1} to $\sim$ 70 km s^{-1}. For comparison, the dot-dashed line in Figure 4.14 shows the best fit of the same form obtained by Nordström et al (2004) for the observations in the solar neighborhood. The result suggests that the effects of CDM substructure

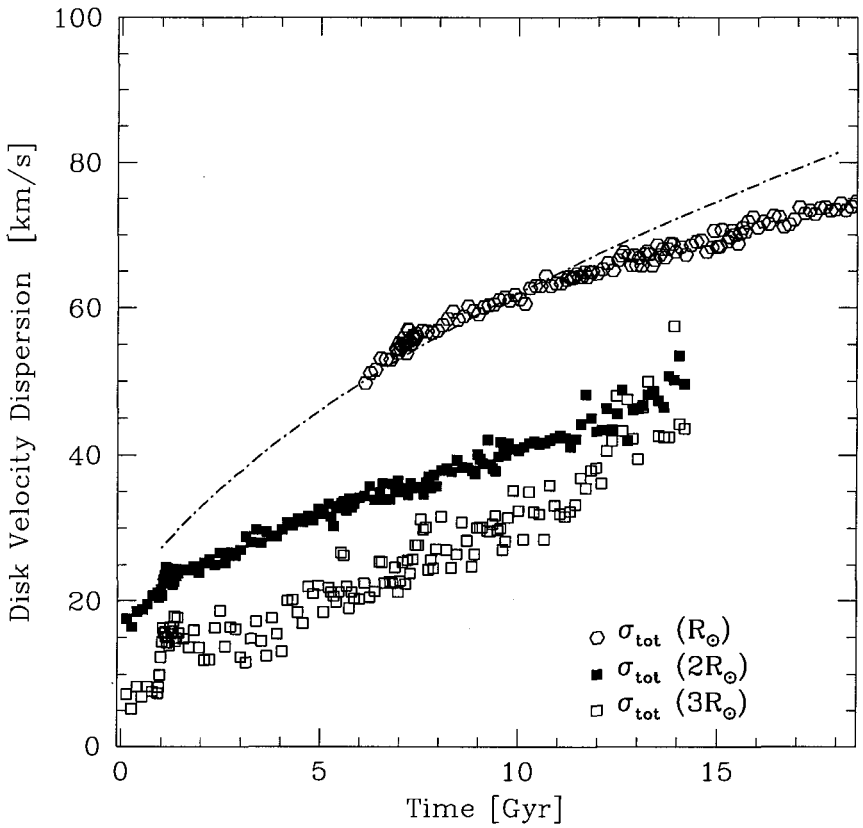

Figure 4.14: Evolution of σ_{tot} measured at several locations in the disk ($R_\odot$ - hexagons; $2R_\odot$ - filled squares; $3R_\odot$ - empty squares) in the case of the simulation with Halo 1 substructure.

may be insufficient to explain the heating rate in the solar neighborhood.[8]

Over a period of about 10 Gyr, the total velocity dispersion at twice the solar radius, $\sigma_{tot}(2R_\odot)$, grows from $\sim$ 20 km s^{-1} to $\sim$ 40 km s^{-1}. Similarly, over a period of about 10 Gyr, the total velocity dispersion at three times the solar radius, $\sigma_{tot}(3R_\odot)$, grows from $\sim$ 5 km s^{-1} to $\sim$ 30 km s^{-1}. This implies that parameters $\sigma_{tot}(2R_\odot)$

[8]Since the other two simulations with substructure (ΛCDM Halo and Halo 2 SCDM) show even less heating than the simulation analyzed here, this value will represent an upper limit to our set of simulations.

and $\sigma_{tot}(3R_\odot)$ have higher rates than $\sigma_{tot}(R_\odot)$, and than Nordström et al (2004) and Wielen (1977) fits (best fits of the form $\sigma \sim t^\alpha$ to these data give values of α larger than 1/2). Although there are currently no velocity dispersion measurements at large distances in the disk to compare our numerical results to, it is an instructive exercise to quantify the effect of substructure in these regions of the disk. Quantifying the effect where it is most important, i.e. at the outskirts of the disk, allows us to understand the way the mechanism of disk heating by dark matter substructure operates and provides testable predictions for future observations.

However, as we have previously demonstrated that two body interactions can have a significant contribution to the overall heating rate, it is important to separate the effect of substructure from the secular evolution of the disk. In doing this, we have removed the evolution of σ_{tot} in the isolated model from the evolution of σ_{tot} in the model with Halo 1 substructure, at each of the sampled radii where the effect of substructure is important, i.e. at $2R_\odot$ and $3R_\odot$: $\Delta\sigma_{tot} \equiv \sigma_{tot}^{with\ substr} - \sigma_{tot}^{no\ substr}$. This new parameter contains only the effect of substructure. The result is shown in Figure 4.15, which includes in the top panel (with full squares) the evolution of $\Delta\sigma_{tot}(2R_\odot)$ and in the bottom panel (with empty squares), the evolution of $\Delta\sigma_{tot}(3R_\odot)$.

In a similar fashion as in the case of the observational AVR analysis, we choose to fit the $\Delta\sigma_{tot}$ data with the function:

$$\Delta\sigma = (\Delta_0^2 + D' \cdot t)^{1/2}, \tag{4.13}$$

where Δ_0 is a constant which has units of km s^{-1} and D' is a coefficient similar to the diffusion coefficient introduced before, this time describing the rate of $\Delta\sigma_{tot}$. This function is similar to the relation (4.9), which was found by Wielen (1977) to be a good fit to the observational σ_{tot} data. The best fits of this form are also shown in Figure 4.15, and have the following parameters:

$$\Delta\sigma_{tot}(2R_\odot) = [(-0.065)^2 + 2.254 \cdot t]^{1/2}$$

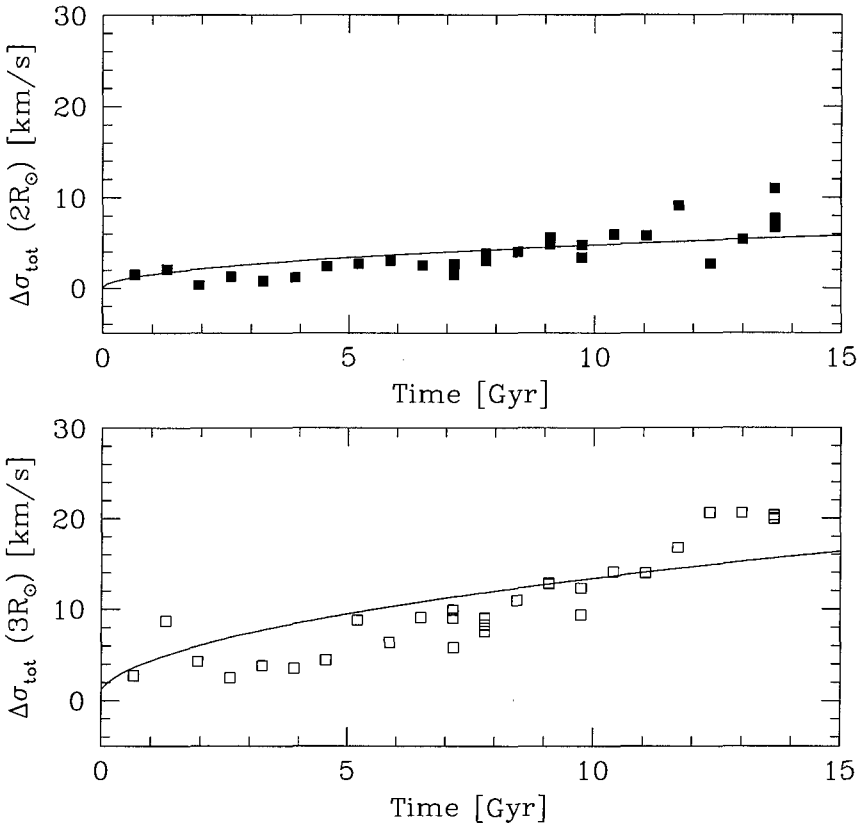

Figure 4.15: The evolution of $\Delta\sigma_{tot}(2R_\odot)$ (top panel; filled squares) and $\Delta\sigma_{tot}(3R_\odot)$ (bottom panel; empty squares) in the simulation with Halo 1 substructure. In both panels, the curves represent the best fits of the form (4.13) to the data (see text for details).

and

$$\Delta\sigma_{tot}(3R_\odot) = [(-1.254)^2 + 17.686 \cdot t]^{1/2}$$

However, the standard errors of these fits are quite large: for example, the D' values have standard errors of $\pm 24.93\%$ (for $2R_\odot$ case) and $\pm 17.43\%$ (for $3R_\odot$ case), respectively. This suggests that equation (4.13) might not provide a good fit to the $\Delta\sigma_{tot}$ data[9]. Therefore, we choose to investigate whether a better fit could be obtained

[9]We note that we also found that Wielen's relation (4.9) is not a good fit in the case of the σ_{tot}

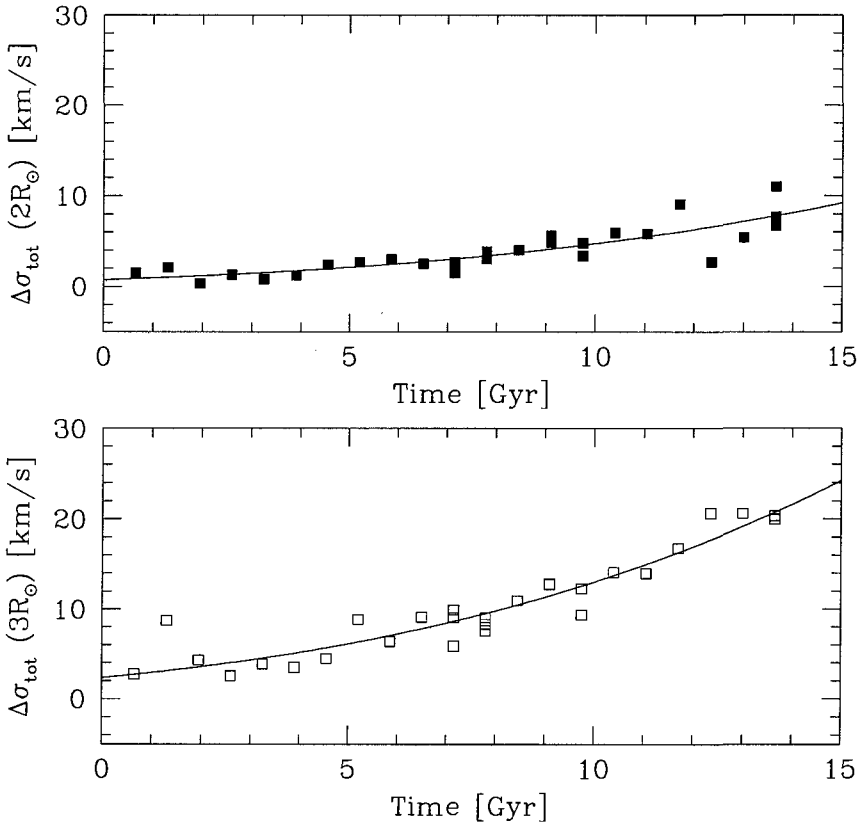

Figure 4.16: The evolution of $\Delta\sigma_{tot}(2R_\odot)$ (top panel; filled squares) and $\Delta\sigma_{tot}(3R_\odot)$ (bottom panel; empty squares) in the simulation with Halo 1 substructure. In both panels, the curves represent the best fits of the form (4.14) to the data (see text for details).

by applying the more general formula:

$$\Delta\sigma = \Delta_0(1 + t/\tau)^{\alpha'}, \tag{4.14}$$

where Δ_0 is a constant with units of km s^{-1}, τ is the timespan (in Gyr) over which the data is fit and α' is a coefficient similar to the heating coefficient introduced before. This relation is similar to the one introduced by Hänninen & Flynn (2002) in order

simulation data, either.

to better fit the observational σ data – recall the relation (4.12). The results of the best fits with this function (in which Δ_0, τ and α' are treated as free parameters) are presented in Figure 4.16, and have the following forms:

$$\Delta\sigma_{tot}(2R_\odot) = 0.740(1 + t/16)^{3.815}$$

and

$$\Delta\sigma_{tot}(3R_\odot) = 2.335(1 + t/16)^{3.539}$$

The function (4.14) provides a better fit[10] to the data. For example, the standard errors in the α' values are: $\pm 14.77\%$ (for the $2R_\odot$ case) and $\pm 7.389\%$ (for the $3R_\odot$ case), respectively. The above fits suggest that, at large radii, the gradient $\Delta\sigma_{tot}$ caused entirely by the interaction with substructure can be described by a function of the form $\sim (1 + t/\tau)^{\alpha'}$, where the coefficient α' takes values between 3-4.

These results suggest that the CDM substructure may be an important factor in the disk heating in the outer parts of the disk, in contrast with its negligible effect in the inner parts ($\sim R_\odot$). Measurements of velocities of stars beyond the solar radius in the Galaxy are needed in order to confirm or refute this result.

4.2.5 Can Other Mechanisms Explain the Observed Heating Rate in the Solar Neighborhood?

The results presented in the previous Section suggest that heating by CDM substructure is inefficient within the solar radius (although it may be more efficient at large radii) and may not be able to reproduce the amount of heating observed in the solar neighborhood. If this result is representative for the CDM models, we have to look to other heating mechanisms as responsible for heating the disk up to the measured

[10]For reference purposes only, we mention that we have also found the analogue formula (4.10) to provide a better fit to the σ_{tot} simulation data. In this case the α values are ~ 0.3 (for $2R_\odot$) and ~ 0.6 (for $3R_\odot$). However, we caution that the σ_{tot} data contains a significant contribution from numerical noise and may not measure accurately the effect of substructure.

values. This may be indeed the case, as other mechanisms are known to operate in the disk and to generate a significant amount of heating. The most likely mechanisms of this kind are reviewed below (see also the discussion in §1.2.4).

One example is heating as a result of collisions between stars and molecular clouds orbiting in the plane of the disk. Spitzer & Schwarzschild (1951, 1953) were the first to propose this mechanism and, also, to predict that in order to reach the required amount of heating molecular clouds have to have masses between $10^5 - 10^6\ M_\odot$. Later observations have discovered that molecular clouds in this mass range indeed exist in the Galactic disk.

A number of analytical and numerical studies have investigated the AVR produced by giant molecular clouds (GMC). Lacey (1984) estimated analytically the effect of GMCs with masses $10^6 M_\odot$ and predicted that the velocity dispersion of the disk stars should evolve as: $\sigma \sim t^\alpha$, with $\alpha = 0.25$. Villumsen (1985) simulated numerically GMCs with masses $M_{GMC} \sim 10^6 M_\odot$ and found that the vertical and radial components of the velocity dispersions can be approximated by: $\sigma_z \sim t^{0.31}$ and $\sigma_R \sim t^{0.25}$. More recently, Hänninen & Flynn (2002) performed another numerical simulation of GMCs with masses of $10^6 M_\odot$ and have found that $\sigma \sim t^{0.21}$. All the above studies agree that the heating index produced by GMCs is below that inferred from observations ($0.3 < \alpha < 0.6$; where $\sigma \sim t^\alpha$). This is because, although the GMCs are very massive, they are too few to be entirely responsible for the observed AVR. In addition, the predicted ratio $\frac{\sigma_z}{\sigma_R}$ is also inconsistent with observations (recall that the observed value is $\frac{\sigma_z}{\sigma_R} \simeq 0.5$). Lacey (1984) finds $\frac{\sigma_z}{\sigma_R} = 0.8$, whereas Villumsen (1985) obtains $\frac{\sigma_z}{\sigma_R} = 0.6$. GMCs seem to produce too much heating in the vertical direction. Nevertheless, these studies suggest that GMCs are a non-negligible source of heating in the disk, which needs to be taken into account in the final estimate of heating rates.

Table 4.2: The age-velocity dispersion relation (AVR) for other heating mechanisms.

Perturber	AVR	$\frac{\sigma_z}{\sigma_R}$	Reference
GMC ($M_{GMC} \sim 10^6 M_\odot$)	$\sigma_{tot} \sim \tau^{0.25}$	0.8	[1]
GMC ($M_{GMC} \sim 10^6 M_\odot$)	$\sigma_{tot} \sim \tau^{0.25}$	~ 0.6	[2]
GMC ($M_{GMC} = 10^6 M_\odot$)	$\sigma_{tot} \sim \tau^{0.21}$	~ 0.5	[3]
Spiral Arms	$\sigma_R \sim \tau^\alpha$; $0.2 \leq \alpha \leq 0.76$		[4]
Massive BH ($M_{BH} = 2 \cdot 10^6 M_\odot$)	$\sigma_{tot} \sim \tau^{0.5}$	$0.55 - 0.67$	[5]
Massive BH ($M_{BH} \sim 10^6 M_\odot$)	$\sigma_{tot} \sim \tau^{0.51}$	$0.47 - 0.67$	[3]
Very Massive BH ($M_{BH} \sim 10^7 M_\odot$)	$\sigma_{tot} \sim \tau^{0.5}$	$0.5 - 0.6$	[3]

References in Table 4.2 are: [1] Lacey (1984); [2] Villumsen (1985); [3] Hänninen & Flynn (2002); [4] de Simone, Wu & Tremaine (2004); [5] Lacey & Ostriker (1985).

Another possible scenario is heating by transient spiral arms. This mechanism is able to produce a heating index in the range $0.2 \leq \alpha \leq 0.76$ (de Simone, Wu & Tremaine 2004), which is consistent with observations. However, most of this heating occurs on the radial direction and almost none on the vertical one (Barbanis & Woltjer 1967). Lacey (1984) and Carlberg (1987) suggest that a combination of the two scenarios may explain the data: the spiral arms may heat the disk radially, whereas molecular clouds will heat it on the vertical direction.

Finally, if black holes (BH) contribute a substantial amount of mass to the halo of the Galaxy, than it might also contribute to the heating rate. Lacey & Ostriker (1985) investigated the impact of massive black holes ($M = 2 \cdot 10^6 M_\odot$) on the disk and estimated the AVR as: $\sigma_{tot} \sim \tau^{0.5}$. A similar result was obtained by Hänninen & Flynn (2002), for massive black holes with masses of 10^6 $M_\odot$ and 10^7 $M_\odot$. Whereas the observed AVR can be well explained by this mechanism, the same cannot be said of the ratio of the vertical and radial velocity dispersions. The $\frac{\sigma_z}{\sigma_R}$ obtained in

these studies are only marginally consistent with the observations (they are usually larger than 0.5). Black holes also pose a more fundamental problem, in the sense that massive BHs in the range $10^6\ M_\odot - 10^7\ M_\odot$ would endager the survival of the disks in dwarf spiral galaxies (Friese, Fuchs & Wielen 1995). The results of the above studies are summarized in Table 4.2.

4.3 Discussion

We have analyzed three simulations including substructure (one satellite realization drawn from a ΛCDM cosmological simulation and two realizations drawn from a SCDM cosmology) and one "control" simulation without substructure. Based on this set of simulations, our results have implications for the viability of CDM halos as hosts of stellar disks. We discuss this further below.

- **Differences between different cosmological models**

Are there any differences in disk heating between the ΛCDM and SCDM cosmologies? Our limited set of simulations does not allow us to draw definitive conclusions on this matter. Our data seem to suggest that it is not the cosmological model, but rather the properties of the most massive satellites that influence the outcome in disk heating. For example, Figure 2.10 shows that the mass functions of the dark matter substructure in the two cosmologies are similar, except at the higher mass end. The ΛCDM mass function (long dashed line) has fewer massive satellites than the two SCDM mass functions (short-dashed lines). The most massive satellite identified within the virial radius of the primary ΛCDM halo has a ratio $V_{peak}/V_{200} \sim 0.2$. In terms of the total mass of the satellite[11], this corresponds to $\sim 10^{10} M_\odot$. In the SCDM mass function, the most massive satellites have $V_{peak}/V_{200} \sim 0.42$, or, in

[11] In this calculation, we take into account the re-scaling of the peak velocities relative to the V_{200} of the galaxy model in our simulations (see §3.4.2 for details).

terms of total masses of satellites, $\sim 10^{11} M_{\odot}$. The fact that the SCDM substructure contains subhalos with initial masses about a factor of 10 larger than the mass of the disk suggests that these subhalos may be capable of critically influencing, in time, the heating of the disk. On the other hand, the most massive satellite in the ΛCDM substructure is a factor of 10 less massive than in the SCDM substructure and therefore, it may have less important effects on the disk.

The above argument also has some theoretical support. Indeed, a rough estimate of the tidal heating rate by substructure shows that it scales as $dE/dt \propto \int n(m_s)\, m_s^2\, dm_s$, where m_s is the subhalo mass (White 2000). Also, according to numerical simulations, the substructure mass function is described by the power law: $n(m_s) \propto m_s^{-1.8}$ (Ghigna et al. 2000, Springel et al. 2001; and our data). From these two relations, we can see that the tidal effects of substructure are dominated by the few most massive subhalos.

Besides the mass function of the substructure, the orbital properties of the satellites are also expected to play an important role. This is because the tidal effect that a satellite has on the disk depends not only the satellite mass, but also on how closely it approaches the disk. Figure 3.17, for example, shows that the SCDM substructure contains quite a few $10^9 - 10^{10} M_{\odot}$ satellites with pericenters reaching the disk (*i.e.* < 30 kpc). Also, the ΛCDM halo has many more small mass ($\sim 10^8$) satellites with pericenters within the 30 kpc limit, but practically no high mass satellites ($M > 10^9 M_{\odot}$).

In conclusion, stochastic effects are expected to dominate the heating rate. Simulations with a large set of substructure distributions (drawn from both SCDM and ΛCDM models) are needed in order to determine whether the heating rate depends also on the details of the CDM models.

- **The relation between structural changes in the disk and the orbital parameters of the dark matter satellites**

The orbital structure of the satellite population and the structural changes in the disk are directly related. Using the impulse approximation, we can estimate that each close encounter between a massive satellite increases the velocities of disk stars by $\delta v \sim GM_b/r_bV$ – where M_b is the bound mass of the satellite, r_b is the impact parameter and V is the instantaneous velocity of the satellite. We can make a rough estimate of the change in the velocity of the disk stars by selecting only the satellites above a certain mass limit (taken, for convenience, to be a few times $10^9 M_\odot$, *i.e.* of the order of the disk mass). Figure 3.17 shows that above this mass limit the SCDM (Halo 1) substructure has a small number of satellites which are both more massive and have lower pericentric radii than the satellites in ΛCDM substructure. We can further approximate the impact parameter of each satellite with its pericentric radius and assume that its mass loss is negligible (*i.e.* the bound mass is equal to the initial mass). For the Halo 1 (SCDM) substructure we obtain that the boost δv introduced by the most tidally important satellites is $\sim 20 - 30$ km s^{-1} (in agreement with the increase in the velocity dispersions obtained in the simulation), whereas for the Halo 2 (SCDM) and ΛCDM substructure the boosts are negligible (in these latter runs there are no satellites able to induce a δv greater than 10 km s^{-1}).

Of course, the impulse approximation and the above simplifying assumptions retrieve only the magnitude of the effect that satellites have on the disk. A more accurate derivation will have to include the efect of the dynamical friction (neglected here).

- **The increase in disk scale height due to tidal interactions with substructure**

As previously mentioned, one of the predicted effects of CDM substructure is the non-uniform heating of the disk. In particular, the vertical structure of the disk

is expected to vary with the radial distance. For example, our Halo 1 simulation suggests that substructure acts so as to increase the disk scale height z_0 with radial distance R in the disk. Specifically, our results show that, over a period of about 10 Gyr, the disk scale height at the outskirts of the disk increases by a factor of ~ 1.8 in the simulation with Halo 1 (SCDM) substructure and by a factor of at least 1.2 in the simulation with Halo 2 (SCDM) substructure. In the same time, the scale height in the inner regions of the disk remains approximately unchanged. Several other studies alluded to this idea before. For example, previous numerical simulations of single satellite interactions have also found a differential increase in the disk height: QHF93 reported an increase in the disk scale height by a factor of 0.5 in the inner regions of the disk ($\sim$ 3.5 kpc) and up to about 400% at the outskirts of the disk; also, HC97 found that, after about 7 Gyr evolution, the scale height at $\sim$ 20 kpc was 1.5 times larger than the scale height near the center of the disk.

The advantage of our study is that it includes a realistic distribution of satellites, which makes this result relevant in the context of cosmological models. This suggests a possible constraint for CDM models. Future studies, using a statistical sample of simulations, need to clarify if the cold dark matter substructure leaves a particular signature in the shape of the $z_0(R)$ dependence (i.e., whether the rate of increase in the $z_0(R)$ function is constrained by the cosmological model or whether it takes a variety of forms). In particular, it would be interesting to compare $z_0(R)$ obtained as a result of CDM substructure with the same dependency produced by other heating mechanisms (such as molecular clouds or spiral arms, which are expected to leave a rather constant scale height with radius).

This prediction of the CDM models can be, in principle, tested against observations. As mentioned before, in our Galaxy, the observations are currently limited to the solar neighborhood. Thus, at present, it is not possible to study the the disk scale height as a function of Galactocentric distance. This situation may change within the next decade with the launching of the Space Interferometer Mission (SIM) satellite, which will measure distances and kinematics of stars much further into the

disk. Alternatively, one can investigate the radial structure of galactic disks in external galaxies which undergo mergers. A large sample of galaxy disks is that of Schwarzkopf & Dettmar (2000; 2001), which includes radial and vertical information on 110 highly inclined and edge-on disk galaxies. Of the total number of galaxies surveyed, 49 are interacting and 61 are isolated. This study finds that galaxies which undergo minor mergers ($M_{sat}/M_{disk} \sim 0.1$) have, on average, disk scale heights ~ 1.5 times larger than the isolated galaxies (Schwarzkopf & Dettmar 2000). Furthermore, in interacting galaxies the scale height is seen to increase systematically with the radial distance, a feature that these authors refer to as a "hallmark" of tidal interactions (Schwarzkopf & Dettmar 2001). According to these authors, the increase in scale height, measured as the gradient between the corresponding values at the edge and inner part of the disk, is estimated to be $\sim 14\% - 22\%$ of the mean scale height.

Our simulations show, for a similar mass range of the mergers, an increase in the disk scale height of the same magnitude as the one measured in the sample of disk galaxies of Schwarzkopf & Dettmar (2000; 2001) (recall Figure 4.5).

We note that a similar conclusion has been reached by Benson et al (2003) in their semi-analytical study of disk heating by substructure in the ΛCDM cosmology. These authors also find that the increase in disk scale height is consistent with the observational data and conclude that the observed thickness distribution of stellar disks is consistent with the ΛCDM model. Their study has the advantage of including a much larger set of substructure realizations than ours, although, on the other hand, it may not accurately follow the full dynamics of the problem. It is however reassuring that both numerical simulations and semi-analytical studies reach similar conclusions.

- **Final remarks on the validity of the CDM models**

Our simulations show that the substructure at $z = 0$ plays only a minor dynamical role in the local heating of the disk (at $\sim R_{\odot}$) over several Gyrs. Specifically, we obtain that the heating rate due to present day CDM substructure in the solar neighborhood is lower than the observed rate in the Galaxy at the same location. If substructure in the halos considered here is representative of galaxy-sized CDM halos (and one has no reason to suspect it is not), this implies that the tidal heating of thin stellar disks by substructure halos is consistent with the observational evidence. Therefore, the substructure does not preclude virialized CDM halos from being acceptable hosts of thin stellar disks like that of the Milky Way.

Summary:

The results in this chapter suggest that concerns regarding excessive tidal heating of thin stellar disks by substructure in CDM halos may be less serious than previously thought. We show that, once subhalos with the mass, structure, and orbital distribution expected in a present ($z \sim 0$) CDM universe are considered, thin stellar disks at the center of such halos may survive the tidal interactions with the subhalos. This is because subhalo masses in galaxy-sized halos are typically below $10^{10} - 10^{11} M_{\odot}$ and have orbits that usually do not take them close to the disk, where their tidal effects are greatest. We conclude that the substructure observed in virialized CDM halos is consistent with the existence of thin stellar disks such as that of the Milky Way.

We have also quantified the disk heating rate caused by the interactions with the satellites (whenever present). We compare the results with similar results derived from observations in the Galaxy and show that the dark matter substructure is inefficient in reproducing the observed heating rate in the solar neighborhood, but can have a significant effect in the outer parts of the disk. Our results suggest that beyond $\sim 2R_{\odot}$, the heating rate due to CDM substructure may become comparable with the rate produced by other heating

mechanisms, such as spiral arms or encounters with molecular clouds or black holes. Future observations, characterized by an improved accuracy in measuring the velocities of stars beyond the solar radius, should be able to test the predicted effects of dark matter substructure with radius.

Chapter 5

The Response of the Disk to CDM Satellite Tides

Abstract

This chapter investigates the tidal coupling between the disk and the surrounding dark matter satellites and its effects on the evolution of the disk. The most noticeable effects of substructure on the disk may be traced to the few most massive subhalos and include: *(i)* tilting of the disk caused by the orbital decay of the satellites; and *(ii)* forcing of short-lived stellar warps in the disk as a result of tidal shocks that arise during the pericentric passages of a given satellite. The signatures of tidal interactions on the structure of the disks are discussed in the context of observational evidence.

Whereas the previous chapter was concerned with the global effect of substructure on the evolution of the disk, the present chapter concentrates on the effect of individual satellite encounters. The aim is to identify the relation between the disk-satellite interactions and changes in the structure of the disk, as well as to describe this relation in a quantitative way. We shall concentrate the analysis on the numerical simulation where the effects of the substructure are clearly defined, i.e. the "Halo 1" SCDM simulation. This chapter is organized as follows: Section §5.1 investigates

the tidal coupling between dark matter satellites and the disk and identifies those satellites which have a significant impact on the structure of the disk. The effect of these satellites is analyzed in detail in the following sections: Section §5.2 considers the tilting of the disk in response to the tidal interactions with the main satellites; §5.3 investigates the generation and maintenance of stellar warps through recurrent tidal interactions with the main dark matter satellites. Section §5.4 discusses the validity of the latter mechanism in the context of observational evidence.

5.1 Tidal Coupling between Satellites and the Disk

Figure 5.1 shows a snapshot of dark matter satellites in Halo 1 run and the Galaxy disk (for clarity, the parent dark matter halo and the bulge are not shown). The box is a (x, z) projection of the Galaxy model, 175 kpc on a side. The snapshot is taken at $t \sim 190$, which is the time when the satellite GS1 is at the first apocenter of its orbit. Particles are color coded in terms of their density values (provided by SKID), the densest regions being colored in yellow and the least dense ones in blue.

Let us denote the largest and the second most massive satellites in the upper half of Figure 5.1 as GS1 and GS2, respectively. These satellites start with initial total mass of $8.62 \times 10^{10}\ M_{\odot}$ and $1.85 \times 10^{10}\ M_{\odot}$, respectively. We will show that these satellites are the main perturbers of the disk and analyze in detail their orbital motion. We will then quantify the strength of their interaction with the disk, as well as correlate their motion with the transient changes in the structure of the disk.

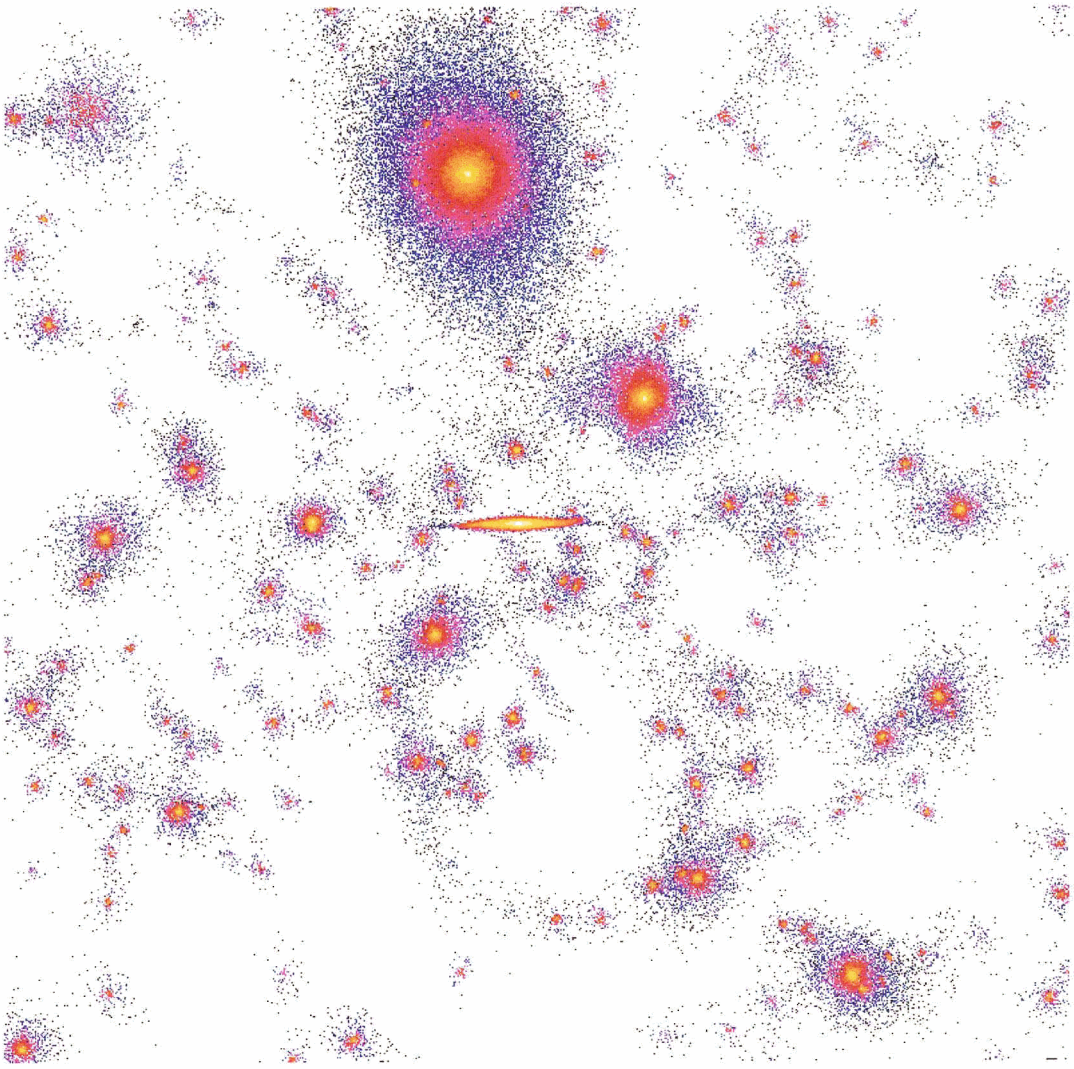

Figure 5.1: The Galaxy model including substructure halos. The box shown here is 175 kpc on the side and for clarity, the parent halo and the bulge are not shown. In this picture, satellites GS1 and GS2 are the largest and the second largest satellites in the upper half of the box, respectively.

5.1.1 Orbital Evolution of the Main Satellites

We turn our attention to the detailed orbits of the largest satellites in Figure 5.1. In doing so, we need to identify the center of mass of those systems at a succession of time snapshots along their orbit. Since the satellite may lose mass, the mass which is relevant for tracking the orbit is that of the material self-bound to the satellite. Therefore, satellite particles that are tidally stripped need to be excluded from the computation.

In order to calculate the self-bound mass of a given satellite at a certain time t, we use a numerical algorithm similar to the one proposed by Tormen, Diaferio & Syer (1998). The basic steps of this algorithm are as follows:

1) At time t, select all particles pertaining to the satellite (or, alternatively, the bound particles determined in a previous time step, if available). Then determine the position and velocity of the center of mass of these particles.

2) Calculate the kinetic energy K and the potential energy U for each of the selected particles. The kinetic energy is calculated in the reference frame where the center of mass is at rest. The potential energy includes the 2-body gravitational interactions between all selected satellite particles.

3) Determine which particles are unbound in this frame, i.e. those that have a positive total energy $K+U>0$. Discard those particles and calculate the new center of mass of the remaining particles. If no particles are unbound, the computation ends.

4) Iterate the steps *2)* and *3)* with respect to the newly determined center of mass until there are no more unbound particles. At this point the computation has converged and the final set of particles represents the bound material of the satellite at t.

We have applied the above algorithm to a few of the most massive satellites within the ~ 175 kpc distance from the disk. Figure 5.2 illustrates the orbital evolution of

the most massive satellite, GS1, in the (x, z), (y, z) and (x, y) planes, followed for about a Hubble time. Orbital decay is apparent in these plots.

Figure 5.3 is the same as Figure 5.2 but for satellite GS2, the second largest satellite within $\sim$ 175 kpc from the disk. The orbit of this satellite passes very frequently near the outskirts of the disk (recall that the outermost radius of the disk is at $R \sim 30$ kpc). The orbital decay of this satellite is less drastic than that of GS1. This is illustrated in Figure 5.4, which shows the evolution of the distance between satellites GS1 (top panel) and GS2 (bottom panel) and the center of the disk, and demonstrates that the rate of orbital decay is closely connected with the original mass of the satellite. Figure 5.5 shows the bound mass of the two satellites versus time. The large initial mass loss of GS2 also results in a slower orbital decay at later times. GS1 retains a larger self-bound mass at the end of the simulation in comparison with GS2, and proves remarkably resilient to mass loss[1]: after completing 8 full orbits (the equivalent of $\sim$ 10 Gyr) it still retains a well-defined core with $\sim$ 10% of its original mass. In comparison, the bound mass of satellite GS2 is after $\sim$ 10 Gyr (or 9 full orbits) only about 5% of the initial mass. The resilience of the satellite GS1 to mass loss (together with its proximity to the stellar disk) suggests that this satellite can leave its imprint on the structure of the disk for an extended period of time.

[1]Note that dark matter satellites are modeled with an NFW profile.

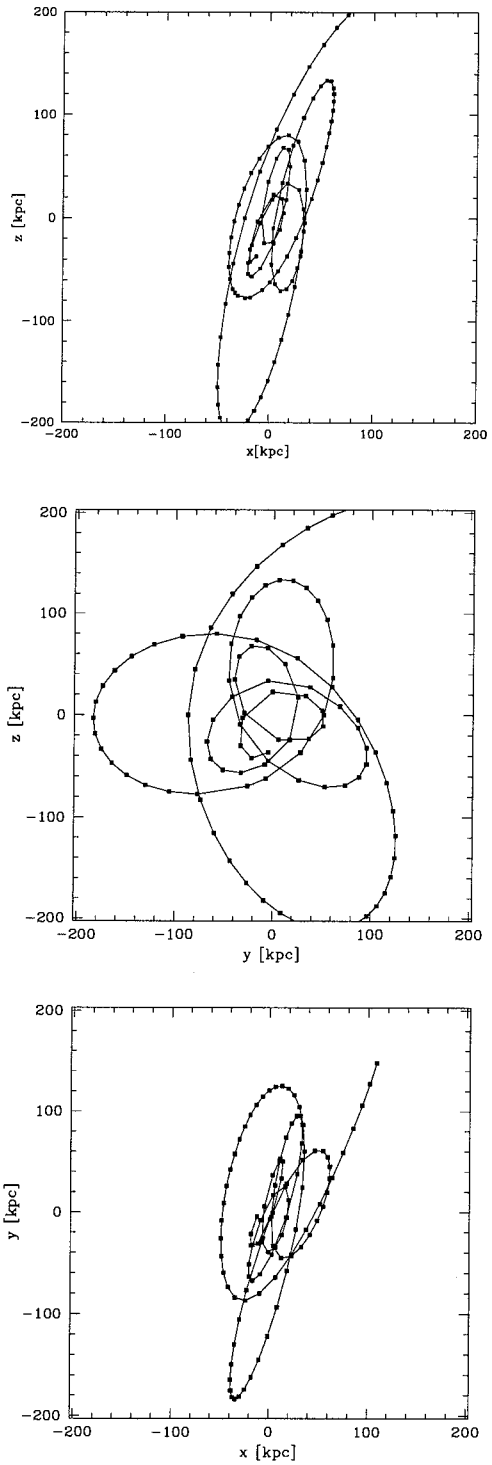

Figure 5.2: The orbital evolution of the most massive satellite (GS1), followed for about a Hubble time.

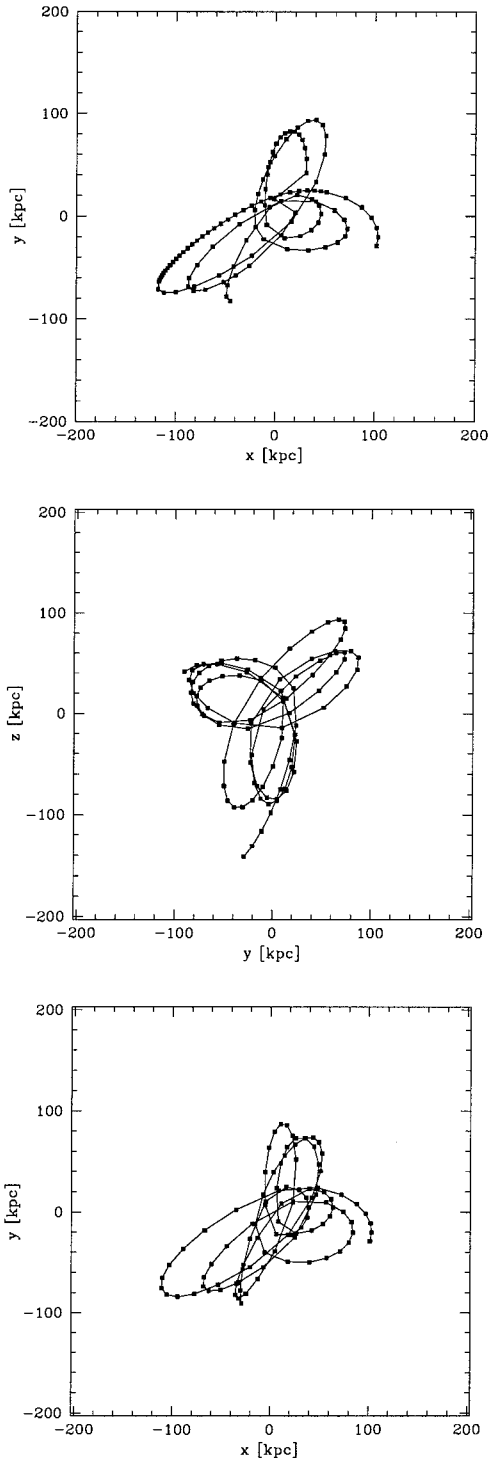

Figure 5.3: The orbital evolution of satellite GS2, followed for about a Hubble time.

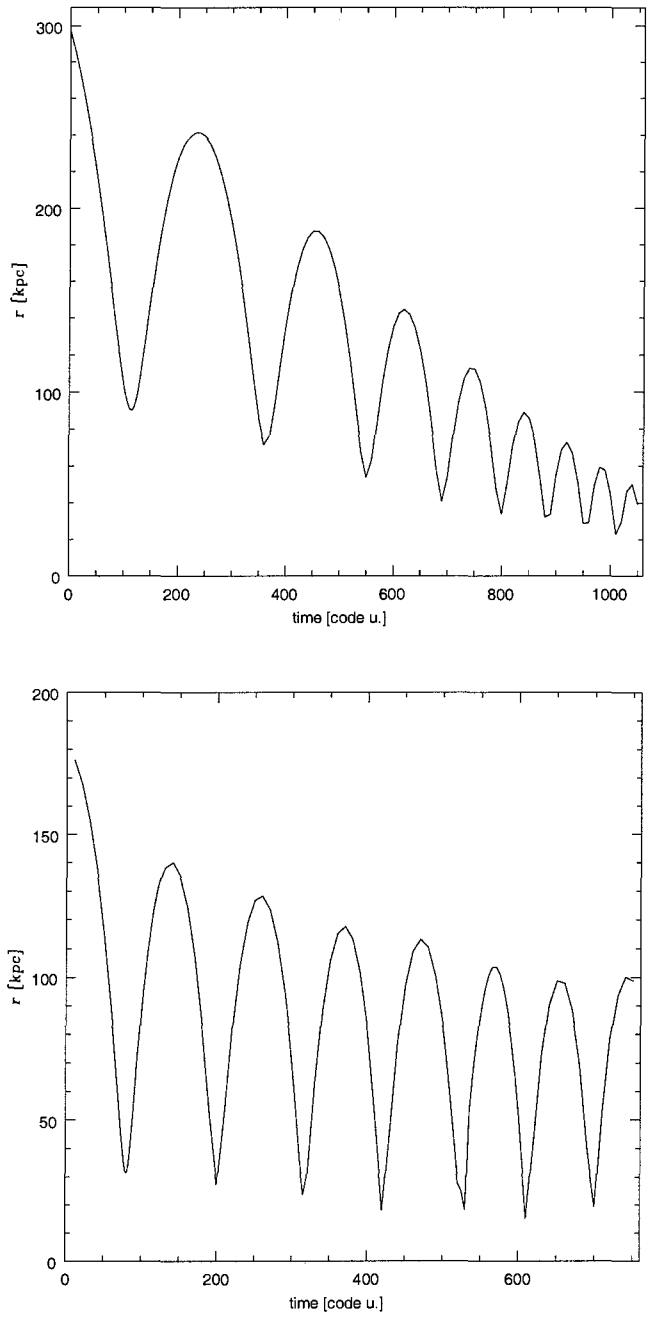

Figure 5.4: Distance of satellites GS1 (top panel) and GS2 (bottom panel) to the center of the disk as a function of time.

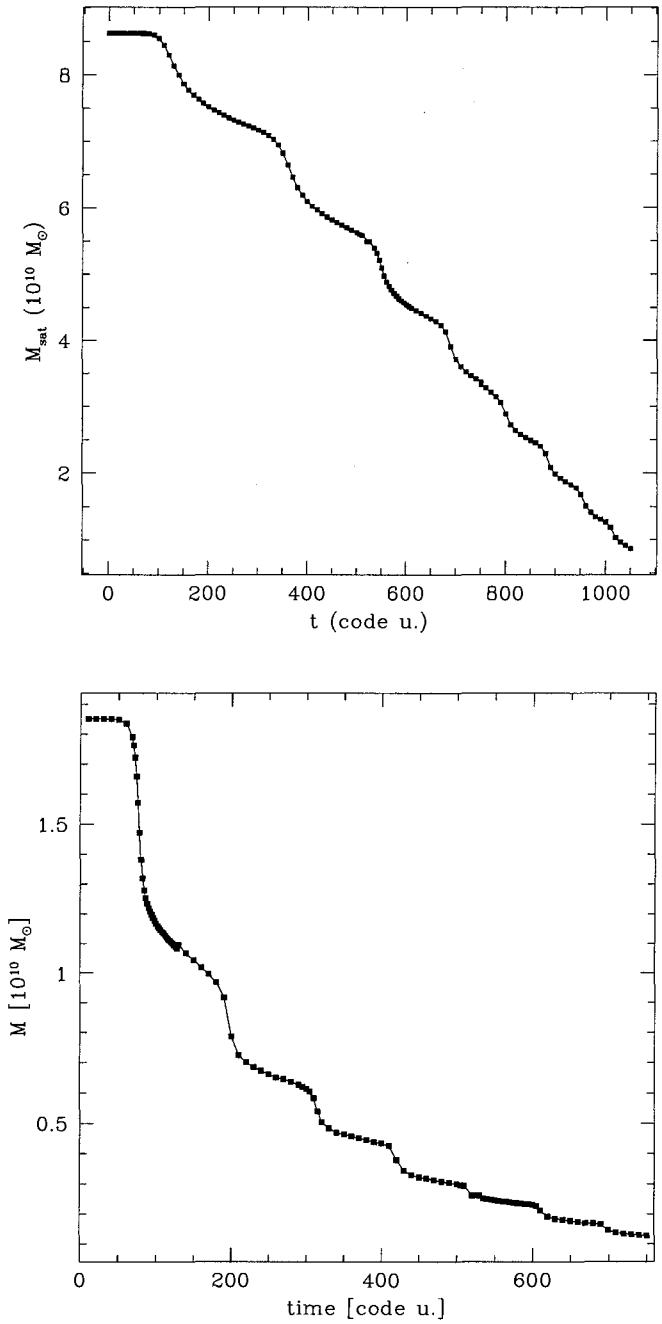

Figure 5.5: Mass loss of satellites GS1 (top panel) and GS2 (bottom panel). Plotted here is only the bound mass of each satellite.

5.1.2 Gravitational Forces

We wish to quantify the interaction between the dark matter satellites and the disk and to identify those satellites which couple more strongly with the disk. To this aim, we use the gravitational force between a satellite and the disk at pericenter assuming, for simplicity, the two systems are point masses. This rough approximation is not intended to be a precise calculation of the interaction between the two bodies, but rather to give us an idea about the order of magnitude differences between different satellite-disk interactions.

We proceed to identify the dark matter satellites which have the strongest impact on the disk by comparing their corresponding gravitational forces as defined above. For this, we select from Figure 3.16 the first ten satellites which meet the requirement of being most massive and having smallest pericenters (specifically, we order them in terms of quantities M_{sat}/R^2_{peri} and select the ten satellites with the highest values). We then run the algorithm presented in §5.1.1 to all those satellites and calculate the evolution of their gravitational forces. Among those ten satellites, we find that only two interact significantly with the disk: these are satellites GS1 and GS2 discussed in §5.1.1. Figure 5.6 shows the gravitational forces, due to these two satellites[2]. The second satellite, GS2, is initially about four times less massive than GS1 but its pericentric radius is smaller, so its effect on the disk dominates at early times ($t \leq 400$). The larger mass of GS1, on the other hand, leads to rapid decay of its orbit and, as a consequence, its gravitational (hence tidal) effects become strongest at late times.

Figure 5.7 shows the gravitational forces for the next two satellites, selected in order of importance of their gravitational interactions. Note that F_{grav} peaks are about a factor of three smaller than those in Figure 5.6. This implies that GS1 and,

[2]Henceforth, the time will be expressed in code units (1 code unit = 0.013 Gyr), in order to avoid any associations between individual tidal events and the history of the Galaxy. This is to emphasize that the timing of the disk-satellite interactions depends primarily on the given orbital parameters of the satellites and may vary from realization to realization.

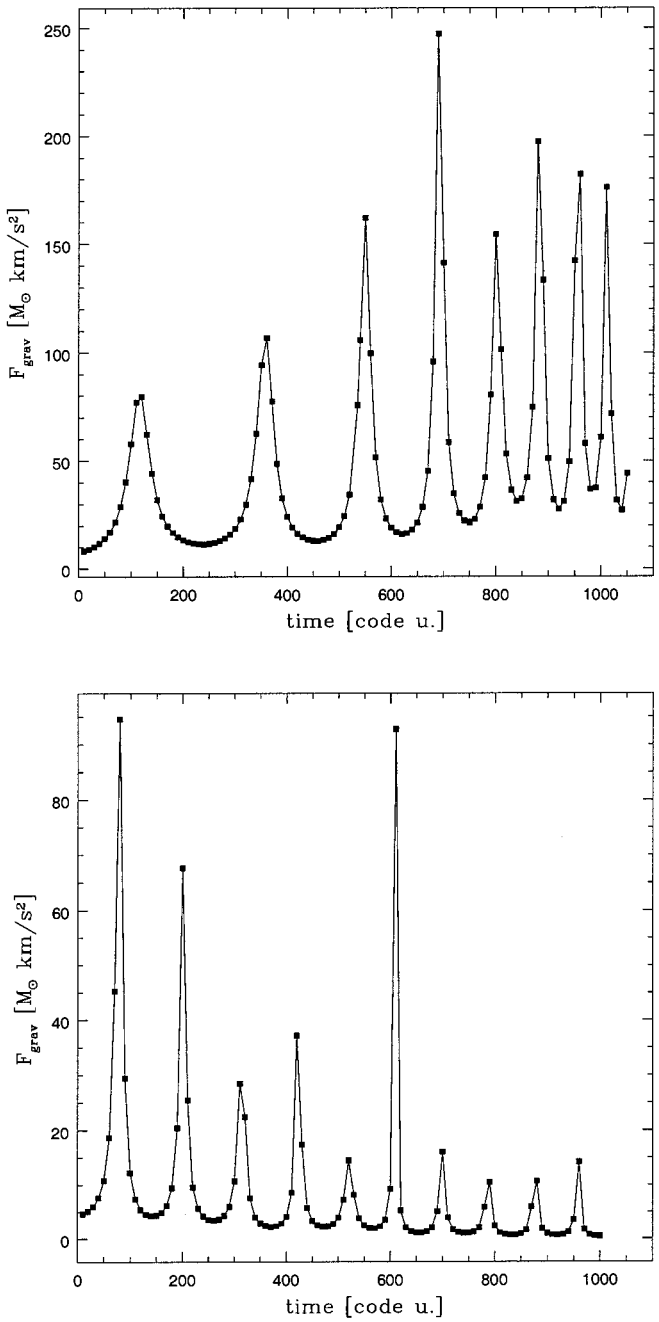

Figure 5.6: Tidal accelerations of satellites GS1 (top panel) and GS2 (bottom panel).

for a brief period of time, GS2, dominate the dynamical evolution of the disk. One may thus expect that the structural changes in the disk to coincide with the time of pericentric passages of these two satellites.

The importance of GS1 is further confirmed in Figure 5.8, which shows that the fluctuations in the radius of this satellite are well-synchronized with the potential energy of the disk (i.e., the pericenter passages of this satellite correspond to minima in the total potential energy of the disk), particularly after $t \sim 400$, when GS1 becomes the main perturber of the disk. In this figure, the potential energy of the disk is calculated as: $E_{d,pot} = \sum_{i=1}^{N_{disk}} m_i \Phi_i$, where m_i and Φ_i are the mass and the potential of each disk particle (as given directly by PKDGRAV), respectively.

The following Sections will investigate the effects that these repeated tidal interactions leave on the disk: §5.2 discusses various aspects of the tilting of the disk; and §5.3 proposes a model for the formation of tidally induced warps.

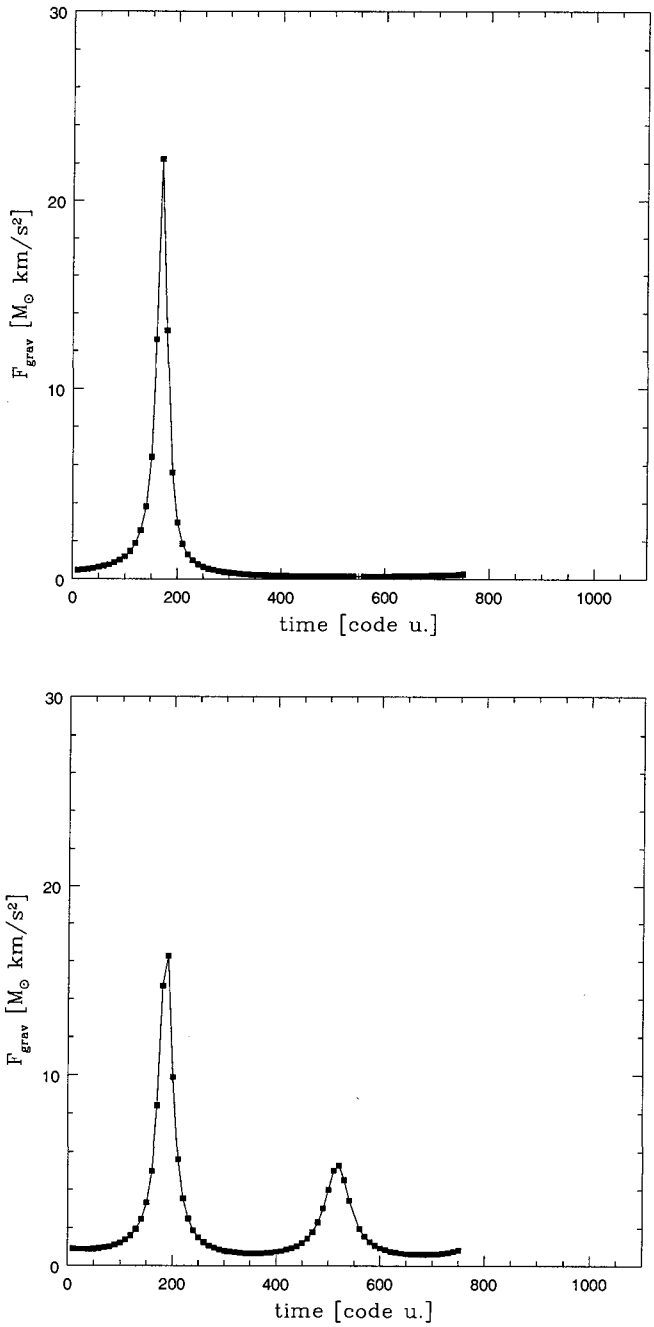

Figure 5.7: Gravitational forces of next two satellites, in order of importance. The F_{grav} values are at least a factor of three lower than the corresponding values of the previous two satellites.

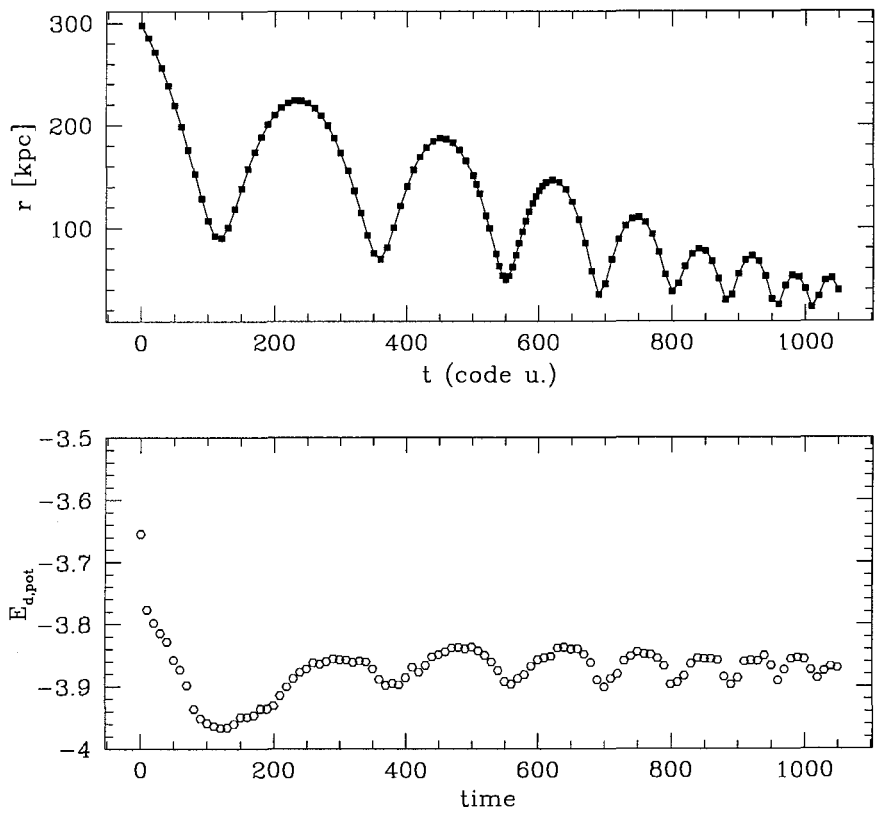

Figure 5.8: Radial evolution of the satellite GS1 (upper panel) and of the potential energy of the disk (lower panel). The synchronization between the minima in the potential energy of the disk and the pericenter passages of the satellite implies that this satellite is the major perturber of the disk.

5.2 Tilting of the Disk by Tidal Interactions

One response of the disk to the action of tidal forces is to tilt its spin axis. This response is global, in the sense that it involves all disk particles, in contrast with more localized reactions of the disk, such as flaring or warping. From theoretical arguments it can be shown that, in a given a satellite-disk gravitational interaction, the disk will tilt in the direction perpendicular to the orbit of the satellite in order to ensure the conservation of the total angular momentum of the disk + satellite system[3]. Therefore, if the tidal interactions in our simulations are strong enough, we should expect to detect a tilting of the disk and to find a correlation between the tilting and the orbits of the responsible satellites.

Figure 5.9 compares the tilting of the disk in the two simulations with SCDM substructure, "Halo 1" and " Halo 2", respectively. The measurements are recorded over the time span of $\Delta t \simeq 1000$ code u. (approximately 13 Gyr). Here θ is the angle subtended by the vertical component of the angular momentum of the disk (measured in the initial, non-tilted frame), $\vec{L}_z$, and the direction of the total angular momentum of the disk, $\vec{L}_{disk}$. The tilting of the disk is more dramatic in the Halo 1 simulation: in about a Hubble time, the tilt angle reaches 10^o relative to its original position. In comparison, in the Halo 2 simulation the disk tilts only by about 1^o or 2^o over the same period of time. The difference in tilting between the two simulations can be easily understood in terms of the differences in the behavior of the dark matter substructure (as previously discussed, the tidal effect of the substructure is negligible in the Halo 2 simulation and significant in the Halo 1 simulation).

In the following, we will concentrate our analysis only on the simulation which displays the tilt effect most clearly (i.e. the Halo 1 simulation). In this case, we will

[3]Of course, the assumption that the disk+satellite is a closed system, i.e. a system that conserves energy and angular momentum, is only a rough approximation. In reality, both disk and satellite lose energy and angular momentum through a variety of mechanisms - for example, through local collisions with halo particles, mass loss (mainly in the case of the satellite) and tidal interactions with other neighboring satellites.

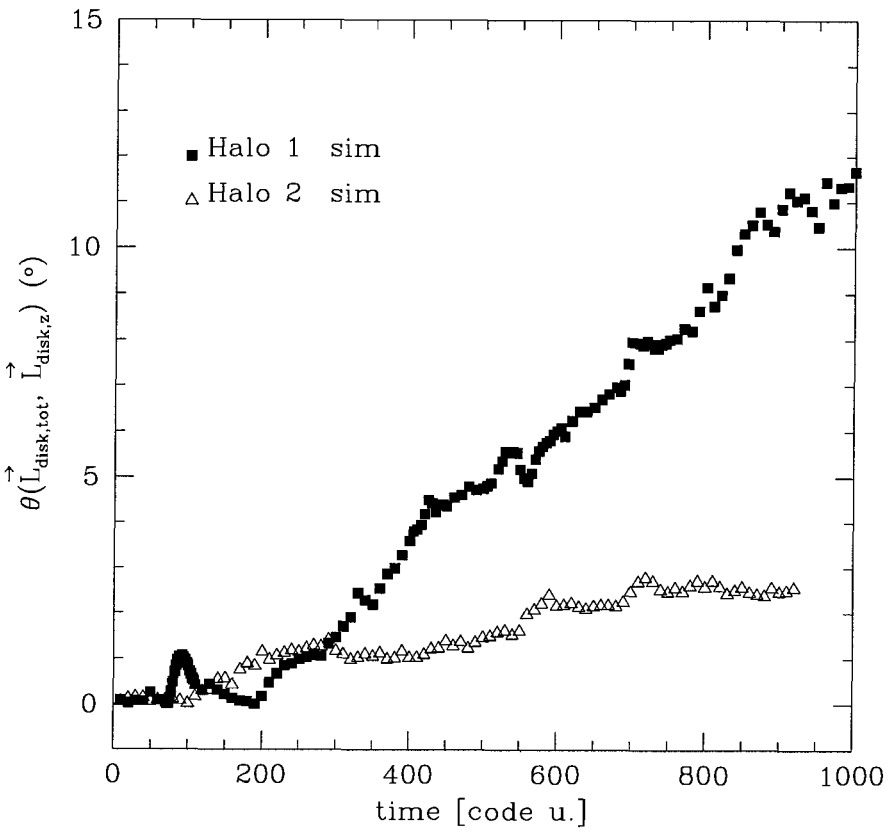

Figure 5.9: The tilting of the disk with time in the two simulations with SCDM substructure ("Halo 1" and "Halo 2"). θ is the angle subtended by the vertical component of the angular momentum of the disk (measured in the initial, non-tilted frame), $\vec{L}_z$, and the direction of the total angular momentum of the disk, $\vec{L}_{disk}$.

proceed to identify the satellite (or satellites) responsible for the tilt and investigate the relation between the orientation of the spin axis of the disk and orbital angular momentum of the satellites.

As the main tidal perturber, GS1 is expected to dominate the tilting of the disk. In terms of the magnitude of the effect, we can see that, after the time $t \sim 400$ (when the tides of satellite GS1 become more pronounced), the tilt angle increases progressively by an amount of $\Delta\theta \sim 1^o$ per each orbital period of satellite GS1. (Note

that the first "peak" in the tilt angle around $t \sim 100$ can be associated with the first pericentric passage of satellite GS2).

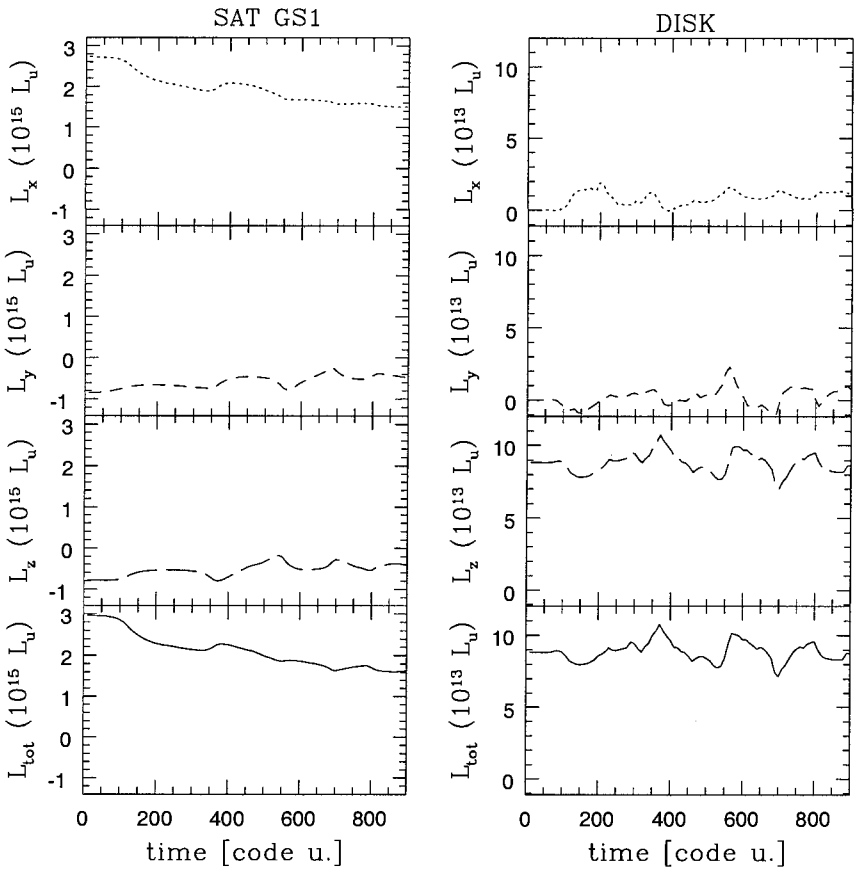

Figure 5.10: Angular momentum components of the disk (top panel) and that of the bound particles of the GS1 satellite (bottom panel). Dotted lines correspond to the x -components of the angular momenta, short dashed lines to the y -components, long dashed lines to the z -components and solid lines to the total angular momenta.

Is the tilting of the disk the result of angular momentum transfer between the main satellite to the disk? Figure 5.10 shows the x, y and z -components of angular momentum of the disk (top panel) and of the GS1 satellite (bottom panel), respectively. Here $\vec{L}_{GS1}$ is calculated taking into account only the bound particles of the satellite. This figure suggests that, although the GS1 satellite loses a significant frac-

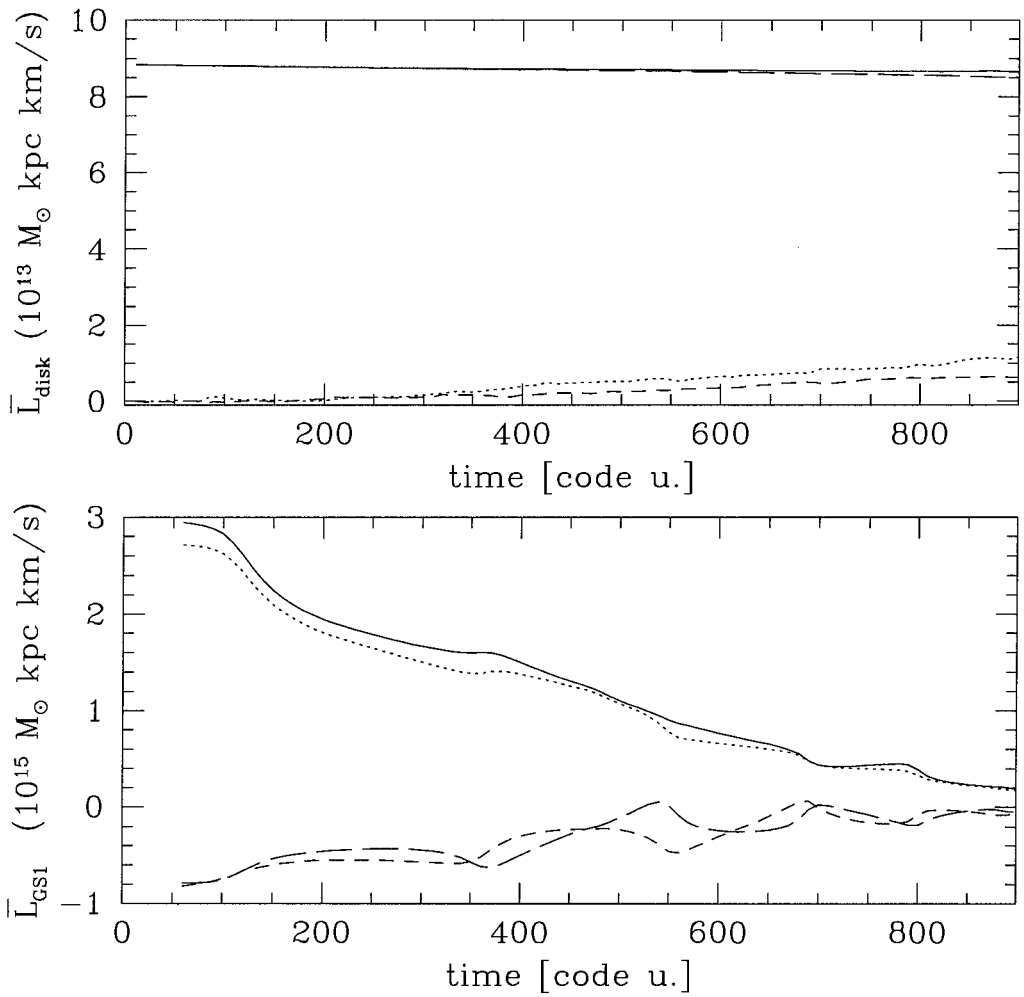

Figure 5.11: Angular momentum components of the total number of particles (i.e. both bound and unbound) in the GS1 satellite (left panel) and that of the disk (right panel). Dotted lines correspond to the x -components of the angular momenta, short dashed lines to the y -components, long dashed lines to the z -components and solid lines to the total angular momenta.

tion of its total angular momentum through mass loss and dynamical friction, only a minor part of this angular momentum is transferred to the disk. Note that the angular momentum of the satellite is several orders of magnitude larger than that of

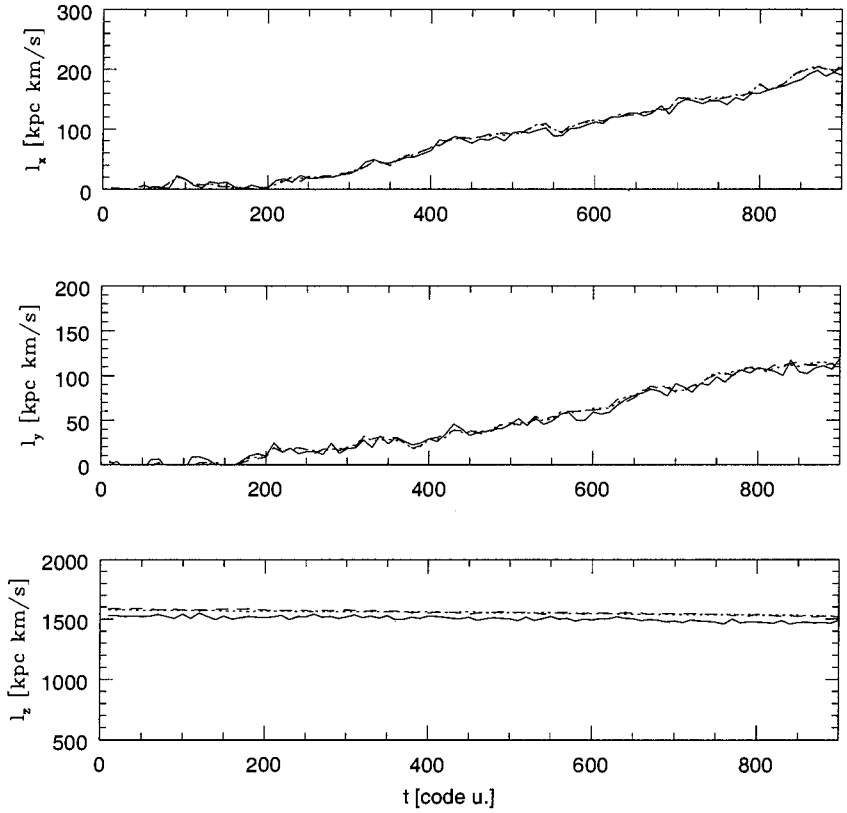

Figure 5.12: Specific angular momentum components of different radial bins in the disk: Dotted lines correspond to the specific angular momenta in the $0 < R < R_\odot$ radial bin, dashed lines to the $R_\odot < R < 2R_\odot$ radial bin, and solid lines to $2R_\odot < R < 3R_\odot$ radial bin. The top panel illustrate the x components; the intermediate panel the y components; and the bottom panel the z components.

the disk. Any major "deposition" of angular momentum from the satellite to the disk would be apparent in the disk angular momentum plot. A similar conclusion can be drawn from Figure 5.11, where we compare the angular momenta of the GS1 satellite, this time summed over all particles (i.e both bound and unbound) and that of the disk. In this case, the angular momentum components are calculated in the system of reference centered on the center of mass of the (satellite+disk) system. As in Figure

5.10, the angular momentum of the GS1 satellite is about two orders of magnitude larger than that of the disk and no major deposition of angular momentum into the disk is noticeable as a consequence of the pericentric passages of this satellite.

As we will show later in this chapter (§5.3.1 and §5.3.2), the disk is acquiring some angular momentum, although mainly locally. Figure 5.10 suggests that, globally, the disk is redistributing its angular momentum: Whereas the z component of disk angular momentum drops, the x and y components rise.

The redistribution of the angular momentum in the disk is mainly a reflection of the tilting of the disk and results in a change in the direction of the total angular momentum vector, $\vec{L}_{disk}$. Thus, the tilting of the disk is achieved mainly through the variation of the direction of the total angular momentum with little variation of its magnitude.

Figure 5.12 investigates the three components of the specific angular momentum in three different regions of the disk: the inner region $0 < R < R_{\odot}$ (dotted lines), the intermediate region $R_{\odot} < R < 2R_{\odot}$ (dashed lines), and the outer region $2R_{\odot} < R < 3R_{\odot}$ (solid lines). The three panels illustrate the $l_x = \Sigma_i^{N_d} L_{x,i}/M_d$ components (top panel); the $l_y = \Sigma_i^{N_d} L_{y,i}/M_d$ components (middle panel); and the $l_z = \Sigma_i^{N_d} L_{z,i}/M_d$ components (bottom panel) of the three regions in the disk. As Figure 5.12 shows, all three regions in the disk participate in the tilting. Since the parameters presented here describe the global changes in the angular momentum components, smaller effects such as warping of the disk cannot be identified in this kind of plot. However, these effects exist and as we will show later, they influence different regions in the disk with different magnitudes.

The direction of the tilt also contains important information. Thus, the disk tilts gradually in the direction perpendicular to the orbit of GS1. This is shown in Figure 5.13, which shows the evolution of the angle θ subtended by the total angular momentum of the disk, $\vec{L}_{disk}$ and the corresponding orbital angular momentum of the satellite GS1, $\vec{L}_{sat}$:

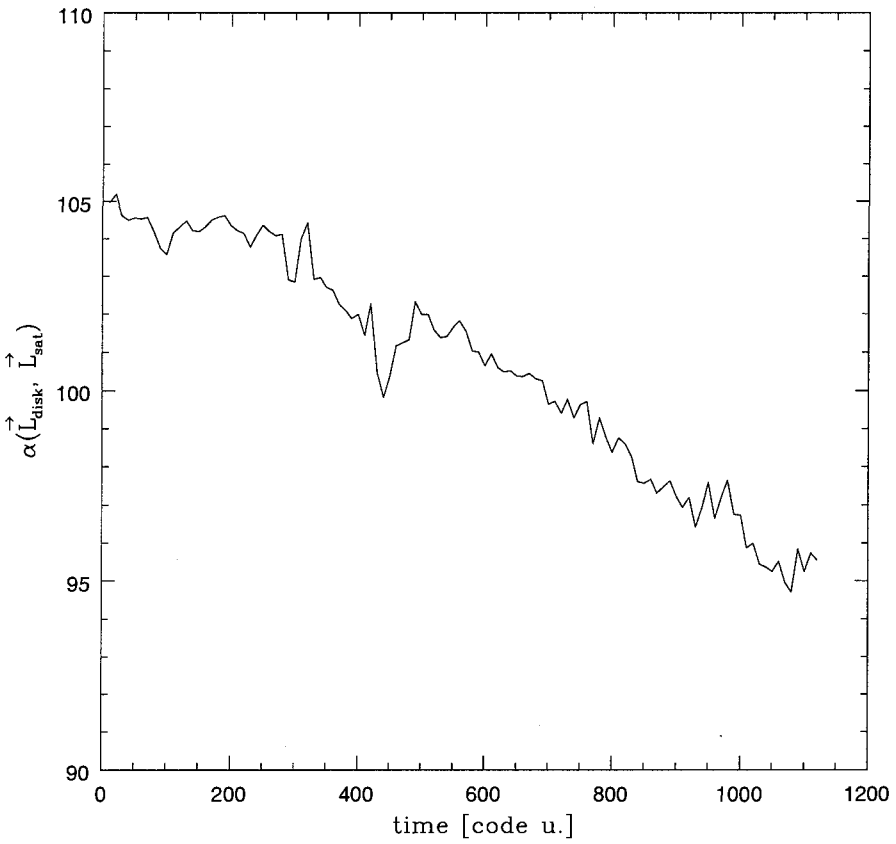

Figure 5.13: Alignment of the angular momenta of the disk and GS1 satellite. The angle α between the two angular momenta tends towards 90^0.

$$\cos\alpha = \frac{\vec{L}_{disk} \cdot \vec{L}_{sat}}{|\vec{L}_{disk}| \cdot |\vec{L}_{sat}|} \tag{5.1}$$

From this figure one sees that the angle θ decreases steadily from its initial value of $\alpha = 105^o$ towards $\alpha = 90^o$. This means that the satellite does not decay into the plane of the disk, but moves in the opposite direction. The result implies that the tidal interaction between these two systems may be able to redirect the satellite towards a high inclination orbit. If tidal interactions can not only bring, but also stabilize a satellite into a polar orbit, this could suggest a general mechanism of co-

herent alignment of the angular momenta of the satellites with respect to the angular momentum of the disk.

5.3 Tidally Triggered Warps

In this Section, we investigate another consequence of tidal interactions: the warping of the disk. During a satellite pericentric passage, the disk is likely to experience local changes in its structure, besides global ones. The largest disturbances in the stellar disk are expected to occur at or near the pericenter passages of the main satellites described before. Moreover, the outer parts of the disk will be more prone to changes, since the tidal torques from the neighboring substructure will have the greatest effect there.

Before proceeding with our analysis, we illustrate in the various panels of Figures 5.14 - 5.15 the (x, z) - projections of the disk at different time snapshots ($t =$ 580, 820, 980, 1040) following several of the pericentric passages of satellite GS1 (the corresponding pericentric passages occur at $t \sim$ 550, 800, 960, 1020).

We find that the tidally triggered disturbances occur both in the radial and vertical direction in the disk. The vertical structure of these features is most often asymmetric and one-sided. These feature do not represent the flaring of the disk, since they are typically very localized and disappear in short dynamical times. In terms of usual classifications of disk disturbances used in observational studies, these features are best qualified as irregular (Irr) warps (see, for example, Schwarzkopf & Dettmar 2001). We find that most of the tidally triggered vertical perturbations are one-sided, asymmetric features (in agreement with observations of tidally triggered warps - see Schwarzkopf & Dettmar, 2001). For the remaining of this chapter, whenever we use the term "(tidally triggered) warps", we implicitly assume that the warps have the characteristics described above. We therefore caution that our results and interpretations do not extend to other types of warps, such as the symmetric S-shape (one side up, the other side down) and U − shape (both sides oriented in the same

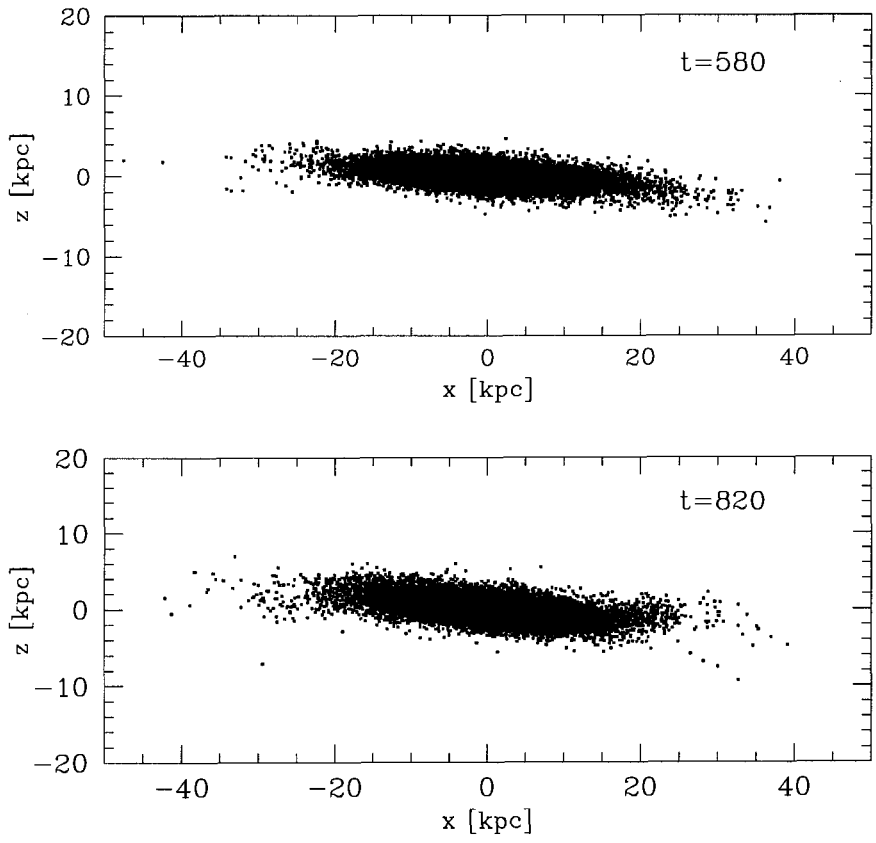

Figure 5.14: Edge-on view of the disk at time snapshots t = 580, 820. Transient features in the disk are tidally triggered around the times of nearby pericentric passages of satellite GS1, i.e. $t \simeq$ 550 and 800, respectively.

direction) warps, which do not seem to be produced frequently in our simulations.

From Figures 5.14 - 5.15, it is evident that the transient warps, tidally triggered by the pericentric passages of satellite GS1, are better defined in some snapshots and less well defined in others. This suggests that the strengths of the warps may be correlated with the strengths of the tidal interactions. Below, we will analyze in more detail how the tidal interactions can generate stellar warps and how they influence the location, amplitude and lifetime of these warps.

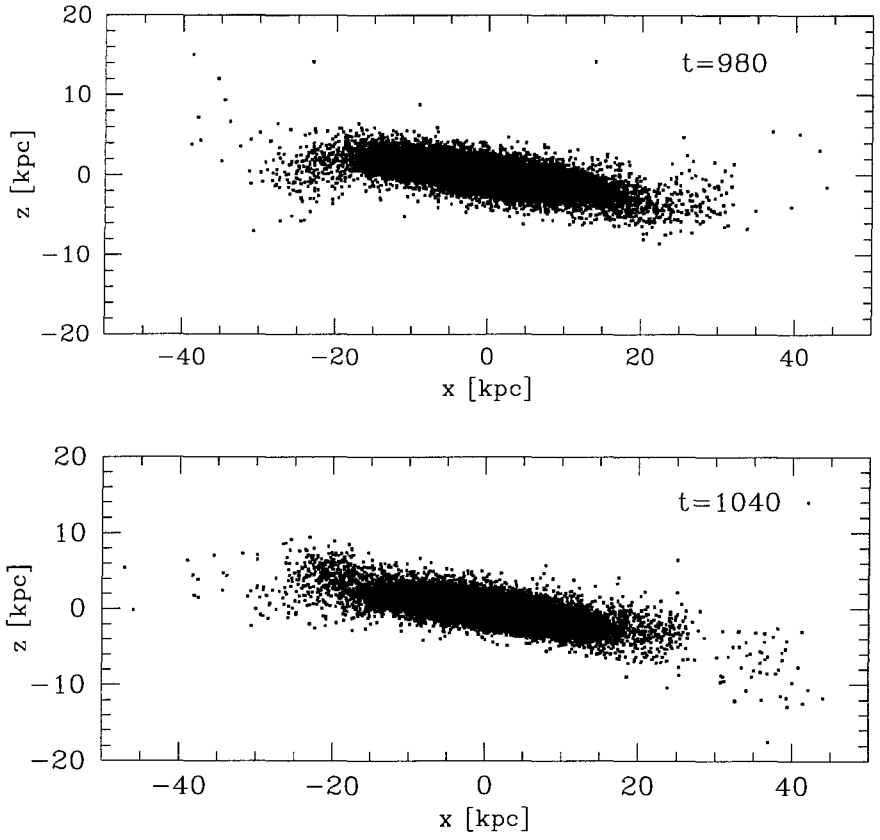

Figure 5.15: Edge-on view of the disk at time snapshots $t = 980$, 1040. Transient features are tidally triggered around the times of nearby pericentric passages of the satellite GS1, i.e. $t \simeq 960$ and 1020, respectively.

5.3.1 The Warp Triggering Mechanism

How do individual passages of satellites near the galaxy disk trigger these transient stellar warps?

As a first test case, we focus our analysis on the first pericentric passage of satellite GS2. The reason for choosing this particular passage is that around this time, $t \sim 70-90$, no other satellite exerts a tide of comparable magnitude with that of satellite GS2 (the satellite GS1 is still further away from the disk and its first pericentric passage

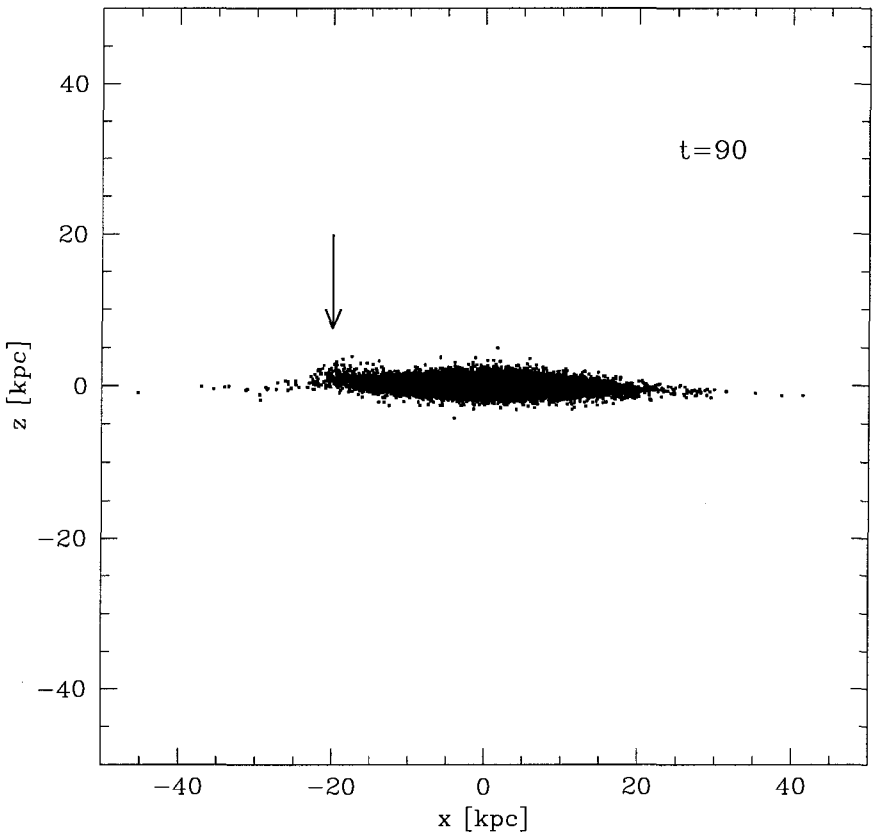

Figure 5.16: Edge-on view of the disk at $t = 90$. The stellar warp triggered around the first pericentric passage of satellite GS2, reaches a maximum height around this time.

occurs after this first warp subsides, i.e. around $t \sim 110$). Moreover, subsequent tides of comparable magnitude occur much later, after $t \sim 400$, when satellite GS1 becomes the important. The relative isolation of this first tide allows us to clearly identify the relation between the gravitational tide and the stellar warp (note that at later times the tides become more frequent, due to the orbital decay of satellites, hence raising the possibility that their effects on the disk may become superimposed).

Figure 5.16 shows an edge-on view of the disk at $t = 90$. The stellar warp triggered around the first pericentric passage of satellite GS2 reaches a maximum height around this time and is visible on the left side of the disk. Figure 5.17 shows the (x, z) and

(x, y) projections of the disk and the satellite GS2's orbit around first passage of this satellite through the plane of the disk. Thus, we can see that the satellite GS2 crosses the plane of the disk, moving upwards, at $t \sim 74$ (see top panel). Also, the satellite is closest to the polar axis as it approaches the disk at $t \sim 82$ (see bottom panel). The orbit of the satellite GS2 during this time is analyzed in more detail in Figure 5.18, which follows the evolution of the total (d), radial (R) and vertical (z) distances, respectively. Coincidently, the pericenter occurs at $t \sim 78$, nearly at the same time as the satellite crosses the plane of the disk (i.e. $z = 0$ in Figure 5.17).

The disk - satellite gravitational interaction (hence the tide) reaches a maximum around the time of GS2's passage through the plane of the disk (and in this case, near the time of the pericenter passage also). The time interval for this maximum ($\Delta t \sim 10$) is less than the orbital time in the parts of the disk which are most affected by the tide ($T_{orb} \sim 40$ at $R \simeq 15$ kpc).

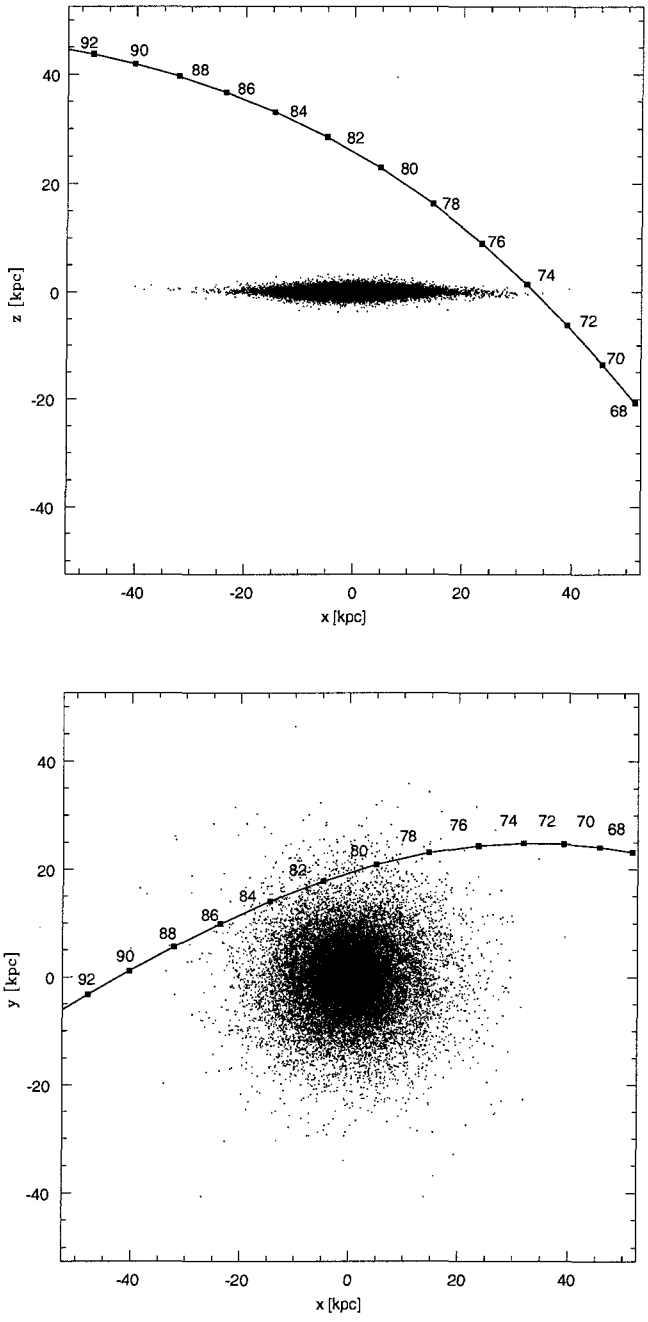

Figure 5.17: First passage of satellite GS2 through the plane of the disk: in the (x, z) projection (top panel) and the (x, y) projection (bottom panel). The orbit of the satellite is plotted in time steps $\Delta t = 2$ code units, during the times $t = 68$ and $t = 92$. The disk snapshot is chosen at time $t = 78$.

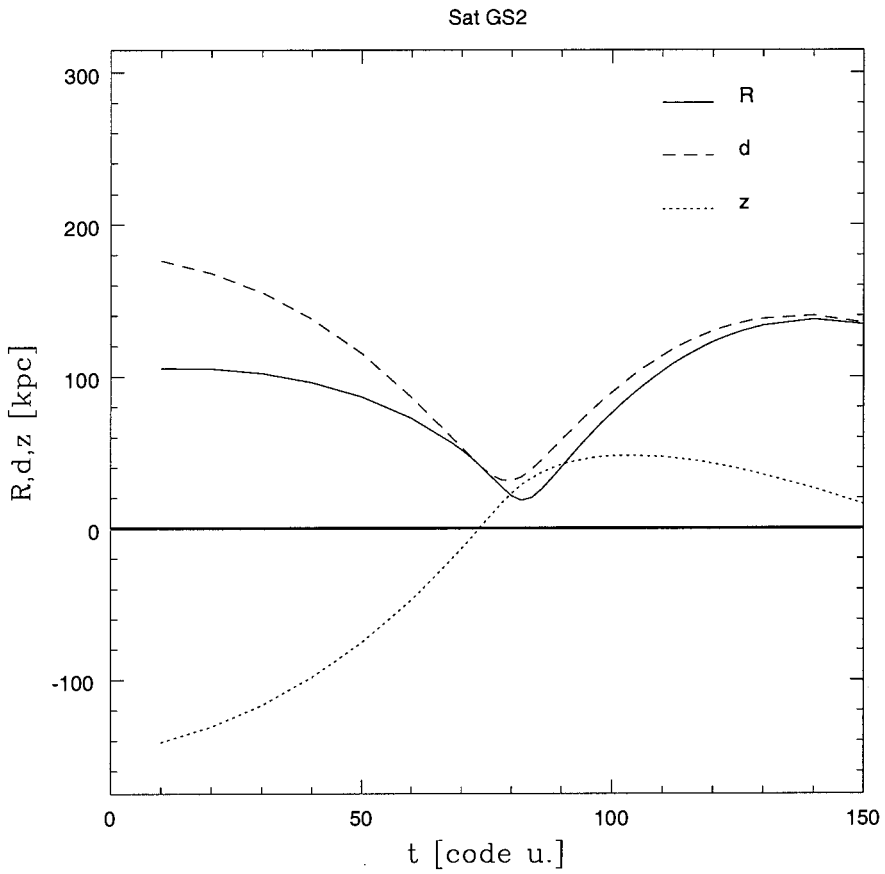

Figure 5.18: Evolution of the orbit of the satellite GS2 near its first passage through the disk. Dashed line corresponds to the total distance (d), the radial distance (R) is shown in full line and the vertical distance (z) in dotted line, respectively.

At the time of crossing, the satellite is relatively close ($\sim$ 45 kpc) to the center of the galaxy. Not all disk particles are affected in the same way by the satellite: Disk particles which are closest to the satellite will experience the largest gravitational acceleration. However, this is a dynamical process and, as the satellite is continuously moving, different disk particles will be affected at different times with different magnitudes.

The reaction of disk particles to the tidal interaction may be examined by estimating the tidal torque exerted by the satellite on the disk. The torque is given

by:

$$\vec{\tau} \equiv \vec{r} \times \vec{F} = m \cdot \vec{r} \times \vec{a}\,. \tag{5.2}$$

This relation can be applied to the case of the gravitational torque exerted by a given satellite – approximated as a point mass containing the bound mass of the satellite M_s – on a individual disk particle of mass m which moves on a nearly circular orbit around the center of the galaxy. From equation (5.2) it is easy to show that, for example, the torque on the x- axis can accelerate the particle only along the y and z -axes. Whereas the a_y acceleration can only influence the motion of the particle in the plane of the disk, the a_z component can move the particle above or below this plane. The same reasoning applies to the τ_y component of the torque. On the other hand, the τ_z component of the torque is not able to take particles out of the plane of the disk, but it can only accelerate the motion within the disk (i.e. changing only a_x and/or a_y). Therefore, the only torque components that can generate a warp are τ_x and τ_y. The magnitude of the warp might therefore be linked to the "radial torque", defined as:

$$\tau_r = \sqrt{({\tau_x}^2 + {\tau_y}^2)} \tag{5.3}$$

The torque is, of course, related to the change in total angular momentum of the disk,

$$\vec{\tau} = \frac{d\vec{L}}{dt}. \tag{5.4}$$

Thus, by measuring changes in the angular momentum of the disk particles one can infer the torque applied by the satellite. Initially, the disk particles move on nearly circular orbits characterized by an almost constant angular momentum: $L \simeq$

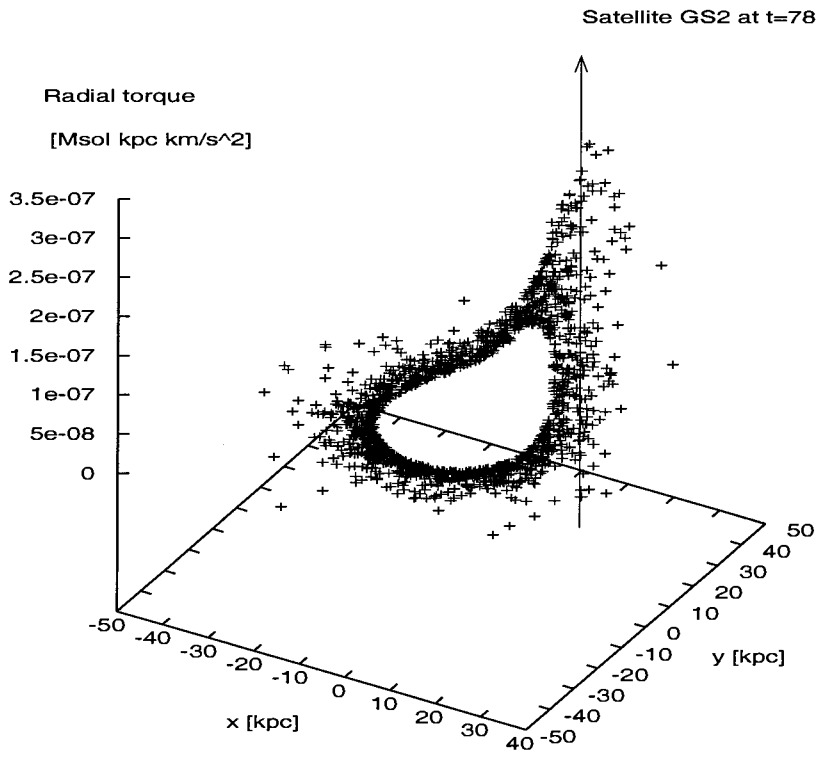

Figure 5.19: Radial torque, τ_r, exerted on the particles in the outer part of the disk ($r > 17.5$ kpc) at a time of the first pericentric passage of satellite GS2. The torque is expressed in units of $M_\odot \frac{kpc\,km}{s^2}$.

$L_z = const$, i.e. with zero net torque. A large non-zero torque is thus more likely to arise from a tidal interaction and can, therefore, reveal the influence of a satellite.

Figure 5.19 shows the radial torque τ_r induced by the satellite GS2 on the disk particles at the time of its first pericentric passage ($t = 78$). For simplicity, only the disk particles beyond $R > 17.5$ kpc are plotted (these are the particles most affected by the tide). As expected, disk particles closer to the satellite (located at $x \sim 10$ kpc and $y \sim 20$ kpc, around $t \simeq 78$) have higher torques.

The change in the angular momentum of the disk particles will be proportional to

the integral of the torque over the duration of the encounter. It can be easily shown that the integral is dominated by the short period of time around the pericentric passage. Thus, the change in the angular momentum should be well approximated by:

$$S_x \equiv \sum_{t=70}^{t=90} \tau_x \delta t; \quad S_y \equiv \sum_{t=70}^{t=90} \tau_y \delta t \tag{5.5}$$

Here S_x and S_y approximate the integral of the x and y torques over the time of interaction. We expect the disk particles with the highest S_x and S_y values to be the ones which will participate in the warp. Since either x and y - components are equally likely to move disk particles out of the plane of the disk, this is equivalent to saying that the warp will be formed by the particles with the highest $S_r = \sum_{t=70}^{t=90} \tau_r \delta t$. However, since we need a general selection criterion which is independent of the number of time outputs, we choose to select particles in terms of $< \tau_r >= (\Sigma_1^{N_{frames}} S_r)/(\Sigma_1^{N_{frames}} \delta t)$, where N_{frames} is the number of time frames used in the analysis. We then select a given number (N_{cut}) of particles with the highest $< \tau_r >$. The number of particles to be selected is somewhat arbitrary (a small value could result in selecting only a few particles in the warp, whereas a large value may include particles which do not belong to the warp). For the case of the particular interaction studied here, we have experimented with different N_{cut} values and we have found that a value of about $N_{cut} \simeq 100$ ensures that our criterion does not include particles which are not clearly related to the warp. The somewhat arbitrary number of particles that we select does not influence our analysis, since we are interested only in average values of the physical parameters of warp particles[4].

Figure 5.20 shows with empty circles all the disk particles with radii $R > 17.5$ kpc in the $S_x - S_y$ locus (particles in the inner regions of the disk are less affected by the tide). The $N_{cut} = 100$ particles with the highest $< \tau_r >$ values stand out in an

[4]In the current and following analyses of the warps, we will use a fixed value of $N_{cut} = 100$ (all warps with less than 100 particles are disregarded from analysis).

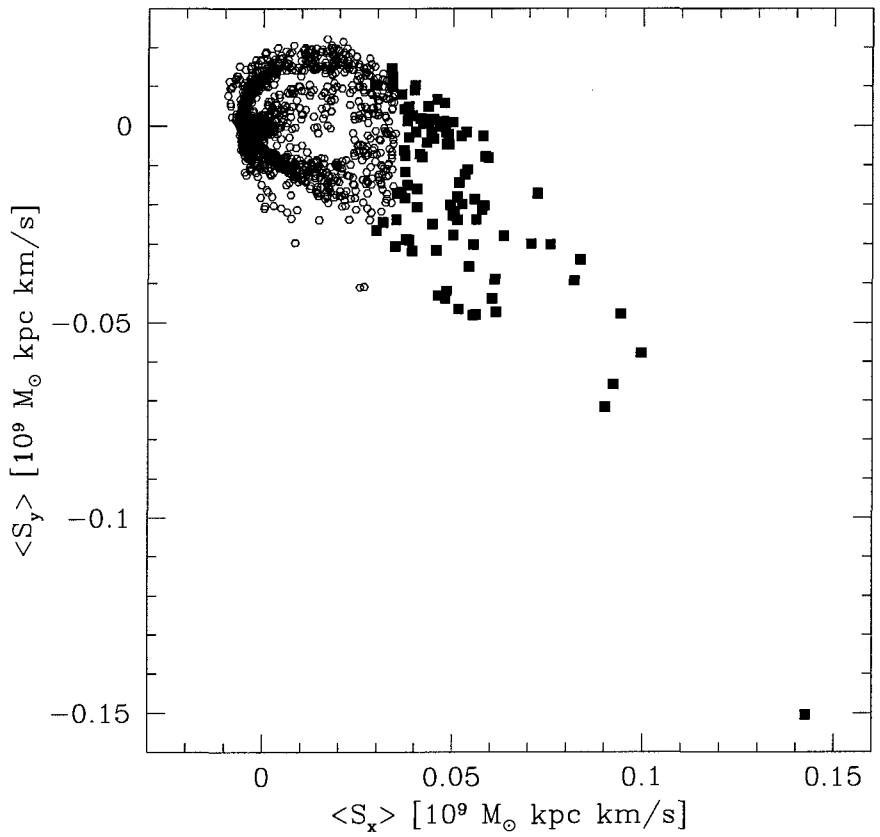

Figure 5.20: Particles in the warp are expected to stand out in the $S_x - S_y$ locus. Indeed, particles selected based on their highest $< \tau_r >$ form a $S_x - S_y$ "plume". Also, for clarity, only the disk particles with $R > 17.5$ kpc are plotted (particles in the inner regions of the disk are less affected by the tide and therefore, show little asymmetry in this type of plot).

asymmetric $S_x - S_y$ feature (resembling a "plume"). These particles, shown in the figure with filled squares, are expected to be part of the tidally triggered warp.

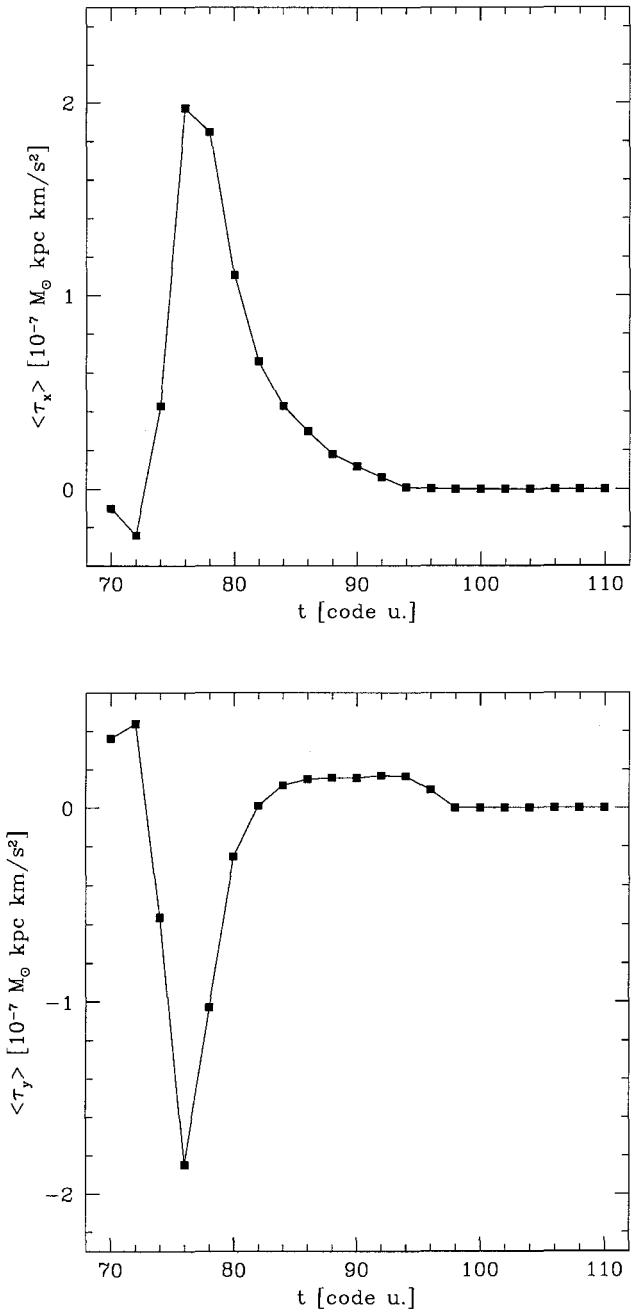

Figure 5.21: The time-evolution of the torques on x and y -axis for the $N_{cut} = 100$ disk particles with the largest tidal torques (maximizing both S_x and S_y values). The torques are found to act simultaneously and over a brief period of time during $t = 72 - 80$.

Figure 5.21 shows the evolution of the average torque components τ_x and τ_y, averaged over the particles selected with the method described above. The torques are found to act over a brief period of time during $t = 72 - 80$, confirming that these particles are indeed influenced by the tide. This implies that if no other external torques act on the disk, the choice of the time interval in (5.5) is not crucial: provided that the peak at $t \sim 76 - 78$ is included, S_x and S_y are well defined.

We expect particles in the warp to also stand out in a separate locus in the $\Delta L_x - \Delta L_y$ plot, with these values being measured in a small time interval just before and after the maximum interaction. As suggested by Figure 5.21, this time interval is $t \simeq 72 - 80$. For each disk particle we then calculate the following parameters:

$$\Delta L_x = \int_{t=72}^{t=80} \tau_x \, dt; \qquad \Delta L_y = \int_{t=72}^{t=80} \tau_y \, dt. \tag{5.6}$$

Figure 5.22 illustrates the $\Delta L_x - \Delta L_y$ locus in this particular tidal interaction. For clarity, only the disk particles in the outer part of the disk – beyond $R > 17.5$ kpc – are plotted. The plot displays a noticeable asymmetry in a form of a "plume" of particles with high gradients in both x and y - directions. These particles are highlighted in the figure with filled squares. We have verified that the particles in the $\Delta L_x - \Delta L_y$ "plume" are the same as the particles undergoing the maximum averaged radial torques. The agreement between these two methods ensures that we have found a reliable algorithm for identifying the warps. In what follows, we shall adopt the one based on the particles with highest $< \tau_r >$, because it is less sensitive to the time interval chosen for the analysis.

We now test whether the particles with highest $< \tau_r >$ actually end up in the transient warp noticed at time $t \sim 90$ in Figure 5.16.

Figure 5.23 shows an edge-on view of the disk at time ~ 90, where the horizontal axis is the line in the plane (x, y) from which the warp can be seen "face-on". Particles with highest $< \tau_r >$ (and forming the $S_x - S_y$ "plume") are highlighted here with filled squares. The figure confirms (visually) that these are the same particles as the

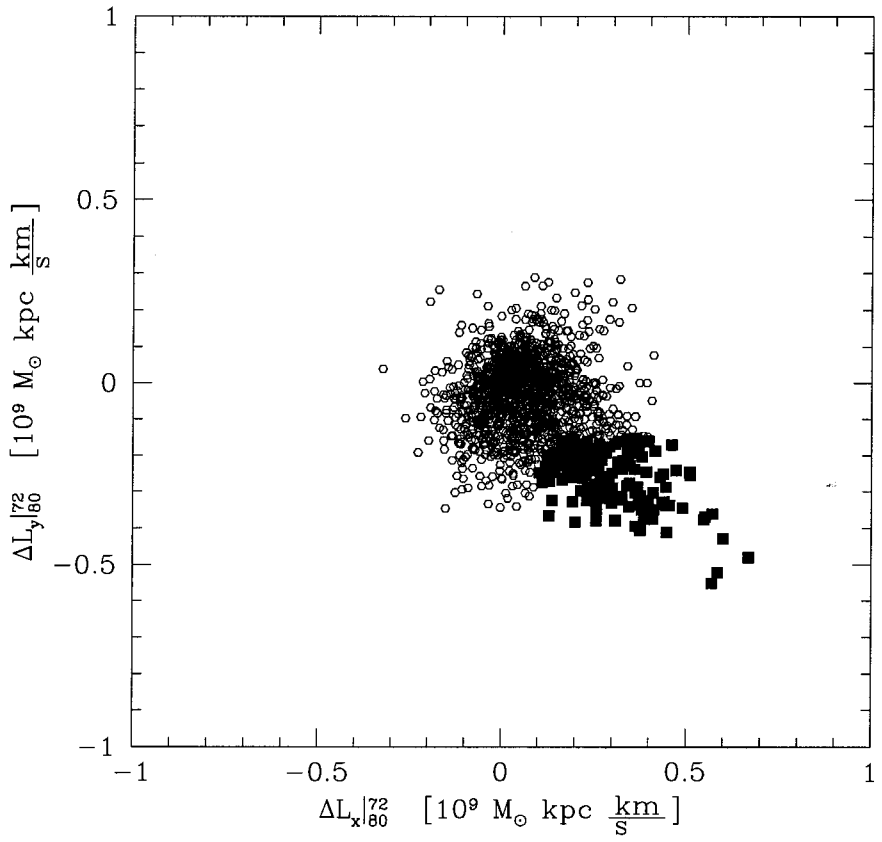

Figure 5.22: Particles in the warp stand out in the angular momentum $\Delta L_x - \Delta L_y$ locus (see equation (5.6)). For clarity, only the disk particles with $R > 17.5$ kpc are plotted.

ones defining the warp. A more detailed inspection of whether this is indeed the case is presented below.

We choose to study the motion of disk particles selected based on their highest $< \tau_r >$ in the following figures. Figures 5.24 and 5.25 show the evolution of these particles (shown in green colors), projected onto the (x, z) and (x, y) planes, respectively. The rest of the disk particles are shown in blue colors. A warp is seen to form around $t \sim 90$ and disperse (i.e. "phase-mix") quickly (in about $\Delta t \sim 40$) due to differential rotation in the plane of the disk.

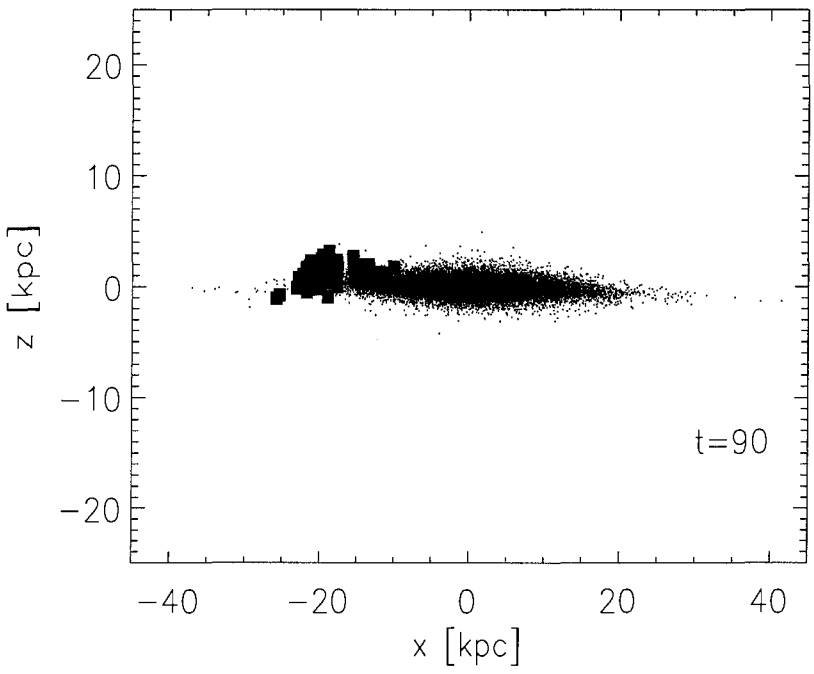

Figure 5.23: Edge-on view of the disk around the time of maximum height of the warp ($t = 90$). Disk particles with the highest $< \tau_r >$ (and pertaining to the the $S_x - S_y$ "plume") are highlighted with filled squares. The figure shows that these are the same particles as the ones pertaining to the warp.

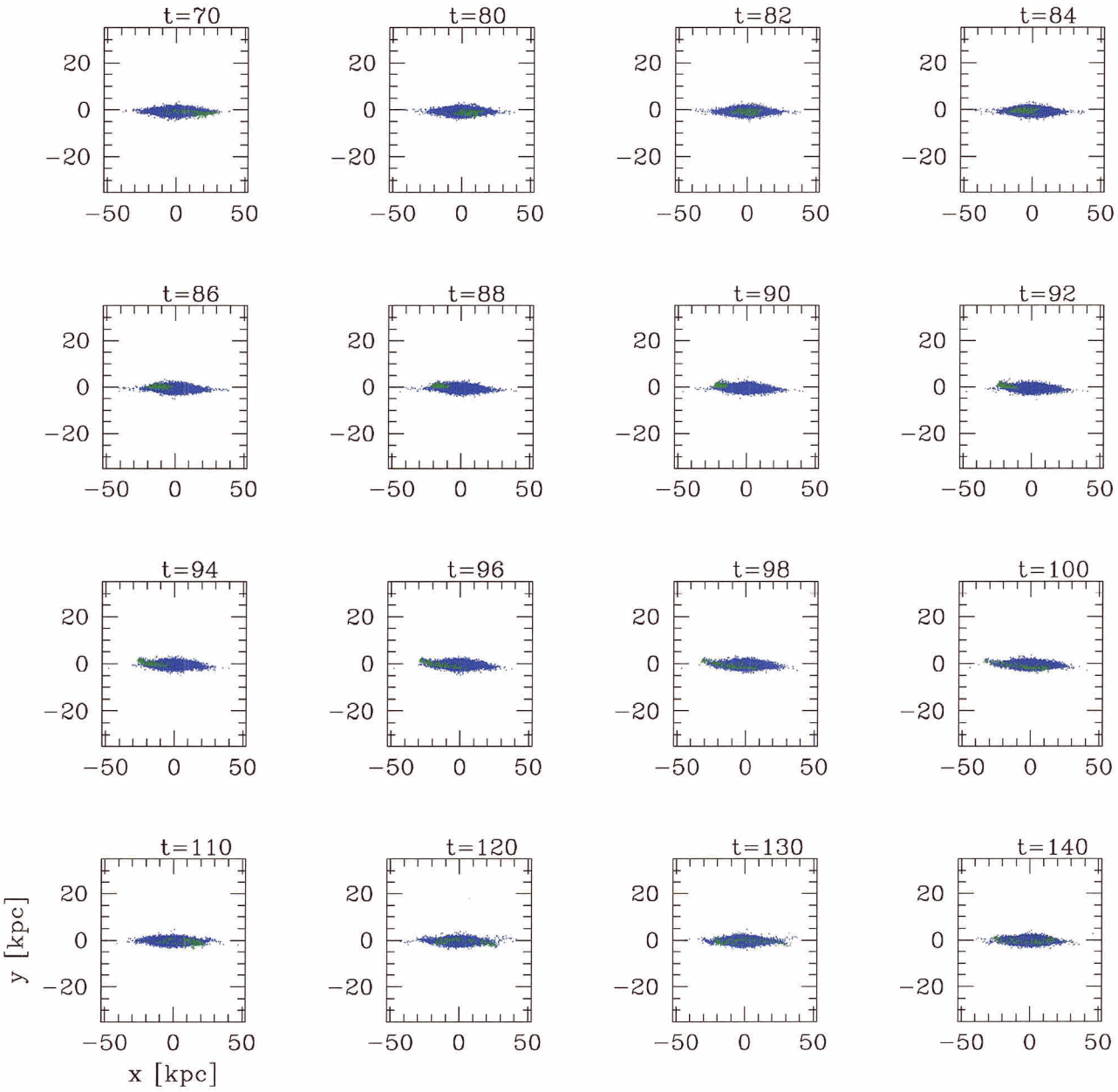

Figure 5.24: The evolution of the warp triggered by the first pericentric passage of satellite GS2, projected onto the (x, z) plane. Warp particles are identified based on their highest $< \tau_r >$ values. The warp particles are shown in green colors and the remaining disk particles are shown in blue colors.

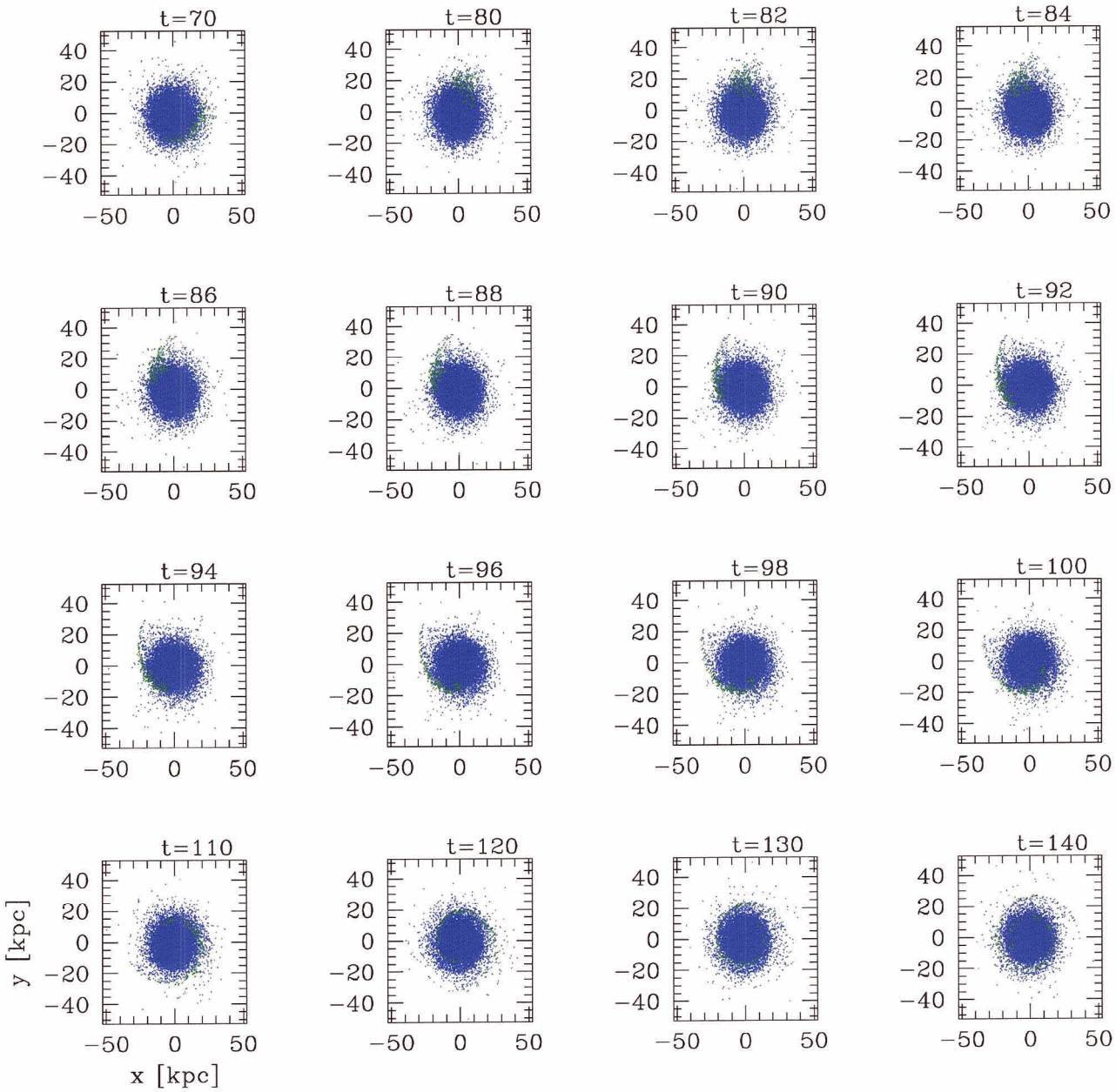

Figure 5.25: The evolution of the warp particles triggered by the first pericentric passage of satellite GS2, projected onto the plane of the disk (x, y). Warp particles are identified based on their highest $< \tau_r >$. The warp particles are shown in green colors and the remaining disk particles are shown in blue colors. After the time $t \sim 92$ – when the warp reaches its maximum height – the warp particles disperse rapidly in the (x, y) plane, due to differential rotation.

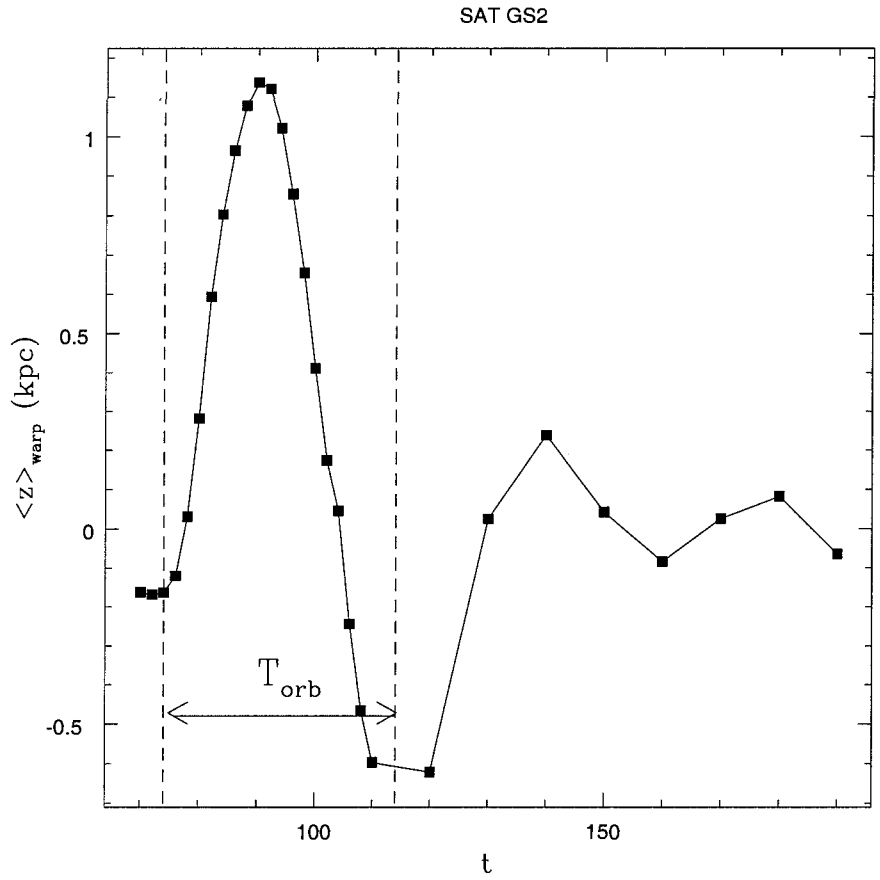

Figure 5.26: The evolution of the z -component of the center of mass of the warp formed at the first pericentric passage of satellite GS2. The orbital time in this region of the disk, $T_{orb} \simeq 40$, is also shown for comparison.

Figure 5.26 follows this more quantitatively, by tracking the evolution of the z -component of the center of mass of the warp particles between the time of interaction, $t \sim 70$, and long after the interaction has subsided, $t \sim 200$. The warp is seen to reach a maximum height of ~ 1.2 kpc at $t = 92$. Later, particles in the warp oscillate vertically and dissipate quickly in about one orbital period, $T_{orb} \sim 40$. This suggests that the warp can disperse very rapidly compared with the lifetime of the galaxy, making this a mechanism that may only explain short lived warps.

In conclusion, we have found a clear connection between the impulsive tidal torque due to satellite encounters with the disk and the formation of asymmetric, transient warps. In the next Subsection we will use this method for identifying the warps generated by close-by passages of satellites GS1 and GS2.

5.3.2 The Recurrence of Tidal Interactions

Once a method for identifying the stellar warps has been chosen, we can apply it to the study of other tidal interactions. In §5.3.1 we analyzed the first pericentric passage of the satellite GS2. We have argued before that this particular passage is the most significant among the other pericentric passages of the satellite GS2. This is further confirmed in Table 5.1, which shows the radial components of the tidal torques induced by the satellite GS2 near different pericenter passages. The radial torques $< \tau_r^{'} >$ are averaged values of $< \tau_r >$ over the number of particles selected in the warp, $< \tau_r^{'} >= (\Sigma_1^{N_{cut}} \tau_r)/N_{cut}$, in order to quantify the average radial torque per particle. Table 5.1 shows that the radial torque $< \tau_r^{'} >$ during the first passage of GS2 is at least a factor of 10 larger than at any other passage. The last two columns in Table 5.1 are the bound mass at pericenter, M_{bound}, and the the total distance from the center of the disk at pericenter, d, of satellite GS2. The results show that, given that the pericenter radius does not have large variations, the mass loss of the satellite is the main factor in the decrease in the importance of the warp.

Satellite GS1 is also expected to generate the most conspicuous warps, given its mass. The orbit of this satellite are illustrated in Figure 5.27. In this figure the radial distance of the satellite with respect to the center of the disk, $R = \sqrt{x^2 + y^2}$, is plotted in full line; the dotted line represents the vertical distance of the satellite from the plane of the disk, z; and the dashed line represents the total distance d. Table 5.2 contains the similar radial components of the tidal torques per particle, $< \tau_r^{'} >$, induced by the satellite GS1 near its pericenter passages. The radial torque occurring at passage VII is the strongest among all. In decreasing order of importance, follow

Table 5.1: Parameters of satellite GS2 near different pericenter passages.

pericentric passage	t_{peri} [code u.]	$< \tau_r' >$ [10^{-7} $M_\odot$ kpc $\frac{km}{s^2}$]	M_{bound} [10^{10} $M_\odot$]	d [kpc]
I	74	2.7	1.57	35.01
II	200	0.07	0.78	27.35
III	320	0.08	0.50	30.35
IV	430	0.19	0.34	46.65
V	530	0.27	0.26	18.12
VI	610	0.07	0.21	15.19
VII	690	0.02	0.16	39.91

passages VIII, IV, VI, III and V. The remaining passages of the satellite GS1 have radial torques of the same order as the passages II - VII of the satellite GS2 (and do not induce any obvious warps).

Table 5.2: Radial torques induced by the satellite GS1 near different pericenter passages.

pericenter passage	t_{peri} [code u.]	$< \tau_r' >$ [10^{-7} $M_\odot$ kpc $\frac{km}{s^2}$]	M_{bound} [10^{10} $M_\odot$]	d [kpc]
I	110	0.2	8.44	91.29
II	350	0.5	6.82	84.48
III	555	1.3	4.97	55.73
IV	690	3.4	3.91	40.89
V	780	1.1	3.15	71.66
VI	890	2.1	2.09	33.70
VII	950	12.6	1.68	28.66
VIII	1020	4.5	1.02	29.29

Note that radial torques will be maximum when the satellite is near pericenter, but also above or below the plane of the disk. Indeed, the radial torque is zero when the satellite crosses the plane of the disk. This means that the largest warps will be triggered when the pericentric passage coincides with the closest approach to the rotation axis: i.e., if $R = 0$ when d is minimum.

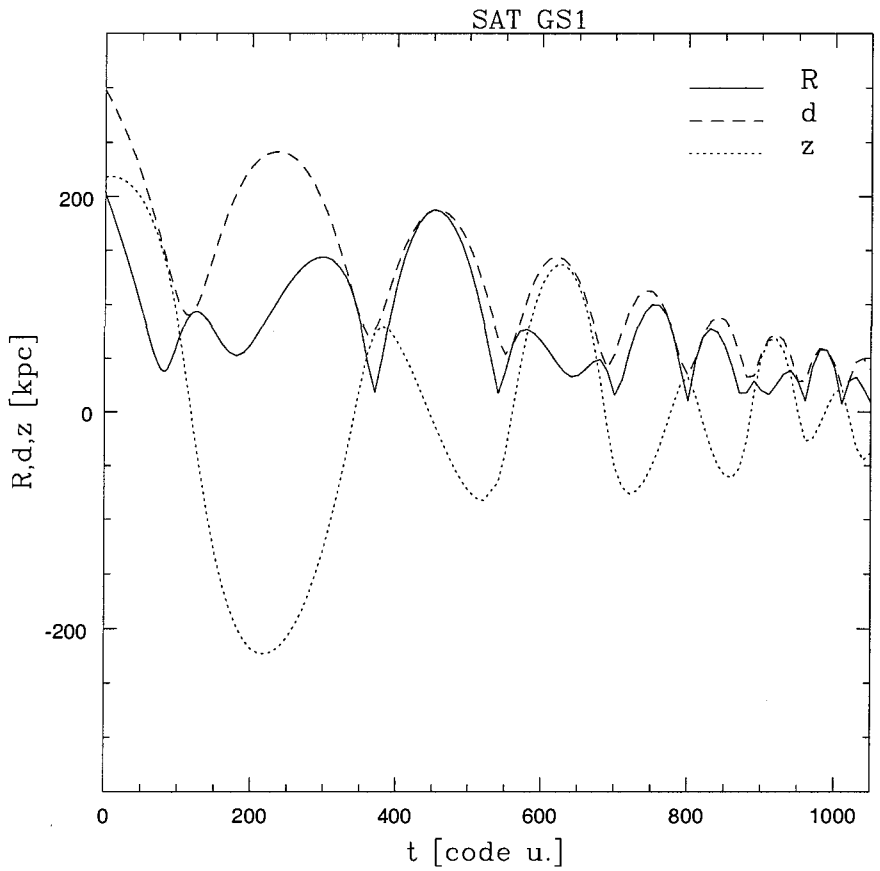

Figure 5.27: Orbital evolution of satellite GS1. Full line denotes the projected distance R in the plane of the disk, between the center of the mass of the satellite and the center of the disk; the dotted line represents the vertical distance z from the plane of the disk; and the dashed line represents the total distance d.

The two last columns in Table 5.2 are, as in Table 5.1, the bound mass at pericenter, M_{bound}, and the pericenter radius, R_{peri}, of the GS1 satellite. In this case the pericenter radius has large variations, specifically, it decreases sharply after $t \geq 500$. Although GS1 losses a significant amount of mass, the radial torques inducing the warps increase in importance, because the pericenter radius becomes very small (This is in contrast with the case of satellite GS2, where we saw that the mass loss correlates

more closely with the importance of the radial torques).

We further apply the warp identification algorithm described in §5.3.1 to all near-disk passages of satellites GS1 and GS2 and select the first $N_{cut} = 100$ particles with the highest $< \tau_r >$ values. We find that asymmetries in the $S_x - S_y$ plots are clearly defined for most of the satellite GS1's passages and only for the first passage of satellite GS2.

Figure 5.28 shows the identification of the particles in the warps formed as a result of passages IV (top panel) and VII (bottom panel) of the satellite GS1, based on their highest $< \tau_r >$. These particles also form asymmetric features (i.e. "plumes") in the $S_x - S_y$ plane. Figure 5.29 shows that, at the peak of the tidal interaction, an asymmetric feature[5] forms also in the angular momentum $\Delta L_x - \Delta L_y$ plot. The time intervals used for calculating the gradients of the angular momenta are: $\Delta t = 710 - 680$ (for passage IV) and $\Delta t = 960 - 930$ (for passage VII), respectively.

Figure 5.30 shows the identification of the particles in the warp formed as a result of passage VIII of satellite GS1, based on their highest $< \tau_r >$. These particles also form a $S_x - S_y$ "plume". Figure 5.31 shows that warp particles triggered at this passage also stand out in the $\Delta L_x - \Delta L_y$ plot. The time interval used for calculating the gradients of the angular momenta is: $\Delta t = 1010 - 1030$, which encompasses the time of maximum radial torque exerted by this passage of satellite GS1.

[5]Particles pertaining to the $\Delta L_x - \Delta L_y$ " plumes" generally correspond to the ones with the highest $< \tau_r >$, and implicitly, to the warp. As explained before, the method of warp identification based on the highest $< \tau_r >$ is more reliable than the one based on $\Delta L_x - \Delta L_y$ asymmetries. The plots including the $\Delta L_x - \Delta L_y$ " plumes" are shown mainly for illustrative purposes.

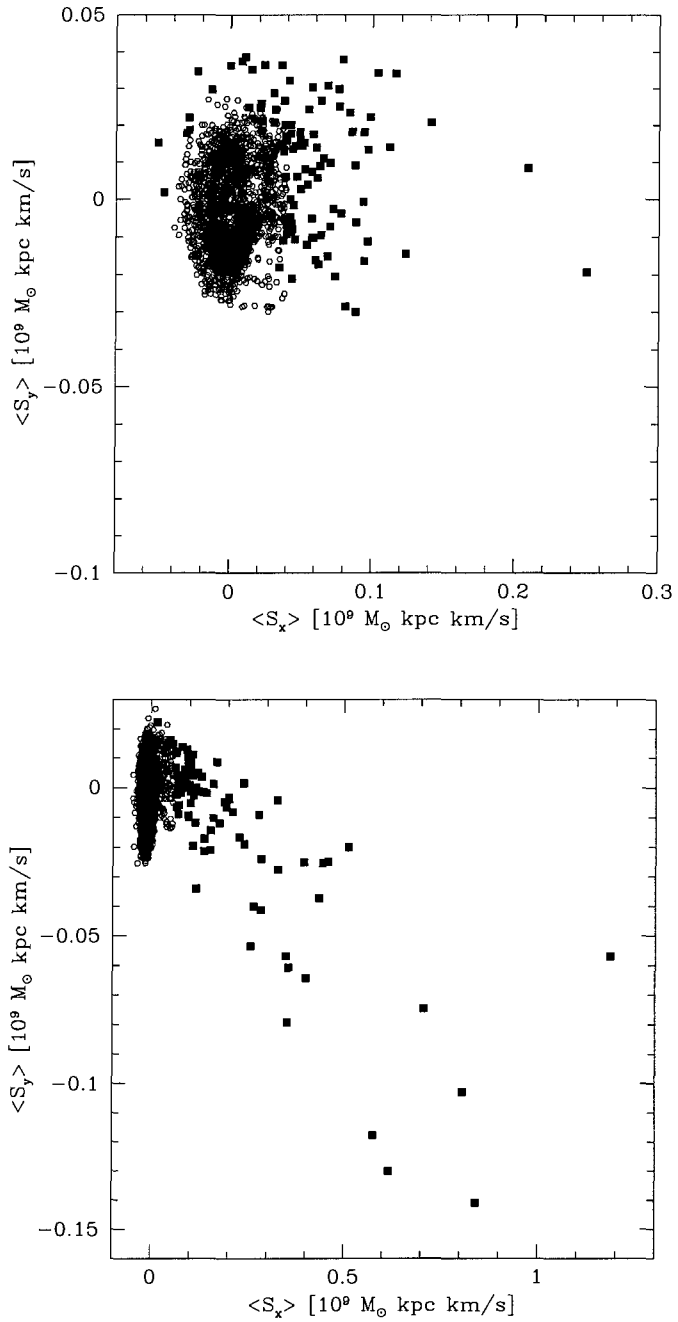

Figure 5.28: Identification of the particles in the warps formed as a result of passages IV ($t \sim 690$) - top panel - and VII ($t \sim 950$) - bottom panel - of the satellite GS1, based on their highest $< \tau_r >$. These particles form asymmetric features in the $S_x - S_y$ locus and are highlighted with full squares. For clarity, only the disk particles in the outer part of the disk – beyond $R > 17.5$ kpc – are plotted (these are shown with empty circles).

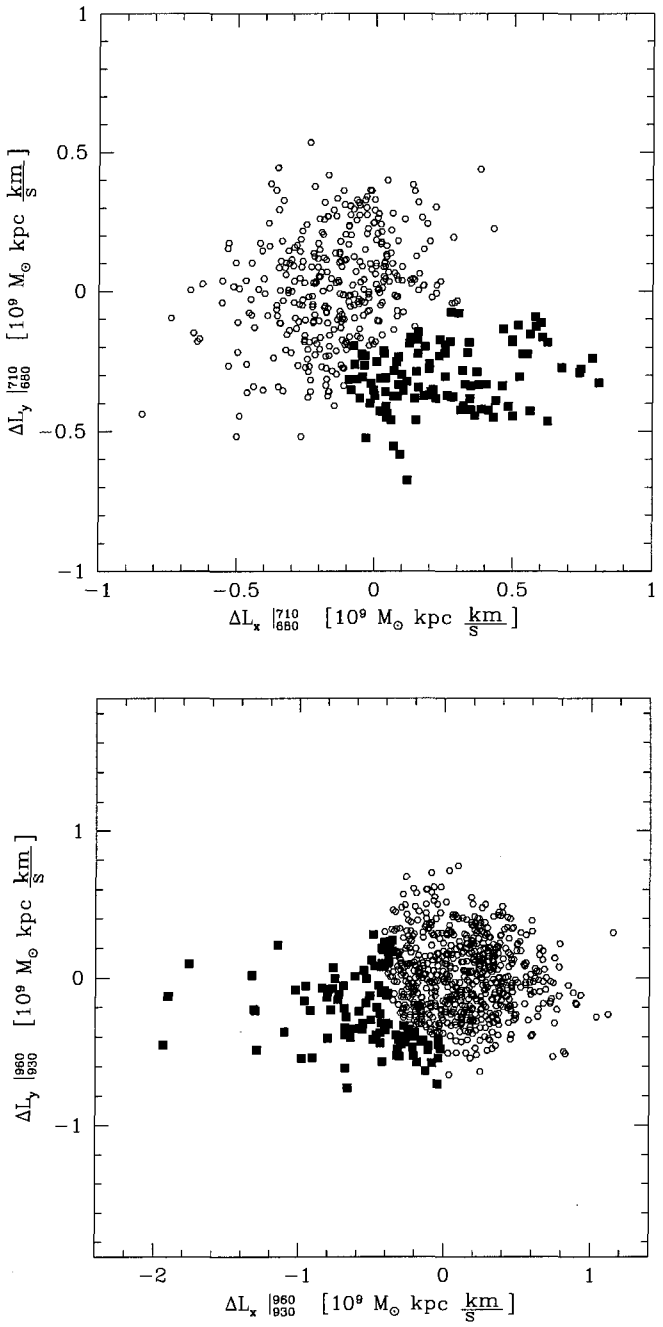

Figure 5.29: At the peaks of tidal interactions during passages IV ($t \sim 690$) - top panel - and VII ($t \sim 950$) - bottom panel - of satellite GS1, similar "plumes" form in the angular momentum $\Delta L_x - \Delta L_y$ plots. For clarity, only the disk particles in the outer part of the disk – beyond $R > 17.5$ kpc – are plotted (these are shown with empty circles). The asymmetric features are highlighted with full squares.

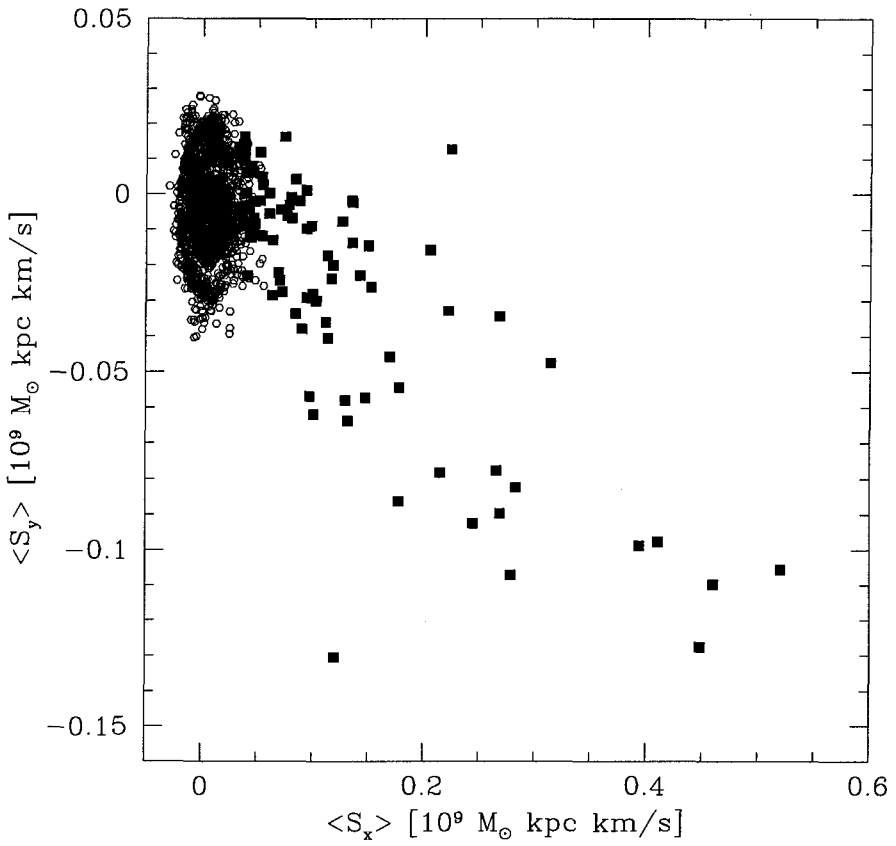

Figure 5.30: Identification of the particles in the warp formed as a result of passage VIII of satellite GS1, based on their highest $< \tau_r >$ (these particles also stand out in a $S_x - S_y$ "plume"). The symbols are the same as in Figure 5.28. For clarity, a circle is drawn to delineate the warp candidate particles.

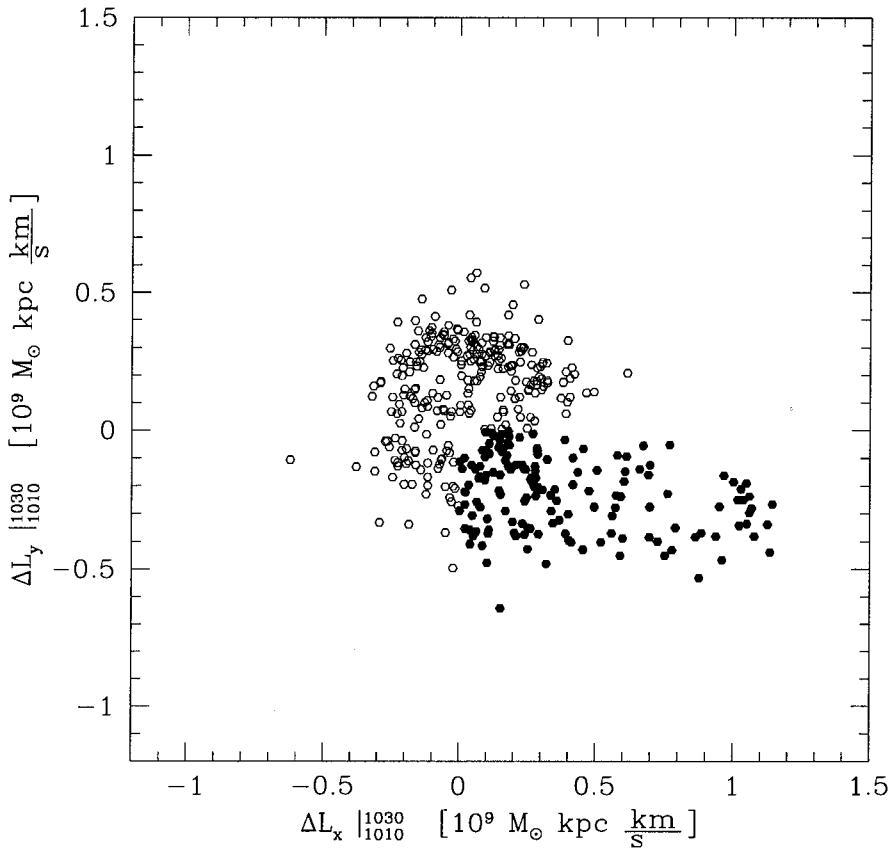

Figure 5.31: At the peak of the tidal interaction ($t = 1010 - 1030$), an asymmetric feature forms also in the angular momentum $\Delta L_x - \Delta L_y$ plot. The symbols are as in Figure 5.29.

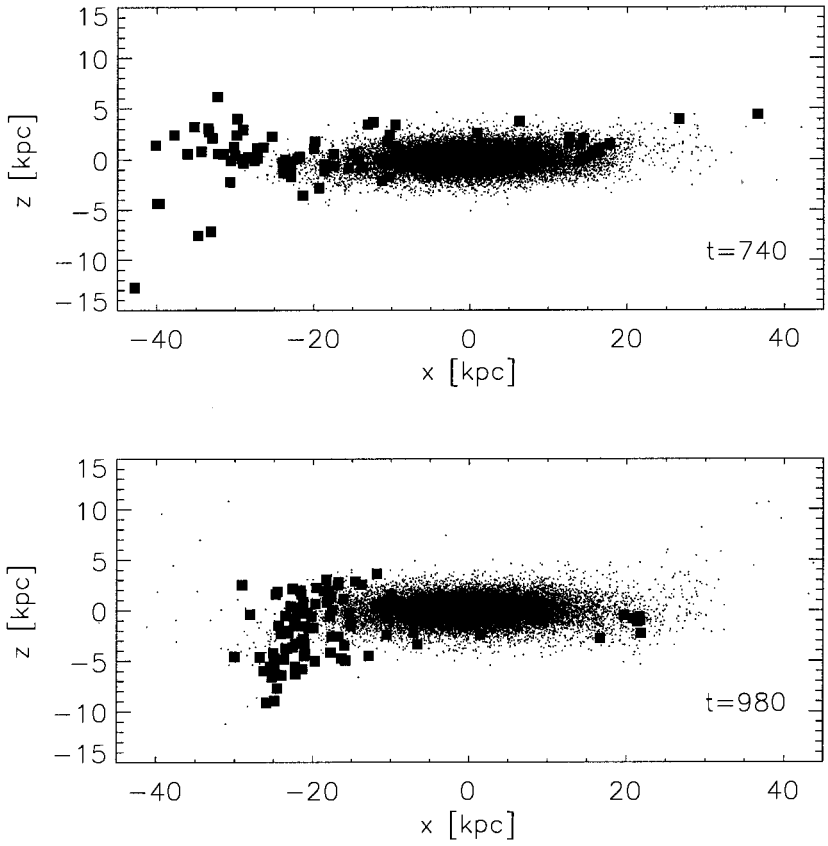

Figure 5.32: Edge-on view of the disk and the particles identified in the warps triggered by the passages IV (top panel) and VII (bottom panel) of the satellite GS1. The snapshots are taken around the time of maximum height of the warp ($t = 740$ and $t = 980$, respectively). The particles with the highest $< \tau_r >$ (and pertaining to the the $S_x - S_y$ "plumes") are highlighted in filled squares. The figure shows that the particles selected by this method are later retrieved in the warp.

Figure 5.32 shows two examples of warps formed during two passages of the satellite GS1 (passage IV – top panel; and passage VII – bottom panel). The identification of the warp particles was made based on their highest $< \tau_r >$. Figure 5.32 shows the edge-on view of the disk at the times when the warps attain their maximum heights: at $t = 740$ and $t = 980$, respectively. The particles with the highest $< \tau_r >$ (and located in the $S_x - S_y$ "plumes") – highlighted here with filled squares – are proven to be part of the warps.

5.3.3 The Warp Strength

So far, we have shown that the disk-satellite tidal interactions are able to trigger stellar warps. We are further interested to know whether these warps resemble in size the warps observed in the Universe. Also, since the results derived so far are strongly dependent on the particular orbital parameters of the satellites under study, we would like to formulate a general model from which one can derive predictions for other satellites.

First, we investigate the relation between the height of the warps and the strength of the tidal interactions in our simulation. Figure 5.33 shows the height of the warps, $< z >_{max}^{warp}$, versus the radial torque τ_r, both values measured at their corresponding maxima. The height of the warp is averaged over the number of particles selected in the warp (i.e $N_{cut} = 100$). As expected, the satellite passages with the highest radial torques induce the more significant warps. For example, the uppermost point in the plot ($< z >_{max}^{warp} \simeq 2.1$ kpc) corresponds to the strongest radial torque exterted by satellite GS1 which occurs around $t \sim 950$ (the warp, however, will not reach its maximum height until time $t = 980$).

As explained before, not all satellite passages result in conspicuous warps (that is, with the height of the warp being comparable with the average disk height at that radius). In Figure 5.33, we have excluded those passages for which their warps were inconclusive. Thus, the plot includes only the first passage of the satellite GS2

(denoted by the empty triangle) and several passages of the satellite GS1 (denoted by the full squares).

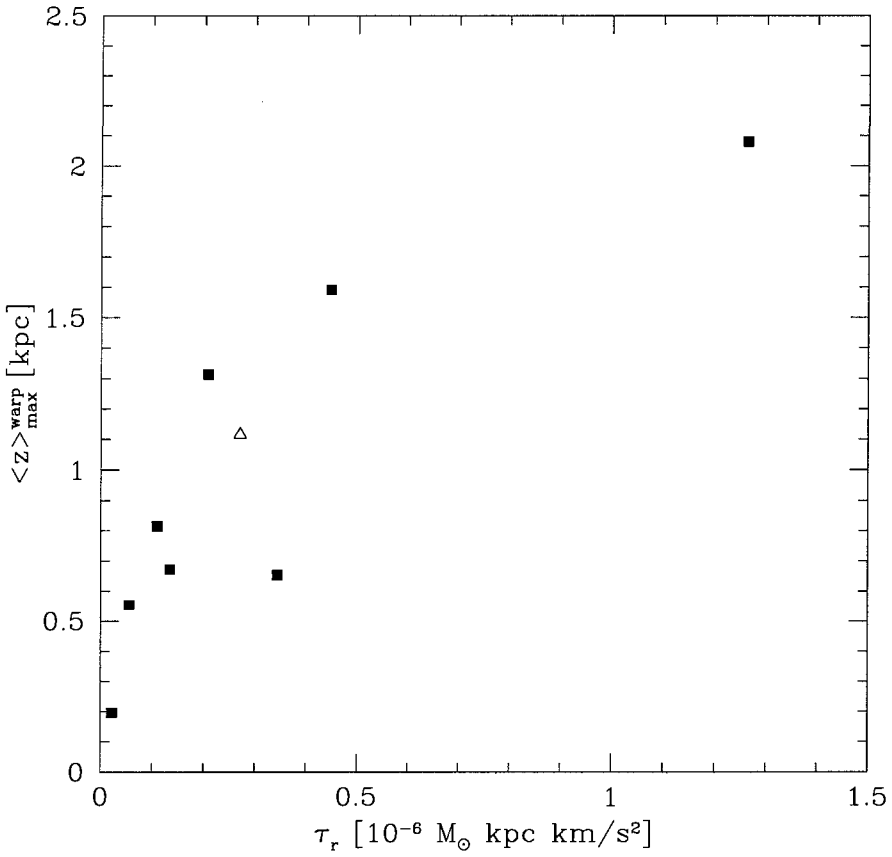

Figure 5.33: The warp heights $< z >^{max}_{warp}$ measured at the time of maximum height increase with the value of radial torque τ_r. The full squares denote passages of satellite GS1 and the empty triangle denotes the first passage of satellite GS2.

The result in Figure 5.33 confirms the basic expectation that the warp height must increase with the strength of tidal interactions (measured here by the radial torque).

$$< z >^{warp} \sim \Delta L_r \sim \tau_r \tag{5.7}$$

However, the maximum height of the warp (in other words, the amplitude of

oscillatory motion on the vertical direction) cannot be infinitely large. If the period of vertical excursions of warp particles, T_z, becomes much larger than the orbital period T_{orb} at that radius in the disk, the particles may become unbound from the disk.

The scaling in Figure 5.33 may be used to infer (or predict) warp heights triggered by any given satellite for which one can estimate its tidal strength (specifically, the magnitude of the radial torque τ_r of the satellite at its maximum strength). In §5.4.2 we will comment on a few examples of observational warps and their relation with neighboring satellites.

As a final note, we mention that the stellar warps obtained in our simulation have typical heights between $1 - 2$ kpc. Warps with higher amplitudes may, however, be created in simulations with different initial conditions for the satellite orbits. Also, the maximum height of the warp should depend on the detailed properties of the disk as well; more massive disks will exert a stronger restoring force which will lead to smaller warps for the same torque. Our results should in principle be restricted to the Milky Way or similar galaxies.

Comparisons of these results with a few cases of observational data will be presented in §5.4.

5.3.4 Differential Tilting of the Disk. Implications for Warp Lifetimes.

Warps are a particular case of differential tilting[6] of the disk. They are also thought to be short lived phenomena, due to the short winding time of such perturbations in disk models (BT87, p. 346). This is clearly the case for warps in our simulations, which damp typically in less than one orbital time (The particular example discussed

[6]The warped region of the disk is expected to be characterized by a distinct direction of its total angular momentum, compared with the unwarped region. We can therefore understand the warp properties in terms of the misalignment of angular momenta within the disk.

in §5.3.1 is not an isolated case. The same conclusion applies to all the other warps we have analyzed.). However, even if the warp itself disappears as a distinct feature, some forms of differential tilting of radial bins in the disk may persist through the "superposition" of several tidal interactions between the satellite and the disk. This phenomenon is most likely to occur at later stages in the evolution of a massive satellite, as in the case of satellite GS1. As Figure 5.27 shows, as satellite GS1 sinks towards the center of the galaxy, its orbital period shortens due to dynamical friction and the disk-satellite tidal interactions occur more frequently.

The short lifetime of warps has always been intriguing. If warps are truly short lived, the likelihood of observing them is low, unless, of course, the warp-triggering mechanism occurs very frequently or warps are "locked in" by some unknown mechanism. The recurrence of disk-satellite interactions may provide a mechanism capable of increasing the persistence of these warps.

To investigate this proposed mechanism, we study the differential tilting of the disk at different times. For our analysis, we choose to divide the surface of the disk in several equidistant radial bins and follow the evolution of their inclination angles θ, defined as:

$$\theta \equiv \theta_{bin} = \theta(\vec{L}_{bin}, \vec{L}_{d,tot})$$

Here the angle θ represents the angle subtended by the total angular momenta of each radial bin, $\vec{L}_{bin}$, and the total angular momentum of the disk, $\vec{L}_{d,tot}$. (Note that these measurements are made after the correction for the global disk tilt. That is, the direction of $\vec{L}_{d,tot}$ is perpendicular to the plane of the disk.)

Various symbols in Figure 5.34 denote the inclination angles of the radial bins in the disk at several times around the pericentric passages at which the warps are more pronounced: $t \simeq t_{peri}$: 580, 820, 980, 1040. For comparison, the empty squares denote the initial inclination of the radial bins (t=0). Two general trends can be inferred from Figure 5.34:

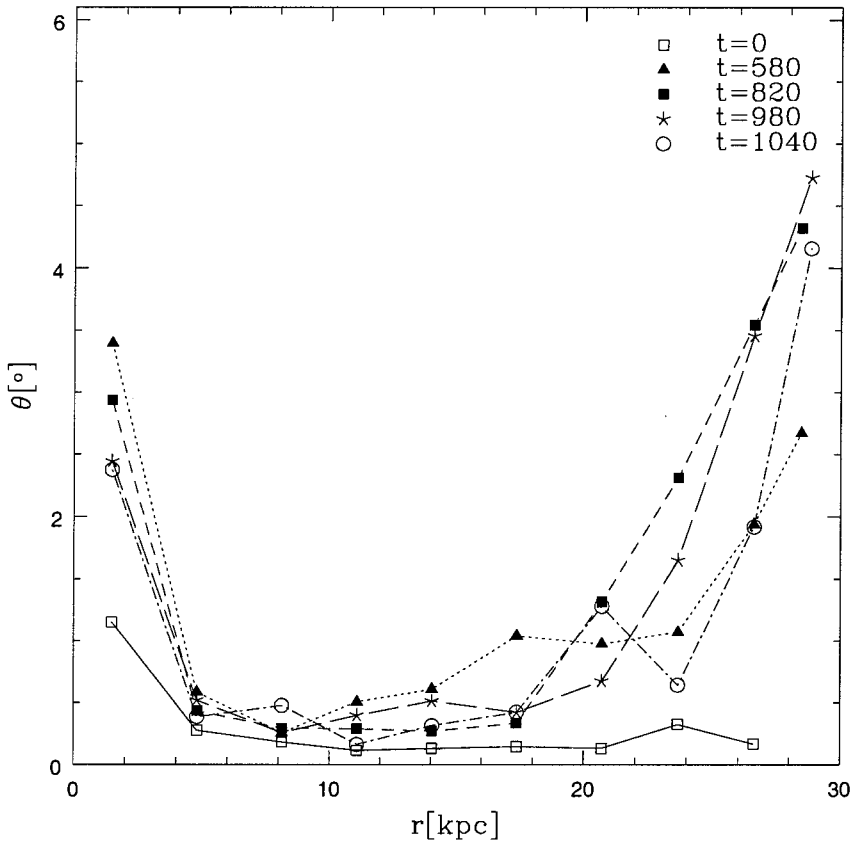

Figure 5.34: Inclination angles, $\theta(\vec{L_{bin}}, \vec{L_{d,tot}})$, of equidistant radial bins in the disk, plotted at different times $t \simeq t_{peri}$: 580, 820, 980, 1040. For comparison, the empty squares denote the initial inclination of the radial bins (t=0).

– firstly, at any fixed time snapshot, the outer radial annuli ($R > 15$ kpc) tilt progressively with radial distance R in the disk. For example, at $t = 980$: $\theta(20\,\mathrm{kpc}) = 0.5^o$ and $\theta(30\,\mathrm{kpc}) = 5^o$ (i.e. an increase by a factor of 10). The inner radial annuli ($R \leq 15$ kpc) are almost leveled with the plane of the disk – with the exception of the first few bins with relatively large θ values (however, these large values are caused by the increased velocity dispersion of the stars near the center of the disk and, therefore, do not reflect a real tilting). For the times t considered here, the outermost region of the disk ($R \simeq 30$ kpc) becomes misaligned by $\sim 3.5 - 5^o$ relative

to the inner disk. We note briefly that these values are in excellent agreement with typical values for the inclination angles inferred from observations of warped galaxies (for example, Reshetnikov et al. (2002) find $\theta \sim 3 - 4^o$).

– secondly, for a fixed radius, the angle θ fluctuates with time. For example, at the radius $R = 25$ kpc, the pericentric passages occurring around $t = 820$, 980 leave more steeply inclined warps than the passages around the times $t = 580$, 1040.

The aspect of inclination of outer parts of the disk is further investigated in Figure 5.35. Here we concentrate on the inner (10 kpc $< R <$ 15 kpc) and outer (25 kpc $< R <$ 30 kpc) regions of the disk, for which we calculate their specific inclination angle:

$$\theta_{inner} = \theta(\vec{L}_{inner}, \vec{L}_{d,tot}),$$

and

$$\theta_{outer} = \theta(\vec{L}_{outer}, \vec{L}_{d,tot}).$$

Here $\vec{L}_{inner}$ and $\vec{L}_{outer}$ represent the total angular momentum of the inner and outer regions of the disk, respectively, and $\vec{L}_{d,tot}$ is the total angular momentum of the disk. The new parameter θ characterizes the vertical changes in the disk (warps included) globally, at any given t.

The top panel in Figure 5.35 shows the evolution of the inclination angle θ for the two regions in the disk (that is, θ_{inner} and θ_{outer}) with time. If the warps were much shorter lived than the time between two successive interactions, we would expect the angle θ to peak around each pericentric passage and then settle back to the initial values $\theta \simeq 0$. Figure 5.35 (top panel) shows that this is indeed the behavior of θ_{inner}. However, θ_{outer} has a different behavior: in particular, after the time $t \sim 600$, θ_{outer} begins to increase continuously. This suggests that vertical perturbations in the disk are strengthened by later passages with predilection in the outer region of the disk.

How does this result compare with the timing of tidal interactions? The bottom

panel in Figure 5.35 shows the time interval T between successive pericentric passages of the satellite GS1. The values T decrease continuously, from about $T = 250$ for the first few pericentric passages to about $T = 50$ after $t = 1000$ code u. $= 13$ Gyr of evolution. The relation between the top and bottom panels in Figure 5.35 is given by a key parameter: the orbital time in the disk, T_{orb}. The orbital period at the radius where the warp occurs is an important time scale for the lifetime of the warp. Thus, if the orbital period of the disk stars is longer than the dynamical time of the interaction, the warp is prone to quickly disappear. On the contrary, if the orbital period is short, the warp will be able to be carried on for several orbits and, therefore, increase the likelihood that it would be observed. Figure 5.35 (bottom panel) shows that at the time $t \sim 600$, the time interval T becomes shorter than the orbital period at the radius in the disk where the warp resides ($T_{orb} = 122$ code u. $\simeq 1.54$ Gyr at $R \sim 25$ kpc).

For $T < T_{orb}$, contributions from successive passages may become superimposed. Individual warps may be therefore difficult to separate and they may contribute to the global differential tilting of the disk. Note that a "superposition" of different warps may not necessarily enhance an existing warp or create a new one; however, vertical disturbances in different regions of the disk may survive for a longer time. These, in effect, may be indistinguishable in an observational study from other asymmetric, irregular warps. One observational consequence of this mechanism would be the increase of the likelihood of detection of these vertical disturbances in the disk.

Finally, we note that this mechanism is expected to operate when a given satellite achieves an orbital period of similar magnitude with the orbital period in the disk, $T \simeq T_{orb} \simeq 1$ Gyr. This satellite would have to be in a late stage of orbital decay. A notable example of a satellite with orbital period of about 1 Gyr is the Sagittarius dwarf galaxy. It is therefore not unlikely that the warp in the Milky Way or other vertical disturbances in the disk are not the result of a "one time tidal interaction" with this satellite companion, but a superposition of events correlated with its last few close passages.

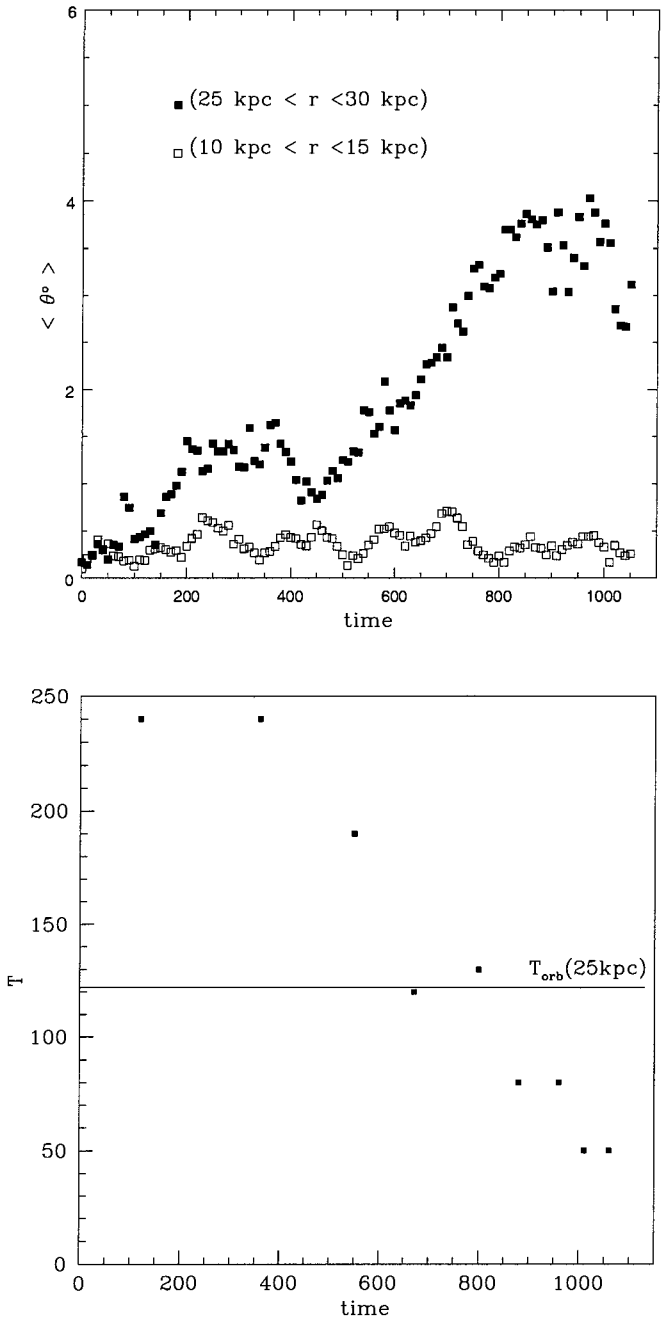

Figure 5.35: *(Top panel)* With filled squares: the inclination angle between the angular momenta of the outer region of the disk and the total angular momentum of the disk, $\theta(\vec{L}_{outer}, \vec{L}_{d,tot})$; empty squares: same angle corresponding to the inner region of the disk, $\theta(\vec{L}_{inner}, \vec{L}_{d,tot})$. *(Bottom panel)* the time T between two successive pericentric passages of GS1 compared with the orbital period, T_{orb} at 25 kpc.

5.3.5 Predicted Signatures in the Line-of-Sight Velocity Curves

The next step in our analysis is to derive some observational predictions for our model. We turn our attention to a method which is commonly used for detected structural changes in the galaxy disks: the measurement of line-of-sight velocities ($v_{l.o.s}$). In particular, we investigate whether the warps formed in our simulation could leave measurable signatures in the $v_{l.o.s}$ curves.

Figure 5.36 shows the $v_{l.o.s}$ curve of the stellar disk plotted at several times around the first close passage of the satellite GS2 (in the time interval $t \sim 70 - 140$). The line-of-sight velocity, $v_{l.o.s}$, is calculated as the average parameter $< v_y >$ of the disk particles in a given x-bin. The disk is viewed edge-on (i.e. the y axis is oriented along the line of nodes), projected in the (x, z) plane, and the radial distance is measured along the x-axis. The $v_{l.o.s}$ is calculated separately for both sides of the disk: the empty squares in Figure 5.36 represent the values along the positive side of the x-axis ($x > 0$) and the full squares denote the values for the negative side of the x-axis ($x < 0$). The location of the warp at the time of its maximum height ($t = 92$) is marked in this figure by an arrow and also highlighted with a full hexagon. The result in this figure implies that the warp triggered in the first passage of the satellite GS2 leaves no obvious signature in the $v_{l.o.s}$ curve: indeed, the warp has a similar $v_{l.o.s}$ as the unwarped, inner regions of the disk.

The warp studied in Figure 5.36 has a amplitude of about $< z >^{warp} \sim 1$ kpc. Would more developed warps be detectable in the line-of-sight velocity curves? To answer this question, we investigate the warp formed in the seventh pericentric passage of the satellite GS1 (i.e. passage VII) – which is the passage with the strongest tidal interaction. The warp triggered during this passage attains a maximum height of $< z >^{warp} \sim 2.1$ kpc at the time $t \sim 980$. In Figure 5.37 we plot the line-of-sight velocity curves of the stellar disk at equal time intervals $\Delta t = 10$ in the time span $t = 920 - 1000$. The empty and full squares are used as in Figure 5.36. The location of the tidally triggered warp at the time of its maximum height, $t = 980$, is marked

by an arrow and also highlighted with a full hexagon. As Figure 5.37 shows, the warp triggered in this passage does not leave an obvious signature in the $v_{l.o.s}$ curve either.

In conclusion, our results predict that the warps triggered by tidal interactions would be difficult to detect in the $v_{l.o.s}$ curves. Therefore, the observational results which report flat velocity curves need to be interpreted with caution before ruling out the presence of warps. As our simulation results show, warps as high as ~ 2 kpc could remain hidden in this type of plot. We note that this problem has been previously alluded to by Sancisi (1976), in the study of the warped galaxy NGC5907. The radial extent of the HI (21 cm) gas in this galaxy is about 100 kpc. The HI warp reaches a height of about 8 kpc above the optical disk, at a radius of about 40 kpc. Even though for this warp is spectacular in size, the circular velocity curve is seen to remain flat up to the end of the optical disk (Sancisi 1976). A possible interpretation is that the warp is mainly caused by changes in L_x and/or L_y components, whereas rotation curves of edge-on galaxies only measure L_z, which is not drastically changed in a tidally triggered warp.

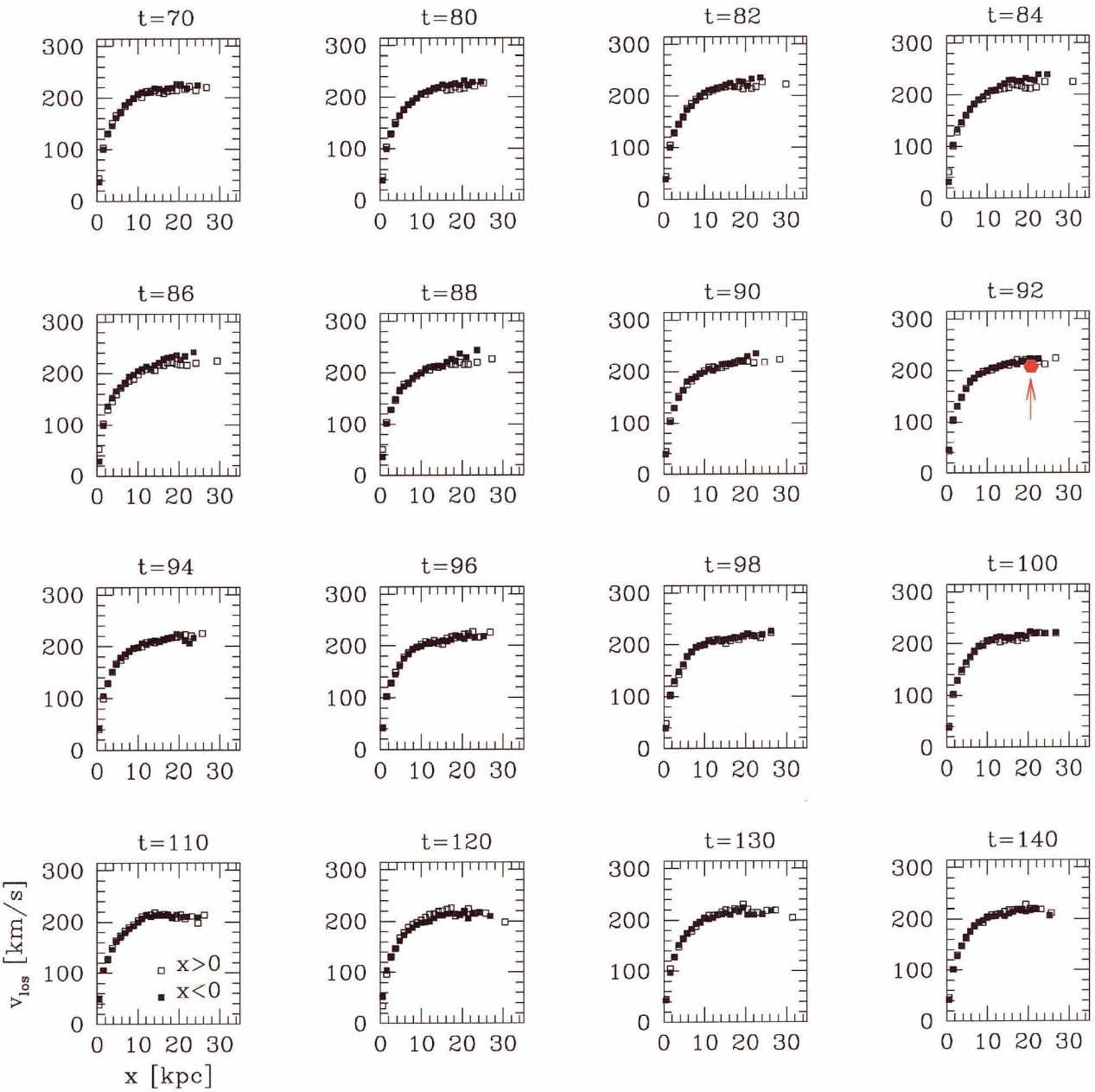

Figure 5.36: Signature of the warp in the line-of-sight velocity curve of the disk, around the first passage of satellite GS2 – plotted for the time interval $\Delta t = 70 - 140$. The values obtained for the positive side of the disk ($x > 0$) are plotted with empty squares and those for the negative side ($x < 0$), with full squares. The location of the warp at the time of its maximum height ($t = 92$) is marked in the figure by an arrow and also highlighted with a full hexagon.

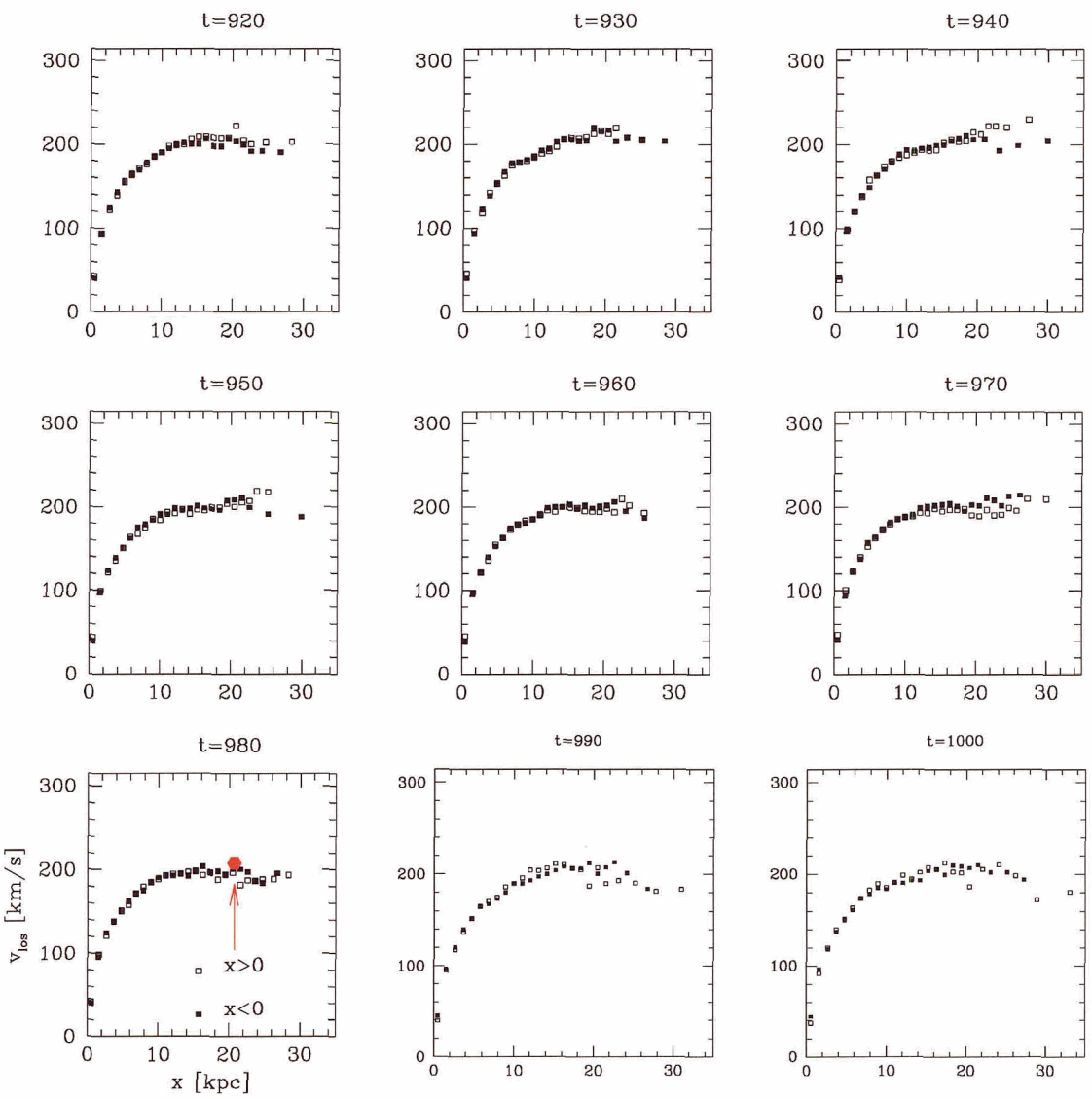

Figure 5.37: Signature of the warp in the line-of-sight velocity curve of the disk, around the passage VII of the satellite GS1 – plotted for the time interval $\Delta t = 920 - 1000$. The values obtained for the positive side of the disk ($x > 0$) are plotted with empty squares and those for the negative side ($x < 0$), with full squares. The location of the warp at the time of its maximum height ($t = 980$) is marked in the figure by an arrow and also highlighted with a full hexagon.

5.4 Observations of Tidally Triggered Warps

In the preceding sections, we have proposed and described a mechanism for generating warps by means of satellite-disk interactions. We ask now whether this is consistent with observational constraints. We will review the evidence from several surveys of external galaxies and then discuss in more detail a few cases of well known warped galaxies.

5.4.1 Surveys of External Galaxies

In recent years, several studies have been conducted with the specific purpose of determining the effect of tidal interactions on the warping of galaxies. One of the largest samples of disk galaxies was collected by Schwarzkopf & Dettmar (2000; 2001). This sample contains 110 highly inclined/edge-on spiral galaxies, of which 49 are interacting with satellite galaxies and 61 are in isolation. Only minor mergers were considered in the selection of the interacting galaxies ($M_{sat}/M_d \sim 0.05 - 0.2$). The authors find that, indeed, interacting galaxies are more frequently warped than non-interacting ones (Schwarzkopf & Dettmar 2001); about 93% of the interacting galaxies have well defined warped disks, in contrast with 45% of the non-interacting galaxies. However, all galaxies seem to have some sort of vertical distortion. In the case of interacting galaxies, the warps are also reported to be larger in amplitude (measured by the scale height of vertical perturbations). This study finds that, on average, warps in interacting galaxies are about 2.5 times larger than those in isolated galaxies (in terms of absolute values, the scale height of the vertical perturbations, i.e. warps, are 340 pc and 140 pc, respectively).

Other statistical studies also find a strong correlation between the presence of warps and that of companion galaxies. For example, two recent studies (Sánchez-Saavedra, Battaner & Florido 1990; Reshetnikov & Combes 1998) report that up to 80% of galaxies in their sample are warped. In agreement with the findings of Schwarzkopf & Dettmar (2001), these authors also find that the fraction of isolated

galaxies displaying warps is around $45 - 50\%$.

The fact that isolated galaxies display noticeable warps (and in a somewhat high percentage) is a surprising result. This suggests that tidal interactions are not the only avenue for generating warps and that alternative mechanisms need to be invoked (some of these mechanisms have been described in Chapter 1). There is, however, no doubt that tidal interactions play a significant role in the creation of warps: as all of the above observational studies show the interacting galaxies are not only characterized by a higher frequency of warps, but also, these warps are brighter and stronger than the ones in non-interacting galaxies.

One can also ask whether the presence of warps varies with the redshift. It is well known that in the model of hierarchical structure formation, the frequency of mergers is believed to have been higher in the past. Intuition then leads us to expect a higher frequency of warps at high redshift. Reshetnikov et al. (2002) investigated this problem by using a statistical sample of galaxies selected from both the North and South Hubble Deep Fields (HDF). The sample consists of 50 edge-on galaxies (32 of those are in the North HDF and 18 in the South HDF) with redshifts ranging from 0 to about 2. The average redshift for the galaxies in the sample is 0.89 (the corresponding average redshifts for the two fields are: 0.85 for HDF-N and 0.95 for HDF-S, respectively)[7]. From their results the authors infer that, indeed, the frequency of warps is higher at higher redshifts ($z \sim 1$) compared with that in the local universe ($z \sim 0$). In fact, they even suggest that all disk galaxies at $z \sim 1$ may be warped. Moreover, the amplitude of the warps is also observed to be stronger for high redshift galaxies than for local ones. These results support the idea that the formation of warps is influenced by the merger rate of the parent galaxies. They also suggest that, in the past, tidal interactions might have had a greater weight compared with other warp generating mechanisms.

In conclusion, observational studies suggest that the frequency of galaxy warps

[7]A redshift of $z \sim 1$ ensures the completeness of the sample. Beyond this redshift the warps are extremely difficult to detect, due to magnitude limitations (see Reshetnikov et al. 2002 for details).

depends strongly on the local environment of the disk. This result supports the model of tidally triggered warps described in the previous Section.

Can the observational data be compared with our numerical results? Unfortunately, a one-to-one comparison between our numerical model and the observational results is not possible, given the statistical nature of the satellite parameters in both cases. However, some general comparisons can be made. One regards the inclination angle of the warps. Reshetnikov et al. (2002), for example, find that the inclination angles of the warps are on average about 3 or 4^o. This value is similar with the θ values obtained in our "Halo 1" SCDM simulation (see Figure 5.34). Secondly, one can compare the amplitude of the warps: in the observations, the warps generated in galaxies interacting with satellites of masses $M_{sat}/M_d = 0.05 - 0.2$ are, on average, about 340 pc. The disk in our numerical simulation develops warps of average amplitudes as high as 2.1 kpc (this occurs at the time of its seventh passage of satellite GS1, characterized by a bound mass of $M_{sat} \sim 1.7 \times 10^{10} M_{\odot} \sim 0.3 M_d$ and pericentric radius of $R_{peri} \sim 29$ kpc).

5.4.2 Individual Cases of Warped Galaxies

One interesting application of our model is the study of a few specific cases of warped galaxies in the local universe, for which we do not have yet an explanation for the formation of their warps. Given that the information about the orbital parameters of satellite galaxies is very limited (i.e. there are poor constraints on the masses, pericentric radii and current 3D positions and velocities of most of satellite galaxies that we know of), our analysis will be valid only as an order of magnitude estimate.

Let us concentrate on a single star in the disk, assumed to have a mass of $\sim 1 M_{\odot}$ and with coordinates (x, y, z). The two torque components acting on this star in the plane of the disk are:

$$\tau_x = y\, F_z - z\, F_y$$
$$\tau_y = z\, F_x - x\, F_z,$$

where $\vec{F} = (F_x, F_y, F_z)$ is the gravitational force exerted on the disk star by a nearby satellite (considered a point mass). Assuming, for simplicity, that the height of the disk star above the plane of the disk is negligible, i.e. $z \simeq 0$, this implies that the radial torque can be approximated as:

$$\tau_r \simeq \sqrt{(x^2 + y^2)}\, F_z = R \cdot F_z = R \cdot \frac{G m_\star M_{sat}}{z_{sat}^2}, \tag{5.8}$$

where $m_\star$ is the mass of the disk star, R is its radial distance in the disk and M_{sat} is the mass of the satellite (note: by calculating the torque on a single star, we can make an easy comparison with the average radial torques $< \tau_r > \equiv \frac{\sum_{i=1}^{N} \tau_r}{N}$ calculated from numerical simulations in §5.3.2).

With the available information of the current mass of a given satellite and its relative position to the disk of the parent galaxy, we can estimate the radial torque exerted on a disk particle (taken, for convenience, to be located at the projection of the satellite onto the plane of the disk). Then, using the scaling obtained in Figure 5.33, we can estimate the height of the warp triggered by the interaction. We mention, however, that the average radial torque values in Figure 5.33 correspond to the interaction between a satellite and a *disk particle*, whereas in this section, we are estimating the interaction between the satellite and a *disk star*. Thus, in order to compare the latter values with the $< \tau_r >$ values presented in Figure 5.33, we need to multiply them by the ratio of the mass of a disk particle in our simulations and the mass of the disk star, $m_d/m_\star = 1.4 \times 10^6 M_\odot / 1 M_\odot = 1.4 \times 10^6$.

The Warp in the Milky Way. Is Sagittarius Dwarf Galaxy the Culprit ?

The Milky Way is one of the most well known warped galaxies. The warp of its gaseous disk has been detected about five decades ago using 21 cm observations of the neutral hydrogen (HI) (Burke 1957; Kerr 1957). The HI warp has an overall S -shape starting beyond the radius $R \sim 14$ kpc: on one side it extends up to the north at a Galactic longitude $l = 90^o$ reaching more than 2 kpc in size at $R = 20$ kpc; on

the other side it curves down to the south at $l = 270^o$ reaching < 1 kpc at a similar distance (Binney 1992). Recent observations indicate that the warp is seen in the stellar component as well (Miyamoto, Yoshizawa & Suzuki 1988; Freudenreich et al. 1994).

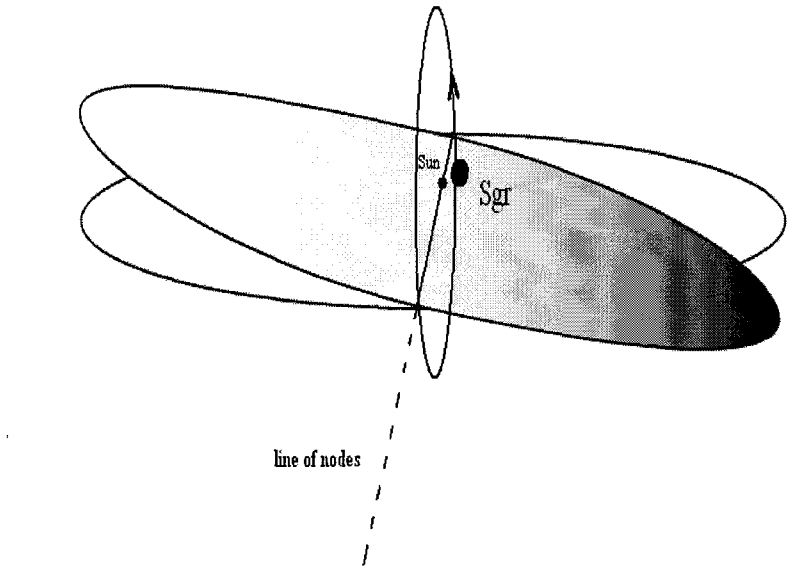

Figure 5.38: The orbit of the Sagittarius dwarf galaxy relative to the warped plane of the Galactic disk.

The origin of the warp is not yet fully elucidated. The recently discovered galaxy companion, the Sagittarius dwarf, is suspected to be the culprit. This satellite is in its late stages of tidal disruption. Currently, it is located at about 25 kpc from the Sun and a corresponding Galactocentric distance of ~ 16 kpc. Its highest density peak is just ~ 5 kpc below the Galactic disk, at $l = 6^o$ and $b = -15^o$, and it is approaching the disk (Ibata et al. 1994). From the shape of its tidal stream, we know that it must be moving on a polar orbit, perpendicular to the plane of the Galaxy (Lin 1996). Velázquez & White (1995) estimate that the satellite is on an orbit with a very short

period ($T_{orb} \simeq 760$ Myr) and with an apocenter of ~ 52 kpc and pericenter of ~ 10 kpc. Helmi & White (2001) obtain a similar result for their set of orbits. For example, the orbit shown in their Figure 1 has an apocenter of 68.3 kpc, a pericenter of 16.3 kpc and a radial period of ~850 Myr. Figure 5.38 illustrates the orbit of Sagittarius.

In order to fully estimate the tidal torque, one needs to know also the mass of the satellite. The mass estimates of Sagittarius still have, unfortunately, large uncertainties. Ibata & Razoumov (1998) suggest that a value of $M_{SG} \sim 5 \times 10^9 M_\odot$ would be sufficient to trigger the warp. Jiang & Binney (2000) argue that the current mass is in the range $1-3\times10^9 M_\odot$. However, the more recent numerical simulations of Helmi & White (2001) constrain the mass of the satellite in the range $0.4-1.7\times10^9 M_\odot$.

If we take in the formula (5.8) the nominal values: $z \simeq 5$ kpc, $R \sim 16$ kpc (the average radial distance of the warp) and $M_{sat} = M_{SG} \sim 1 \cdot 10^9 M_\odot$, we obtain a value for the radial torque exerted on a single star of mass 1 $M_\odot$: $\tau_r \sim 8.9 \times 10^{-14} M_\odot$ kpc km/s^2. This value corresponds to an average radial torque exerted on a single particle in our simulations, of $\sim 1.25 \times 10^{-7} M_\odot$ kpc km/s^2. From Figure 5.33, the corresponding tidally triggered warp would have a height of ~ 0.8 kpc, in rough agreement with the warp height in the Milky Way disk. Better agreement can be obtained, in principle, by assuming a higher mass for the satellite or by modifying either the z or R values in relation (5.8).

The Warp in Andromeda (M31)

The Andromeda galaxy displays a very pronounced warp beyond $R \sim 20$ kpc (Walterbros & Kennicutt 1987). The cause of the warp is still unknown, although the culprit could be one of the nearby satellite galaxies, eg. M32 or NGC 205 or a fully disrupted satellite. The recent detection of a giant stellar stream to the south of the disk suggests that M31 has experienced a recent merger with a dwarf galaxy (Ibata et al 2001). Current observational data and modeling of the orbit seem to suggest that neither M32 nor NGC 205 are consistent with being the progenitor of this stream (eg. Ibata et al. 2004). This progenitor may be a fully disrupted (or a

Figure 5.39: The Andromeda Galaxy (M31)

still undetected) satellite which can also influence the disk. In addition to the warp, the stellar disk of M31 displays other interesting peculiarities. The "Northern Spur" is such an example. This stellar feature has been recently discovered in the vicinity of the gaseous warp, on the north side of the disk (hence its name). Currently, there is some controversy about the origin of this feature, with opinions advocating that is either a part of the warp, or that is a continuation of the stellar stream seen to protrude at the south-east side of M31's disk (Ferguson et al. 2002; Merrett et al. 2003). If it is indeed part of the warp, its large extent and its asymmetric shape (no counterpart has been detected to the other side of the disk) need to be explained in the framework of an appropriate model.

As discussed in the previous case, in order to calculate the radial torque, one needs to know the orbital parameters of the satellite. In the case of M31, however, the observational constraints are not as good as in the case of the Milky Way. For example, the distances of its satellite galaxies relative to the center of M31 have uncertainties of several tens of kiloparsecs. Also, the geometry of the system (eg.

the orbital inclinations, the pericenters and apocenters of the orbits) is not known. Therefore, in this case, our model has limited application. We can, however, make an order of magnitude estimate regarding the strength of the tidal interaction. Thus, we can ask what combination of physical parameters in the definition of the radial torque can explain the magnitude of the observed warp.

For some parameters we have satisfactory observational constraints. The mass of M32 is about $M_{M32} \sim 3 \times 10^9 M_\odot$ and the mass of NGC 205 is estimated to be in the range: $M_{NGC205} = 3.6 - 15 \times 10^9 M_\odot$. Their orbital parameters, however, are not known. New acquired data on the Southern stellar stream has put better constraints on the orbit for its progenitor satellite (Lewis et al. 2004, Ibata et al. 2004, Font et al. 2004). These authors have found a lower mass bound for the satellite of several $10^8 M_{sol}$ and a pericenter of about 5 kpc. The mass of the M31's disk is not very well known, but numerical models suggest $M_d^{M31} = 7 \times 10^{10} M_\odot$ (Widrow, Perrett & Suyu, 2003) – a value similar to that of the Milky Way disk ($M_d^{MWY} = 5.6 \times 10^{10} M_\odot$). The radial distance R can be approximated as $\sim$ 20 kpc, i.e. the distance at which the warp is currently detected.

We further assume the nominal values in formula (5.8): $M_{sat} = 10^9 M_\odot$, $z \sim 5$ kpc, $R \sim 20$ kpc, and obtain a value for the radial torque exerted on a single star of mass 1 $M_\odot$ of: $\tau_r \sim 11.12 \times 10^{-14} M_\odot$ kpc km/s^2. This value corresponds to an average radial torque exerted on a single particle in our simulations, of $\sim 1.56 \times 10^{-7} M_\odot$ kpc km/s^2 and to a corresponding warp height of $\sim$ 1 kpc (see Figure 5.33). This value, admittedly lower than the observed height of the warp in M31 ($\sim 2 - 3$ kpc), has high uncertainties, given the large possible range in mass or orbital parameters. Nevertheless, it suggests that, in order for a satellite to create a warp similar to M31's, its mass should be at least a few $10^9 M_\odot$ and the triggering must have occurred at a very close passage near the disk (with z of a few kpc).

NGC 5907 Galaxy and its Companion: PGC 54419

NGC 5907 is a nearby, edge-on spiral galaxy with a large ($\sim$ 8 kpc) HI warp (Sancisi

1976). In addition, there is evidence that the stellar disk is also warped in the same direction as the HI warp (Morrison et al. 1994). This galaxy has been long cited as the prototype of a warped galaxy in isolation (i.e. without a satellite galaxy). However, recent observations have revealed that this galaxy has indeed a companion: the satellite PGC 54419 (Shang et al. 1998). This small satellite galaxy is severely disrupted along a large tidal stream, suggesting a significant tidal interaction with the disk and, possibly, the cause of the warp.

Some information about the orbital properties of this satellite has been inferred from the data gathered in its tidal stream. Thus, the orbit of PGC 54419 is believed to be nearly polar, and about 45 kpc in diameter. Reshetnikov & Sotnikova (2000) have performed numerical simulations for this system and have constrained the pericenter passage to a value of about 10 kpc. The mass of the satellite is fairly small: $M_{sat} \sim 10^8 M_\odot$ (Shang et al. 1998).

Assume the nominal values: $M_{sat} = 10^8 M_\odot$, $z \sim 10$ kpc, $R \sim 20$ kpc, in formula (5.8), we obtain a value for the radial torque exerted on a single star of mass 1 $M_\odot$: $\tau_r \sim 2.78 \times 10^{-15}\, M_\odot$ kpc km/s^2 (or a corresponding value for the average radial torque on a disk particle in our simulations of $\tau_r \sim 3.9 \times 10^{-9}\, M_\odot$ kpc km/s^2). Figure 5.33 suggests that this radial torque is insufficient to create a significant stellar warp. We note that, alternatively, a triaxial dark matter halo may also create a warp (Sackett 1994) – although, as we have mentioned before (§1.4.2., p22), there is yet no clear evidence for a flattened dark matter halo in the case of NGC 5907.

In conclusion, our model is roughly consistent with different properties of stellar warps detected in observational studies (for example, their average heights and inclination angles). It also gives an order of magnitude agreement in the case of a few particular warps, such as Milky Way's or Andromeda's. However, in order to fully investigate the origin of these warps, one needs to resort to numerical simulations specifically devised for each individual system. In certain cases, these simulations may not be yet possible, given that we still lack good constraints for their satellite

and disk parameters (eg. estimates of their masses, distances or orbital parameters).

Summary:

In this Chapter, the effects of tidal interactions between the disk and neighboring dark matter satellites have been analyzed in detail. The results have shown that the effects of substructure consist mainly of tilting and warping of the stellar disk.

- The tilting of the disk has been shown to be a consequence of the re-adjustment of the angular momenta within the galaxy.

- Our results suggest that tidal interactions between the disk and the neighboring satellites can induce and sustain stellar warps with similar properties as observed in the Universe. The most massive satellites are likely to generate warps of amplitudes up to 2-3 kpc, at a Galactocentric distance comparable with the solar radius. These values are typical for the warp heights measured in interacting galaxies. Also, the inclination angles of the tidally triggered warps are in good agreement with the observed values.

- Moreover, we have developed a general model which describes the formation of stellar warps through tidal interactions with satellites. This model can provide a prediction for the relative location of the warps and their decay time. The model scales with the mass and the orbital parameters of the satellites and can be, in principle, applied to specific cases of warped galaxies.

- We have shown that, despite their short individual lifetime, the warps can be regenerated and strengthened by repetitive tidal interactions, if these occur frequently enough. We predict that this phenomenon will occur when the orbital period of the satellite becomes comparable with the orbital period in the disk measured at the location of the warp.

- We caution that, at this point, we cannot rule out that other mechanisms (such as the interaction of the disk with a flattened dark matter halo) can provide viable, alternative ways of forming warps.

Chapter 6

Discussion and Conclusions

Using high resolution N-body simulations, we have investigated the dynamical effects that substructure and satellite accretions in Cold Dark Matter (CDM) halos have on galactic disks, with particular emphasis on the effects on the long-term stability of stellar disks, as well as the disk heating, tilting and warping. The simulations included here represent the largest and most realistic cosmological simulations of disk heating/ warping to date. The detailed reconstruction of the CDM satellite population makes them distinct from previous numerical simulations, which have focused on the interaction with a single satellite at a time.

The cumulative number of dark matter substructure

We have presented a description of the process of identification of the dark matter substructure on the scale of galaxy halos. Several tests have been performed in order to study how the number of sub-halos that are identified are influenced by: the numerical resolution of the cosmological simulations; the linking length or number of neighbors in the group-finding algorithm; different realizations of satellites; or the type of the cosmological model. This study re-affirms the discrepancy between the number dark matter satellites and the corresponding dwarf satellite galaxies in the Milky Way. Several arguments are also discussed which may help alleviate the

problem.

We also investigate the orbital parameters of the dark matter satellites and argue that the realizations of satellites presented in this thesis are representative of CDM models. This result ensures that the effects these satellites have on the disk may be typical of the CDM paradigm as a whole. Finally, we show that the population of dark matter satellites is evenly distributed between prograde and retrograde orbits, in agreement with observational studies. Based on this result, we suggest that the analytic calculations of heating rates in the disk may need to be revised to allow for this effect.

The effect of particle resolution in the heating of the disk

We have also performed several simulations in order to investigate the importance of the "numerical noise" in the secular evolution of a typical galaxy disk. Quantifying the effect of the artificial heating caused by two-body interactions is important for separating it from the "real" heating caused by physical processes (CDM substructure included). The results of these tests show that the heating of the disk increases with the number of disk and halo particles, roughly as $\sim N_d \cdot N_h$. The results imply that an accurate model of Milky Way galaxy requires a number of at least $N_d N_h \sim 10^{10}$ particles. Simulations with this number of particles are affordable with the present computational resources.

Disk Heating by Cold Dark Matter Substructure

Thin galaxy disks are fragile structures, prone to substantial perturbations during their gravitational interaction with satellite galaxies. Strong tidal interactions leave their trace in the disk structure through increased dispersion in the orbital motions of stars ("heating") and in extreme cases, through the disruption of the disk. We investigated recent arguments that the large number of satellites expected to orbit a typical disk galaxy in the Cold Dark Matter cosmology could alter a typical disk

beyond the observational constraints.

The primary conclusion of this study is that the disk is able to survive in a CDM halo with substructure for at least several Gyrs, thus severely weakening arguments calling for a revision of the present cosmological framework. The only visible adjustments made by the disk to the incoming satellites were through the (angular momentum conserving) tilting, or through changes in the kinematical properties of the stars: minor vertical disk heating, flaring and warping. Through these simulations, it has also become apparent that changes in the disk structure are driven mainly by a few massive satellites rather than by the cumulative effect of many minor mergers.

We also quantify the disk heating rate caused by the interactions with the satellites (whenever present). For the solar neighborhood case, we compare the results from the numerical simulations with the corresponding values derived from observations in the Galaxy and show that dark matter substructure cannot be responsible for the observed heating rate. However, the substructure can have a significant effect in the outer parts of the disk. In this case, we estimate the change in $\Delta\sigma_{tot}$, defined as the gradient in the total velocity dispersion of disk stars induced by dark matter substructure, above the numerical noise. Our results suggest that beyond $\sim 2R_{\odot}$ the heating rate due to CDM substructure can become comparable with the rate produced by other heating mechanisms, such as spiral arms or encounters with molecular clouds or black holes. At the present, the observations are mainly limited to the solar neighborhood. Thus, it is not yet possible to test our predictions regarding variations of the heating rate with Galactocentic distance. However, the situation is about to change in the next few years. Recently, it has become possible to observe the outer regions of stellar disks (such as M31, for example) in unprecedented detail. Ongoing and future surveys of stellar kinematics in the Milky Way and Andromeda will provide a breakthrough in the amount and quality of information we have on these galaxies. Given these current difficulties, we have chosen an alternative route, that compares the radial structure of the galactic disks in our simulations with that in external galaxies which undergo mergers. In particular, we have chosen for comparison the large sample of galaxy disks

collected by Schwarzkopf & Dettmar (2000; 2001). The results of our simulations are in good agreement with the observational results. Thus, for a similar mass range of the mergers ($M_{sat}/M_{disk} \sim 0.1$), we obtain an increase in the disk scale height of roughly the same magnitude as the one measured in the sample of disk galaxies of Schwarzkopf & Dettmar (2000; 2001) –, i.e. on average, disk scale heights ~ 1.5 times larger than the isolated galaxies (recall Figure 4.5). In addition, the CDM substructure can also explain the gradient in which the disk scale height increases with radial distance (see also Figure 4.5 and §4.3.). The overall results support the idea that the thicknesses of the stellar disks are consistent with the distribution of substructure predicted by CDM models.

These conclusions are subject to a number of caveats:

1) The most obvious one is that this study explores only three numerical realizations of a disk galaxy within clumpy dark matter halos, and it is always hazardous to extrapolate from such a small number of cases. As explained before, the main limitation in the construction of a statistical set galaxy models using N-body techniques is caused by the presently insufficient computational power. On the other hand, semi-analytical codes are much less affected by this constraint, are better equipped for constructing large sets of dark matter distributions. The latest advancements in the development of these codes, which include an accurate treatment of the orbital evolution of satellites, of the dynamical friction and tidal heating (Taylor & Babul, 2001; Benson et al. 2004), provide a promising avenue for testing our result. Indeed, the recent semi-analytical study of Benson et al. (2004) has been able to build a large number of realizations, and has, in fact, confirmed the result presented in this thesis (also, Font et al. 2001) – namely, that the heating by dark matter substructure is consistent with the existence of a large number of thin disks, as seen in observations.

2) Another caveat is that we have explored a model motivated by the present-day structure of the Milky Way and by the $z = 0$ substructure of a CDM halo. Models that take into account the ongoing formation of the disk and a more realistic

treatment of the evolution of substructure are clearly desirable in order to refine the conclusions presented here. Finally, for simplicity we have adopted a non-singular isothermal model for the dark halo in the simulations reported here, rather than the 'cuspy' density profiles found in cosmological N-body simulations (NFW). Although we do not expect this assumption to undermine the conclusions reported here, one cannot rule out subtle effects that may arise from the interaction between disk and halo in a cosmological context.

The Disk-Satellite Angular Momentum Connection

Making use of the same high resolution simulations, we have followed the changes that occurred in the distribution of angular momentum, while hundreds of dark matter satellites were tidally stripped during their orbital evolution. The tidal evolution of substructure has proven to be capable, in certain conditions, of redistributing the angular momentum inside the galaxy, with the effect of tilting the disk. Also, as a result of the strong tidal interaction with the disk, the main satellite tends to shift its orbit into a more highly inclined one.

Tidally Triggered Warps

Previous numerical studies have shown that a single satellite, if chosen on an appropriate orbit and if massive enough, can induce warping in a stellar disk (HC97; VW99). However, until now, it has been suspected that tidal forcing by satellites may be too weak to explain the warps observed in the Universe. This opinion has been largely based on a notorious failure of the tidal interaction between the LMC and the Milky Way disk to explain the Galactic warp (see, for example, Garcia-Ruiz, Kuijken & Dubinski 2002). Our simulations suggest that, in the context of CDM substructure, the warping of a typical galactic disk becomes statistically possible.

Our study shows that subsequent pericentric passages of a massive satellite passing in the close vicinity of the disk are able to replenish the energy needed for driving

transient warps. This study also shows that less massive satellites have a negligible effect on the disk during direct encounters (however, their inclusion in simulations is necessary, since they substantially alter the orbits of more massive satellites and, indirectly, influence the evolution of the disk).

We have shown that repetitive tidal interactions between massive dark matter satellites and the disk can create and even enhance vertical disturbances in the disk. Our results suggest that substructure in the dark matter halo may offer a plausible mechanism for creating irregular warps and thus offering an alternative formation scenario for some of the warps observed in the Universe.

Furthermore, the warps induced by tidal interactions in our simulations have similar properties as the warps observed in the Universe. The most massive satellites are more likely to create the warps of magnitude observed in nearby galaxies (up to 2-3 kpc in height at about the solar radius). Also, the inclination angles of the warps obtained in our simulations are similar to the values obtained in the observations ($\sim 3 - 4^o$).

Finally, we have developed a general model describing the formation of stellar warps through tidal interactions with satellites which can provide a prediction for the relative location of the warps and their decay time. The model scales with the mass and orbital parameters of the satellites and can be applied to specific cases of observed warps. The model gives an order of magnitude agreement in the cases of a few particular warps, such as Milky Way or Andromeda.

Our conclusions are subject to a number of caveats:

1) The present model is limited only to asymmetric warps. As discussed in §5.3.1 (and illustrated in Figures 5.24 – 5.25), the tidally triggered warps are likely to disappear in less than the orbital period (measured at the corresponding radius in the disk). It is well known, however, that warps come in different shapes: besides one-sided warps, there are also warps with symmetric S (one side up, the other side down) and U – shapes (both sides oriented in the same direction). Since our model

cannot easily explain these shapes, alternative scenarios need to be investigated for these cases.

2) Our results are based on a limited set of simulations. A full check of the validity of this mechanism in the cosmological setting can be provided only from a statistical study, which should include a consistent set of simulations on the galactic scale. Improved simulations also need to take into account the continuous infall of material at earlier epochs, in parallel with the growth of the disk. However, this type of simulations are computationally prohibitive at the present, due to the high numerical resolution that is required.

3) Our results cannot rule out that other mechanisms (such as flattened halos) could provide viable ways of forming warps. However, our aim was to investigate the tidally triggering mechanism separately from other possible warp-inducing mechanisms. Simulations which include the flattening of the dark halo are needed to compare the efficiency of the two mechanisms.

Astronomical Units

Parsec	1 pc = $3.08567802(2) \times 10^{18}$ cm
Hubble constant	$H_0 = 100h$ km s^{-1} Mpc^{-1}, where $0.5 \leq h \leq 1$
Hubble time	$1/H_0 = 9.78h^{-1} \times 10^9$ yr
Solar Mass	$M_\odot = 1.989(2) \times 10^{33}$ g
Metallicity	$Z \equiv \log[Fe/H] - \log[Fe/H]_\odot$

Bibliography

[1] Abadi, M. G., Navarro, J. F., Steinmetz, Matthias; Eke, Vincent R. 2003, ApJ, 591, 499

[2] Abe, F. et al. 1999, AJ, 118, 261

[3] Ardi, E., Tsuchiya, T., Burkert, A 2003, ApJ, 596, 204

[4] Azzaro, M., Prada, F., Gutiérrez, C. M., astro-ph/0310487

[5] Babul, A. & Rees, M. J. 1992 MNRAS, 255, 346

[6] Bahcall, N. A., Ostriker, J. P., Perlmutter, S., Steinhardt, P. J. 1999, Science, 284, 1481

[7] Bahcall, J. N. & Soneira, R. M. 1980

[8] Bahcall, J. N., Schmidt, M., Soneira, R. M. 1982, ApJ, 258, L23

[9] Bailin, J. 2003, ApJ, 583, L79

[10] Barbanis, B. & Woltjer, L. 1967, ApJ, 150, 461

[11] Bellazzini, M., Ferraro, F. R. & Ibata, R. 2003, AJ, 125, 188

[12] Bennett, C. L. et al. 2003, ApJS, 148, 1

[13] Benson, A. J., Lacey, C. G., Frenk, C., Baugh, M., Cole, S. 2004, MNRAS, 351, 1215

[14] Binney, J. 1992, ARAA, 30, 51

[15] Binney, J. J. & Evans, N. W. 2001, MNRAS, 327, 27

[16] Binney, J. & Tremaine, S. 1987, "Galactic Astronomy", Princeton University Press

[17] Binney, J. J. et al. 1998, MNRAS 298, 387

[18] Binney, J., Dehnen, W., Bertelli, G. 2000, MNRAS, 318, 658

[19] Bosma, A. 1981, AJ, 86, 1791

[20] Briggs, F. 1990, ApJ, 352, 15

[21] Bullock, J. S., Kravtsov, A. V., Weinberg, D. H., 2000, ApJ, 539, 517

[22] Burke, B. F. 1957, AJ, 62, 90

[23] Burstein, D. 1979, ApJ, 234, 829

[24] Burton, W. B. 1988, in "Galactic and Extragalactic Radio Astronomy", Second Edition, eds G. L. Verschuur & K. I. Kellermann. Springer-Verlag

[25] Buser, R. 2000, Science, 287, 69

[26] Carlberg, R. G. 1987, ApJ, 322, 59

[27] Carlberg, R. G. & Sellwood, J. A. 1985, ApJ, 292, 79

[28] Cen, R. 2001, ApJ, 546, L77

[29] Chandrasekar, S. 1942, "Principles of Stellar Dynamics", Chicago: Univeristy of Chicago Press. Reissued by Dover 1960

[30] Chiba, M. & Beers, T. C. 2000, AJ, 119, 2843

[31] Choi, P. I., Guhathakurta, P. & Johnston. K. V. 2002, AJ, 124, 310

[32] Coté, S., Carignan, C., Freeman, K. C. 2000, AJ, 120, 3027

[33] Dalal, N. & Kochanek, C. S. 2002, ApJ, 572, 25

[34] Dalcanton, J. J. & Bernstein, R. A. 2002, AJ, 124, 1328

[35] Davé, R., Spergel, D. N., Steinhardt, P. J., Wandelt, B. D., 2001, ApJ, 547, 574

[36] de Blok, W. J. G. et al. 2001, ApJ, 552, L23

[37] Dehnen, W. & Binney, J. J. 1998, MNRAS, 298, 387

[38] Dekel, A. & Silk. J. 1986, ApJ, 303, 39

[39] de Simone, R., Wu, X. & Tremaine, S. 2004, MNRAS, 350, 627

[40] de Vaucouleurs, G. & Pence, W. D. 1978, AJ, 83, 1163

[41] Dubinski, J., Mihos, J. C., Hernquist, L. 1996, ApJ, 462, 576

[42] Dubinski, J. & Carlberg, R. G. 1991, ApJ, 378, 496

[43] Edvardsson B., Andersen J., Gustafsson B., Lambert D. L., Nissen P. E., Tomkin J. 1993, A&A 275, 101

[44] Efstathiou, G. P. 1992, MNRAS, 456, 43

[45] Eggen, O. J., Lynden-Bell, D. & Sandage, A. R. 1962, ApJ, 136, 748

[46] Einasto, J. & Lynden-Bell, D. 1982, MNRAS, 199, 67

[47] Eke, V. R., Navarro, J. F., & Steinmetz, M. 2001, ApJ, 554, 114

[48] Ferguson, A. M. N., Irwin, M. J., Ibata, R. A., Lewis, G. F. & Tanvir, N. R. 2002, ApJ, 124, 1452

[49] Flores, R. & Primack. J. R. 1994, ApJ, 427, L1

[50] Font, A. S., Navarro, J. F., Stadel, J., Quinn, T. R. 2001, ApJ, 563, L1

[51] Font, A. S., Johnston, K. V., Guhathakurta, P., Majewski, S. M. & Rich, R. M. R, 2005, AJ, submitted

[52] Freeman, K. C. 1991 in "Dynamics of Disk Galaxies", ed. B. Sundelius, Göteborg: Univ. Göteborg

[53] Freeman, K. C. & Bland-Hawthorn, J. 2002, ARAA, 40, 487

[54] Frenk, C. S., White, S. D. M., Efstathiou, G. & Davis, M. 1985, Nature, 317, 595

[55] Freudenreich, H. T. et al. 1994, A&A, 429, 69

[56] Friese, V., Fuchs, B. & Wielen, R. 1995, in "Stelar Populations", Proc. IAU Symp 164, eds. van der Kruit, P. C., Gilmore, G. F.; Kluwer, Dordrecht, p. 414

[57] Fuchs, B., Dettbarn, C., Jahreiss, H., Wielen, R. 2001, in "Dynamics of Star Clusters and the Milky Way", eds. Deiters et al., ASP Conference Series, Vol. 228, San Francisco, p. 235

[58] Garcia-Ruiz, I., Kuijken, K., Dubinski, J. 2002, MNRAS, 337, 459

[59] Gardiner, L. T. & Noguchi, M. 1996, MNRAS, 278, 191

[60] Ghigna. S. et al. 1998, MNRAS, 300, 146

[61] Gilmore, G. & Reid, N. 1983, MNRAS, 202, 1025

[62] Gilmore, G., Wyse, R. F. G. & Norris, J. E. 2002, ApJ, 574, L39

[63] Goodman, J. 2000, New Astron., 5, 103

[64] Governato, F. et al. 1997, NewA, 2, 91

[65] Hänninen, J. & Flynn, C. 2002, MNRAS, 337, 731

[66] Hayashi, E., Navarro, J. F., Taylor, J. E., Stadel, J., Quinn, T. R. 2003, ApJ, 584, 541

[67] Helmi, A. et al. 2003 ApJ, 592, L25

[68] Helmi, A., White, S. D. M., de Zeeuw, P. T. & Zhao, H. 1999, Nature, 402, 53

[69] Helmi, A. & White, S. D. M. 2001, MNRAS, 323, 529

[70] Hernquist, L. 1993, ApJS, 86, 389

[71] Holmberg, J. 2001, PhD Thesis, Lund University

[72] Hogan, C. J. & Dalcanton, J. J. 2000, PhRevD, 63, 3511

[73] Huang, S. & Carlberg, R. G. 1997, ApJ, 480, 503, (HC97)

[74] Hunter, C. & Toomre, A. 1969, ApJ, 155, 747

[75] Ibata, R. A., Gilmore, G., Irwin, M. J. 1994, Nature, 370, 194

[76] Ibata, R. A. et al. 2001, Nature, 412, 49

[77] Ibata, R. A. et al. 2003, MNRAS, 340, L21

[78] Ibata, R. A. et al. 2004, MNRAS, 351, 117

[79] Ibata, R. A. & Razoumov, A. O. 1998, A&A, 336, 130

[80] Jiang, I. -G., Binney, J. 1999, MNRAS, 303, L7

[81] Jiang, I. -G., Binney, J. 2000, MNRAS, 314, 468

[82] Johnston, K. V., Hernquist, L. & Bolte, M. 1996, ApJ, 465, 278

[83] Johnston, K. V., Sackett, P. D., Bullock, J. S. 2001, ApJ, 557, 137

[84] Kamionkowski, M. & Liddle, A. R. 2000, PhRevL, 84, 4525

[85] Katz, N. & White, S. D. M. 1993, ApJ, 412, 455

[86] Kauffmann, G., White, S. D. M., Guiderdoni, B. 1993, MNRAS, 264, 201

[87] Keeton, C. R. 2001, ApJ, 561, 46

[88] Kerr, F. J. 1957, AJ, 62, 93

[89] Kleyna et al. 2003, ApJ, 588, 21L

[90] Klypin, A., Gottlöber, S., Kravtsov, A. V., Khokhlov, A. M. 1999, ApJ, 516, 530

[91] Klypin, A., Kravtsov, A. V., Valenzuela, O., Prada, F. 1999, ApJ, 522, 82 (K99)

[92] Lacey, C. G. 1984, MNRAS, 208, 687

[93] Lacey, C.G., Ostriker, J.P. 1985, ApJ, 299, 633

[94] Lee, Y. -W. et al., 1999, Nature, 402, 55

[95] Lewis, G. F. et al. 2004, in 5^{th} Galacto Chemodynamics conference, PASA, *in press* (astro-ph/0401092)

[96] Lilly et al. 1998, ApJ, 500, 75

[97] Lin, D. N. C. 1996, Proceedings of the 36th Herstmonceaux Conference, eds. O. Lahav, E. Terlevich, R. J. Terlevich; Cambridge University Press

[98] Lin, D. N. C., Jones, B. F. & Kremola, A. R. 1995, ApJ, 439, 652

[99] Lynden-Bell, D., Cannon, R. D., Godwin, P. J. 1983, MNRAS, 204, 87

[100] Mac Low, M. -M. & Ferrara, A. 1999, ApJ, 513, 142

[101] Majewski, S. R., Munn, J. A., Hawley, S. L. 1994, ApJ, 427, L37

[102] Majewski, S. R. et al. 1999, AJ, 118, 1709

[103] Mateo, M. 1998, ARAA, 36, 435

[104] Merrett, H. R. et al. 2003, MNRAS, 346, L62

[105] Mihalas, D. & Binney, J. 1981 "Galactic Astronomy", Second Edition, ed. W. H. Freeman & Co., San Francisco

[106] Miyamoto, M.; Yoshizawa, M.; Suzuki, S. 1988, A&A, 194, 107

[107] Moore, B. 1994, Nature, 370, 629

[108] Moore, B. et al., 1999, ApJ, 524, L19 (M99)

[109] Moore, B., Lake, G. Quinn, T. Stadel, J. 1999, MNRAS, 304, 465

[110] Morrison, H. L., Boroson, T. A.& Harding, P. 1994, AJ, 108, 1191

[111] Morrison, H. L., Miller, E. D., Harding, P., Stinebring, D. R. & Boroson, T. A. 1997, AJ, 113, 2061

[112] Morrison, H. L., Harding, P., Hurley-Keller, D. & Jacoby, G. 2003, ApJ, 596, L183

[113] Murai, T. & Fujimoto, M. 1980, PASJ, 32, 581

[114] Navarro, J. F., Frenk, C., White, S. D. M. 1996, ApJ, 462, 563

[115] Navarro, J. F., Frenk, C. S., White, S. D. M. 1997, ApJ, 490, 493

[116] Navarro, J. F. & Steinmetz, M. 1997, ApJ, 478, 13

[117] Navarro, J. F. & Steinmetz, M. 2000, ApJ, 538, 477

[118] Navarro, J. F. & White, S. D. M. 1994, MNRAS

[119] Newberg. H. J. et al. 2002, ApJ, 569, 245

[120] Nordström, B. et al. 2004, A&A, 418, 989

[121] Norris, J. E. & Ryan, S. G. 1989, ApJ, 336, L17

[122] Ostriker, J. P. & Steinhardt, P. J. 2003, Science, 300, 1909

[123] Peebles, J. E. 2000, ApJ, 534, 127

[124] Perlmutter, S. et al. 1999, ApJ, 517, 565

[125] Power, C. et al. 2003, MNRAS, 338, 14

[126] Press, W. H. & Schechter, P. 1974, ApJ, 187, 425

[127] Quinn, P. J. & Goodman, J. 1986, ApJ, 309, 472

[128] Quinn, P. J., Hernquist, L. & Fullagar, D. P. 1993, ApJ, 403, 74 (QHF93)

[129] Rees, M. J. 1986, MNRAS, 218, 25

[130] Reshetnikov, V., Battaner, E., Combes, F. & Jiménez-Vicente, J. 2002, A&A, 382, 513

[131] Reshetnikov, V. P. & Combes, F. 1998, A&A, 337, 9

[132] Reshetnikov, V. P. & Sotnikova, N. Ya. 2000, ApJ, 26, L277

[133] Riess, A. G. et al. 2001, ApJ, 560, 49

[134] Roman, N. G. 1954, AJ, 59, 307

[135] Sackett, P. D., Morrison, H. L., Harding, P., Boroson, T. A., 1994, Nature, 370, 441

[136] Sackett, P. D. 1999 in "Galaxy Dynamics", ASP Conf. Series vol. 182 (San Francisco: ASP), ed. D. R. Merritt, M. Valluri & J. A. Sellwood

[137] Sánchez-Saavedra, M. L., Battaner, E., Florido, E. 1990, MNRAS, 246, 458

[138] Sancisi, R. 1976, A&A, 53, 159

[139] Sand, D. J., Treu, T., Smith, G. P., Ellis, R. S. 2003, ApJ, 604, 88

[140] Schwarzkopf, U., Dettmar, R. -J. 2000 A&A, 361, 451

[141] Schwarzkopf, U., Dettmar, R. -J. 2001 A&A, 373, 402

[142] Searle, L. & Zinn, R. 1978, ApJ, 225, 357

[143] Sellwood, J. A. & Carlberg, R. G. 1984, A&A, 89, 296

[144] Shang, Z. et al. 1998, ApJ, 504, L23

[145] Sparke, L. 1986, MNRAS, 219, 657

[146] Spitzer, L. & Schwarzschild, M. 1951, ApJ, 114, 385

[147] Spitzer, L. & Schwarzschild, M. 1953, ApJ, 118, 106

[148] Springel, V. & White, S. D. M 1999, MNRAS, 307, 162

[149] Springel, V., White, S. D. M., Tormen, G., Kauffmann, G. 2001, MNRAS, 328, 726

[150] Stadel, J. G., PhD Thesis, 2001, University of Washington

[151] Stoehr, F., White, S. D. M., Tormen, G., Springel, V. 2002, MNRAS, 335, 84L

[152] Taylor, J. E. & Babul, A. 2001, ApJ, 559, 716

[153] Thacker, R. J. & Couchman, H. M. P. 2000, ApJ, 545, 728

[154] Toomre, A. 1963, ApJ, 138, 385

[155] Tormen, G., Diaferio, A. & Syer, D. 1998, MNRAS, 299, 728

[156] Tóth, G. & Ostriker, J. P. 1992, ApJ, 389, 5 (TO92)

[157] Tsikoudi, V. 1980, ApJS, 43, 365

[158] Tsuchiya, T., Dinescu, D. I., Korchagin, V. I. 2003, ApJ, 589, L29

[159] Tyson, J. A., Kochanski, G. P. & Dell'Antonio, I. P. 1998, ApJ, 498, 107

[160] van den Bergh, S. 2000, PASP, 112, 529

[161] van den Bosch, F. C. & Swaters, R. A. 2001, MNRAS, 325, 1017

[162] van den Bosch, F. C. et al. 2000, AJ, 119, 1579

[163] van der Kruit, P. C. & Searle, L. 1981a, A&A, 95, 105

[164] van der Kruit, P. C. & Searle, L. 1981b, A&A, 95, 116

[165] van der Kruit, P. C. & Searle, L. 1982, A&A, 110, 61

[166] Velázquez, H. & White, S. D. M. 1995, MNRAS, 275, L25

[167] Velázquez, H. & White, S. D. M. 1999, MNRAS, 304, 254 (VW99)

[168] Villumsen, J. 1985, ApJ, 290, 75

[169] Walker, I. R., Mihos, J. C., Hernquist, L. 1996, ApJ, 460, 121

[170] Wang, L., Caldwell, R. R., Ostriker, J. P., Steinhardt, P. J. 2000, ApJ, 530, 17

[171] Weinberg, M. 1989, MNRAS, 239, 549

[172] White, S. D. M. 1976, MNRAS, 177, 717

[173] White, S. D. M. & Rees, M. J. 1978, MNRAS, 183, 341

[174] White, S. D. M. 2000, "ITP Conference on Galaxy Formation and Evolution", `http://online.itp.ucsb.edu/online/galaxy_c00/white/`

[175] Widrow, L. M., Perrett, K. M. & Suyu, S. H. 2003 ApJ, 588, 311

[176] Wielen, R. 1974, in "Highlights of Astronomy", vol. 3, XVth General Assembly of the IAU, ed. G. Contopoulos, Dordrecht: Reidel

[177] Wielen, R. 1977, A&A, 60, 263

[178] Zaritsky, D., Olszewski, E. W., Schommer, R. A., Peterson, R. C., Aaronson, M. 1989, ApJ, 345, 759

[179] Zaritsky, D., Smith R., Frenk, C., White, S. D. M. 1997, ApJ, 478, 39 (a)

[180] Zaritsky, D., Smith R., Frenk, C., White, S. D. M. 1997, ApJ, 478, L53 (b)

[181] Zinn, R. 1993 in "The globular clusters-galaxy connection", ASP Conf. Series, Volume 48, ed. Graeme H. Smith & Jean P. Brodie, San Franciso: PASP

[182] Zwicky, F. 1937, ApJ, 86, 217