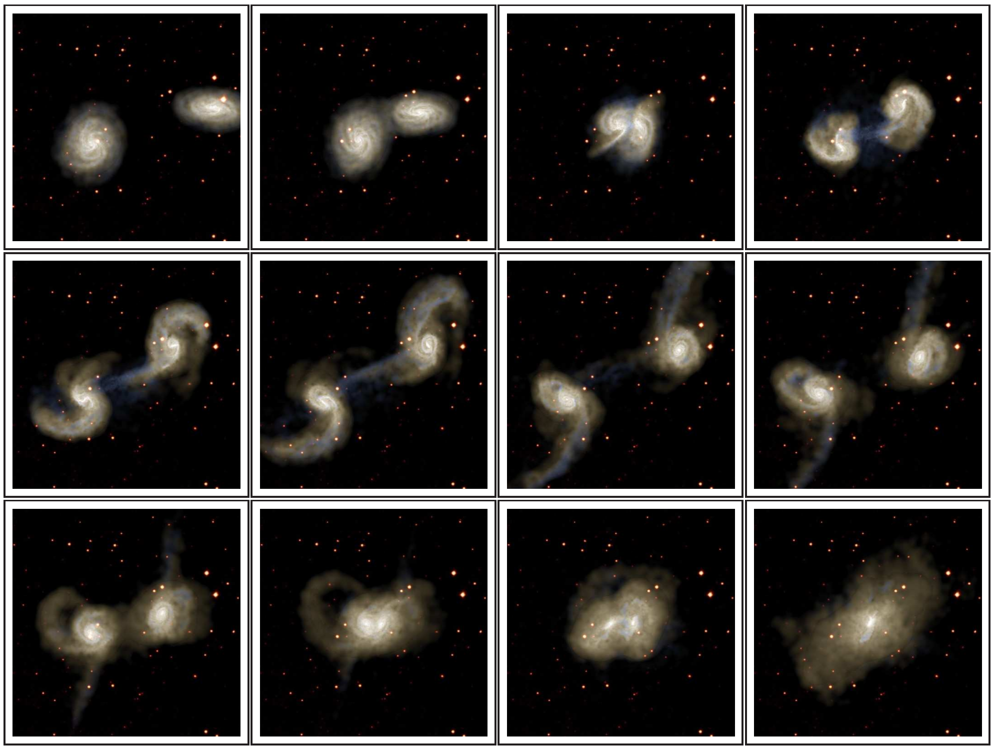
On the Formation and Evolution of Galaxies

Dissertation der Fakultät für Physik
der
Ludwig-Maximilians-Universität München



vorgelegt von Volker Springel
aus Backnang

München, den 12. Juli 1999

On the Formation and Evolution of Galaxies

Dissertation der Fakultät für Physik
der
Ludwig-Maximilians-Universität München

vorgelegt von Volker Springel
aus Backnang

München, den 12. Juli 1999

1. Gutachter:	Prof. Dr. Simon D. M. White
2. Gutachter:	Prof. Dr. Ralf Bender
Tag der mündlichen Prüfung:	9. Februar 2000

Für meine Eltern.

Contents

Zusammenfassung

In den derzeit favorisierten kosmologischen Modellen für die Entstehung von Galaxien wird angenommen, daß der Großteil der Masse im Universum aus dunkler Materie besteht, bei der es sich etwa um ein lediglich schwach wechselwirkendes, noch unbekanntes Elementarteilchen handeln könnte. Experimentell konnte die dunkle Materie bisher auf der Erde nicht nachgewiesen werden, allerdings gibt es viele indirekte Hinweise auf ihre Existenz. Diese indirekten Beweise stützen sich auf die gravitative Wirkung der dunklen Materie, wie sie in vielen astrophysikalischen Systemen beobachtet wird, etwa in den Bahnbewegungen von Galaxien in Galaxienhaufen, oder im Gravitationslinseneffekt, bei dem Bilder von Hintergrundgalaxien durch massereiche Vordergrundobjekte verändert werden. Es wird angenommen, daß die dunkle Materie zur heutigen Zeit nur noch an der gravitativen Wechselwirkung teilnimmt und sich daher wie ein stoßfreies Fluid verhält.

Das Wachstum von Struktur im Universum erfolgt durch die gravitative Instabilität von primordialen Fluktuationen in der Energiedichte des frühen Universums. Die derzeit erfolgreichsten theoretischen Modelle nehmen an, daß die dunkle Materie "kalt" ist, das heißt ihre thermische Bewegung ist vernachlässigbar zum Zeitpunkt der Gleichheit der Energiedichte in Strahlung und Materie. In Kosmologien, die von kalter dunkler Materie (CDM, englisch *cold dark matter*) dominiert sind, erfolgt die Entstehung von Galaxien hierarchisch. Zunächst bilden sich kleine Galaxien, welche sich dann in einem hierarchischen Verschmelzungsprozeß zu immer größeren Einheiten zusammenfügen. In diesem Bild der Strukturbildung "von unten nach oben" enstehen große Galaxienhaufen, welche die gegenwärtig größten virialisierten Strukturen darstellen, erst relativ spät.

In dieser Arbeit beschäftige ich mich mit einer Reihe verwandter Aspekte der Theorie der hierarchischen Entstehung von Galaxien. In einem ersten Schwerpunkt meiner Arbeit konzentriere ich mich auf Verschmelzungsprozesse von Scheibengalaxien. Wenn zwei Spiralgalaxien vergleichbarer Masse zusammenstoßen, schleudern gravitative Gezeitenkräfte Sterne und interstellares Gas aus den dynamisch kalten Scheiben heraus. Dieses Material führt zur Bildung langer Gezeitenarme, wie sie in vielen wechselwirkenden Paaren von Galaxien beobachtet werden. In manchen Systemen erreichen die Gezeitenarme Längen von mehreren hundert Kiloparsec.

Vor kurzem wurde in numerischer Arbeit gezeigt, daß die Anfälligkeit von Scheibengalaxien für die Entstehung von Gezeitenarmen stark von der internen Struktur der Galaxien abhängt. Aus diesem Grund könnten Gezeitenarme interessante Hinweise auf die Menge dunkler Materie in Galaxien und auf ihre Verteilung relativ zu den leuchtenden Sternen liefern. Die aktuellen CDM-Modelle für die Entstehung von Galaxien sagen voraus, daß der leuchtende Teil der Galaxien in einen sehr ausgedehnten und massereichen Halo aus dunkler Materie eingebettet ist. Die Gesamtmasse des Halos überschreitet die Masse der Sterne dabei um ein Vielfaches. Je massereicher der Halo allerdings ist, desto größer werden die auftretenden Geschwindigkeiten in Galaxienkollisionen, was zu kürzeren Wechelswirkungszeiten und zu einer insgesamt geringeren Störung der Scheiben führt. Außerdem hat das durch Gezeitenkräfte gestörte Material dann ein tieferes Gravitationspotential zu überwinden. Sehr massereiche Halos könnten daher die Entstehung ausgeprägter Gezeitenarme verhindern.

Um diesem Punkt weiter nachzugehen, habe ich detaillierte numerische Modelle von Scheibengalaxien konstruiert. Die strukturellen Eigenschaften dieser Galaxien wurden aus analytischen Modellen für die erwartete Verteilung von Scheibengalaxien in CDM-Universen abgeleitet. Ich habe dann Kollisionen und Verschmelzungen solcher Scheibengalaxien in einem großen Satz von Computersimulationen studiert. Dabei zeigte sich, daß Modelle mit kleinen Scheiben kaum in der Lage sind, Gezeitenarme zu bilden, während Galaxien, die große Scheiben relativ zu ihrem Halo besitzen, viel leichter massereiche und ausgedehnte Gezeitenarme produzieren. Allerdings scheint es auch so zu sein, daß die aktuellen CDM-Modelle eine hinreichend große Zahl von Galaxien voraussagen, die Gezeitenarme bilden können. Aus diesem Grund kann die beobachtete Häufigkeit solcher Systeme zwanglos erklärt werden. Diese Schlußfolgerung ist praktisch unabhängig von den gewählten kosmologischen Parametern der CDM-Modelle. Aus diesem Grund erscheint es unwahrscheinlich, daß Gezeitenarme besonders nützlich zu deren Bestimmung sind, wenn sie auch weiterhin ein sehr interessantes Werkzeug zur Erkundung der inneren Struktur von Scheibengalaxien darstellen.

In einem zweiten Schwerpunkt meiner Forschungsarbeit habe ich mich mit Problemen der Sternentstehung und ihrer Rückwirkung auf das interstellare Medium in Scheibengalaxien beschäftigt. Beobachtungen zeigen, daß die Sternentstehungsrate in Scheibengalaxien zeitlich relativ konstant ist. Man nimmt an, daß dies als Ergebnis von Rückwirkungsprozessen durch Supernova-Explosionen, durch ultraviolette Strahlung junger Sterne, oder durch starke Sternenwinde zustande kommt. Diese Prozesse führen zu einer Selbstregulierung der Sternentstehungsrate. In Simulationen ohne solche Rückwirkungsmechanismen führen die unbehinderten Kühlungsprozesse des Gases zu einem ungebremsten Anstieg der Sternentstehungsrate. Allerdings sind die detaillierten Wirkungsmechanismen der Rückwirkungsprozesse bisher weitgehend unverstanden. Da sich das Problem der Sternentstehung auch durch einen besonders großen dynamischen Bereich auszeichnet, lassen sich die relevanten physikalischen Prozesse auf der Skala ganzer Galaxien nicht direkt simulieren. In dieser Arbeit entwickle ich daher ein phänomenologisches Modell, das die Einbeziehung von Sternentstehung und ihrer Rückwirkung in hydrodynamischen Simulationen von Galaxien erlaubt. Dieses Modell benutzt einen effektiven, turbulenten Druck, um die turbulenten Gasbewegungen auf kleinen, nicht aufgelösten Skalen zu beschreiben.

Das Modell zur Beschreibung von Rückwirkung wurde so konstruiert, daß es die beobachtete Abhängigkeit der Sternentstehungsrate von der Gasdichte in Spiralgalaxien reproduziert. Man kann nun auch versuchen, das Modell auf Verschmelzungen von

Spiralgalaxien mit hohem Gasgehalt anzuwenden. Nachdem die Galaxien ihre erste Begegnung durchlaufen haben, kommt es zu einem starken Strom von Gas in das Zentrum der Galaxien, da gravitative Drehmomente während der Kollision dem Gas Drehimpuls entziehen. Durch die rasch ansteigende Konzentration von Gas im Zentrum kommt es zu einem starken Ausbruch der Sternentstehung, ähnlich wie er in vielen superhellen Infrarotgalaxien beobachtet wird. Ich zeige, daß die Flächendichte der Sternentstehungsrate in den zentralen Regionen der Modellgalaxien konsistent mit derjenigen in beobachteten Ausbrüchen der Sternentstehung ist.

Nachdem die Galaxien ein erstes Mal zusammengestoßen sind, entfernen sie sich zunächst wieder. Allerdings führen die dynamischen Reibungsverluste zu einer raschen Dissipation der Energie der Bahnbewegung, so daß die Galaxien erneut zusammenfallen und schließlich verschmelzen. Nach der abschließenden Verschmelzung entsteht ein einziges Sternsystem, in welchem die Galaxien ihre ursprüngliche Identität verlieren. Diese Verschmelzungsprodukte weisen große strukturelle Ähnlichkeiten mit elliptischen Galaxien auf. Ich vergleiche nun die Struktur der entstehenden Galaxien in Simulationen mit rein stoßfreier Dynamik mit derjenigen aus äquivalenten gasdynamischen Simulationen. Es zeigt sich, daß die Form der Isophoten systematisch unterschiedlich ist. Galaxienverschmelzungen, bei denen starke dissipative Prozesse und Sternentstehung beteiligt sind, führen zu Isophoten, die scheibenartige Abweichungen von einer perfekten Ellipsenform aufweisen. Umgekehrt tendieren dissipationsfreie Verschmelzungen dazu, rechteckförmige Abweichungen zu produzieren. Dies bestätigt eine speziell in Beobachtungsstudien der Entstehung elliptischer Galaxien gehegte Vermutung, daß scheibenartige Isophoten mit Dissipation verbunden sind. Ein potentielles Problem der untersuchten dissipativen Simulationen ist allerdings, daß sie möglicherweise zu effektiv im Produzieren eines zentralen Ausbruchs der Sternentstehung sind. Die dabei auftretende Erhöhung der zentralen Phasenraumdichte führt zu einer zentralen Spitze in der Helligkeitsverteilung, die möglicherweise nicht in Übereinstimmung mit den beobachteten Helligkeitsverteilungen elliptischer Galaxien steht.

In einem dritten Schwerpunkt meiner Arbeit studiere ich die Entstehung von Galaxien innerhalb eines Modells, das deren kosmologische Einbindung korrekt berücksichtigt. Zu diesem Zweck habe ich die Entstehung eines großen Galaxienhaufens in einem flachen Universum geringer Dichte in hochaufgelösten Computermodellen simuliert. Diese Simulationen wurden auf einem parallelen Supercomputer ausgeführt und erreichen wahrscheinlich den größten dynamischen Bereich, der bisher für solche kosmologischen Simulationen verwendet wurde. Indem ich eine große Zahl von Simulationsausgaben speichere, können die Simulationen zur direkten Analyse der Verschmelzungsprozesse der dunklen Materie verwendet werden. Die hohe Auflösung der Simulationen erlaubt dabei auch erstmals, das Schicksal von Halos zu verfolgen, die in größere Halos hineinfallen. Dabei werden die kleineren Halos nicht sofort aufgelöst, sondern deren innerer Kern kann die starken gravitativen Gezeitenkräfte des Haufens überleben und sich als unabhängiges physikalisches Objekt im Potential des entstehenden Haufens bewegen. In dieser Arbeit entwickle ich neue Techniken, um solche Unterstruktur in Halos der dunklen Materie aufzuspüren und deren Entwicklung zu verfolgen.

In den Simulationen der Galaxienhaufen wird nur die Entwicklung der stoßfreien dunklen Materie berechnet, welche den gravitativ dominierenden Anteil an der Masse im Universum darstellt. Um auch die Entwicklung der leuchtenden Teile der Galaxien zu beschreiben, benutze ich sogenannte semi-analytische Modelle der Galaxienentstehung. In diesen Modellen werden Prozesse wie die Strahlungskühlung des Gases, die Sternentstehung und ihre Rückwirkung, oder die Anreicherung des Gases durch Metalle mit vereinfachten, jedoch physikalisch motivierten Modellen beschrieben. Ich erweitere diese Modelle, so daß die Informationen über die Unterstruktur in Halos aus dunkler Materie mit einbezogen werden kann. Wie ich zeigen kann, führt dies zu einer deutlichen Verbesserung einiger Ergebnisse dieser semi-analytischen Modelle. So wird etwa die Leuchtkraftfunktion durch eine Schechter-Funktion nun gut beschrieben, und das Problem zu heller Galaxien im Haufen, welches in einfacheren Verfahren auftritt, wird vermieden. Ich zeige, daß diese Verbesserungen in erster Linie durch genauere Schätzungen der tatsächlichen Verschmelzungsrate von Galaxien im Haufen zustandekommen, so wie sie die detaillierte Unterstrukturanalyse liefert.

Insgesamt liefern die semi-analytischen Modelle in ihrer jetzigen Ausgestaltung eine bemerkenswert erfolgreiche Beschreibung der Galaxienpopulation in großen Haufen. Mit einem relativ kleinen Satz freier Parameter wird die Leuchtkraftfunktion von Haufengalaxien und deren Masse-Leuchtkraft-Verhältnis gut beschrieben. Gleichzeitig findet man eine ausgeprägte Dichte-Morphologie-Beziehung, und die Verteilung des $B-K$ Farbindex zeigt eine Dichotomie zwischen Spiralen und Ellipsen, ähnlich wie es beobachtet wird. Außerdem wird die Tully-Fisher Relation von Feldgalaxien gut beschrieben und die vorausgesagte Geschichte der Sternentstehungsrate ist in guter Übereinstimmung mit aktuellen Beobachtungsdaten.

Ein erheblicher Teil der oben beschriebenen Arbeit macht intensiven Gebrauch von Computersimulationen, sowohl mit rein stoßfreier Dynamik als auch mit Hydrodynamik. Diese Simulationen wurden mit einem neuen Simulationsprogramm berechnet, welches ich für diese Arbeit entwickelt habe. Das Programm kombiniert ein hierarchisches Multipolverfahren zur Berechnung der Gravitation mit einer teilchenbasierten Beschreibung hydrodynamischer Fluide (sogenannte *smoothed particle hydrodynamics*). Das Programm verwendet eine Reihe algorithmischer und numerischer Verbesserungen im Vergleich zu früheren Verfahren. Beispielsweise werden alle Teilchen mit individuellen, beliebig variablen Zeitschritten integriert und ein neues, effizienteres Kriterium für die Verwendbarkeit der Multipol-Näherung wird eingeführt. Ich habe auch eine Programmversion zur Verwendung auf parallelen Supercomputern mit verteiltem Speicher entwickelt. Hierbei wurde ein expliziter Kommunikationsansatz verwendet, welcher erhebliche algorithmische Änderungen im Vergleich zu den seriellen Berechnungsverfahren erforderlich machte. Aufgrund seiner Lagrangeschen Eigenschaften schränkt das implementierte numerische Verfahren das zu untersuchende Problem prinzipiell weder in seiner Geometrie noch in seinem dynamischen Bereich ein. Das Simulationsprogramm hat dadurch einen sehr weiten Einsatzbereich. In einem vierten Schwerpunkt meiner Forschungsarbeit beschreibe ich die eingesetzten numerischen Techniken und ich stelle eine Reihe von Tests der Geschwindigkeit und Genauigkeit des entstandenen kosmologischen Werkzeugs vor.

In einem abschließenden, fünften Teil meiner Forschungsarbeit untersuche ich das großräumige Dichtefeld des Universums anhand des 1.2-Jy Rotverschiebungskatalogs. Dieser Galaxienkatalog wurde im infraroten Spektralbereich ausgewählt, und zwar von einer Liste infraroter Punktquellen, die vom *IRAS*-Satelliten beobachtet worden sind. Der 1.2-Jy Rotverschiebungskatalog läßt sich zur Konstruktion einer Karte der Galaxienverteilung im lokalen Universum heranziehen, welche fast die gesammte Himmelskugel abdeckt. Die modernen Theorien zur hierarchischen Galaxienentstehung in CDM-Universen gehen von winzigen primordialen Störungen in der Massendichte des frühen Universums aus. Die Aufklärung der statistischen Eigenschaften dieser Fluktuationen stellt eine der grundlegenden Herausforderungen der Kosmologie dar. "Inflationäre" kosmologische Modelle schlagen vor, daß die anfänglichen Störungen letztlich auf Quantenfluktuationen zurückgehen, die während der inflationären Phase um ein vielfaches verstärkt wurden. In solchen Modellen sind die Phasen verschiedener Fouriermoden des Fluktuationsspektrums statistisch unabhängig, so daß die Störungen ein Gaußsches Zufallsfeld bilden. Es gibt allerdings auch alternative Szenarien zur Bildung der primordialen Störungen, die als Saatkörner der kosmischen Strukturbildung aufgefaßt werden können. Topologische Defekte wie magnetische Monopole, oder kosmische Fäden (englisch *cosmic strings*) sind solche Möglichkeiten. Allerdings sagen solche Modelle mit topologischen Defekten Phasenkorrelationen zwischen den Störungen voraus.

In dieser Arbeit benutze ich das topologische Genus-Maß, um die Geometrie von Flächen gleicher Dichte im großräumigen Dichtefeld des Universums zu untersuchen. Diese statistische Methode kann benutzt werden, um die Hypothese eines Gaußschen Zufallsfelds für die anfänglichen Störungen direkt zu testen. Ich entwickle eine ausgedehnte statistische Methodik, um theoretische Computermodelle für die Verteilung der dunklen Materie mit beobachteten Galaxienkatalogen wie dem 1.2-Jy Katalog zu vergleichen. In diesem Verfahren benutze ich große Ensembles künstlicher Galaxienkataloge, welche die Auswahleffekte und die statistischen Unsicherheiten in beobachtenden Daten nachbilden. Ich entwickle dann eine mehrdimensionale Hauptkomponentenanalyse, um die beobachteten Genuskurven mit theoretischen Modellen zu vergleichen. Ich zeige, daß das 1.2-Jy Dichtefeld auf großen Skalen mit der Hypothese eines Gaußschen Zufallsfelds konsistent ist. In dem schwach nichtlinearen Bereich des gravitativen Wachstums von Struktur entwickeln sich Phasenkorrelationen, die mit der Genus-Statistik gemessen werden können. Es wird gezeigt, daß die Genus-Statistik des 1.2-Jy Katalogs inkonsistent mit dem 'Standard'-CDM (SCDM) Modell ist, und zwar mit einem Konfidenzbereich von über 99%. Dies ist in erster Linie eine Folge einer falschen Form des angenommenen Fluktuationsspektrums, welches auf großen Skalen zu schwache Störungen aufweist. Alternative Varianten der CDM-Modelle mit ausreichend starken Fluktuationen auf großen Skalen sind dagegen in guter Übereinstimmung mit dem 1.2-Jy Katalog. Ein wesentlicher Vorteil der Genus-Analyse im Vergleich mit anderen Verfahren zur Untersuchung der Großraumstruktur ist ihre relative Unabhängigkeit vom sogenannten *Bias* der Galaxienverteilung. Im allgemeinen wird angenommen, daß die Galaxienverteilung der Massenverteilung folgt, jedoch gibt es vermutlich Abweichungen von einer strengen Proportionalität, den sogenannten Bias der Galaxien. Die Ergebnisse der Genus-Statistik sollten sich nicht oder nur sehr unwesentlich ändern, falls ein solcher Bias existiert.

Teile dieser Arbeit (Kapitel 2 und 6) sind in den Artikeln Springel et al. (1998) und Springel & White (1999) bereits veröffentlicht worden oder befinden sich im Druck.

Summary

In the currently favoured cosmological models for the formation of galaxies it is assumed that most of the mass in the Universe consists of 'dark' matter, perhaps in the form of a yet to be identified, weakly interacting massive particle. While the dark matter has not been detected directly on Earth so far, there is a lot of indirect evidence for its existence. This indirect evidence stems from the dark matter's gravitational effects on its surroundings, seen for example in the orbital motions of galaxies within clusters, or in the gravitational lensing of background galaxies by massive foreground objects. At the current epoch, dark matter is assumed to interact only gravitationally, behaving as a collisionless fluid. The growth of structure is then thought to occur via gravitational instability of primordial fluctuations in the energy density of the early Universe. The most successful theoretical models assume that the dark matter is *cold*, i.e. it has only a small velocity dispersion at the epoch of matter-radiation equality. In these cold dark matter (CDM) cosmologies, galaxy formation proceeds hierarchically. Small galaxies are born first, and a hierarchy of merging processes then assemble ever more massive systems. The largest virialized objects at the present time, rich clusters of galaxies, are expected to form rather recently in this 'bottom-up' paradigm of galaxy formation.

In this thesis, I work on a number of related aspects of the theory of hierarchical galaxy formation. In one line of my research, I focus on merging processes of disk galaxies. When two spiral galaxies of comparable mass collide, gravitational tidal forces lead to the ejection of tails of stars and gas out of the dynamically cold disks. Such tidal tails are observed in many interacting pairs of galaxies, sometimes extending to distances as large as several hundred kiloparsec from the main bodies of the galaxies.

Recent numerical work suggested that disk galaxies of differing structure can differ strongly in their susceptibility to tail formation. Tidal tails may thus provide important clues about the amount of dark matter in galaxies, and about its distribution relative to that of the visible light. According to the current CDM models of galaxy formation, the luminous part of galaxies is embedded in a fairly extended halo of dark matter, with total mass much larger than that of the stellar component itself. Galaxies with more massive haloes experience higher encounter velocities, leading to a smaller overall strength of the perturbation of the disk. In addition, the perturbed material has to climb out of a deeper potential well, possibly preventing the formation of massive tails.

To elucidate this point further, I have constructed detailed numerical models of disk galaxies with structural properties derived from analytic models for the population of disk galaxies expected in CDM universes. I then studied collisions and mergers of these models in a large set of computer simulations. It was found that models with small disks have difficulty ejecting tails, while larger disks much more easily produce massive and extended stellar tails. However, it appears that in the context of current CDM theories a relatively large number of galaxies should be capable of tail formation. Hence the abundance of observed systems with tails can be explained relatively easily. This conclusion is practically independent of the cosmological parameters of the CDM models. It thus seems unlikely that tidal tails are useful to constrain cosmological parameters.

In a second line of my research, I have addressed problems related to gas cooling and star formation in disk galaxies. Observationally, the star formation rate in isolated disk galaxies appears relatively constant in time. It is thought that this results from feedback processes due to supernova explosions, UV radiation of young stars, or stellar winds, leading to a self-regulation of star formation. Simulations without modeling of such processes give rise to a run-away of the star formation rate due to unimpeded cooling. Unfortunately, the physical mechanisms constituting feedback are poorly understood in detail. There are also severe technical constraints arising from the extraordinary large dynamic range encountered in this problem, precluding any brute force attempt to simulate star formation on the scale of whole galaxies. In this study, I have therefore developed a phenomenological model for the incorporation of star formation and supernova feedback into hydrodynamical simulations of galaxies. The model employs the notion of an effective turbulent pressure arising from motions in the interstellar medium on unresolved scales.

By construction, the feedback model results in star formation rates in agreement with the observed global 'Schmidt-law' for disk galaxies. I also apply this modeling to mergers of gas-rich spiral galaxies. During the first encounter of the galaxies, gravitational torques extract angular momentum from the gaseous component, leading to a strong central influx of gas. As a result, a burst of star formation develops in the galaxies' centers, much like that seen in ultraluminous infrared galaxies. I show that the surface density of the star formation rate in the central regions is consistent with the one observed in starbursting systems.

In these collisions, the galaxies separate again after their first encounter. However, because of energy losses due to dynamical friction, they fall back together again and eventually merge. After the final coalescence of the galaxies, they form a single pile of stars and lose their individual identity. These merger remnants resemble elliptical galaxies. By comparing the structure of remnants between dissipationless simulations and equivalent gasdynamical collisions I show that the isophotal shapes are systematically different. Mergers involving strong dissipation and star formation lead to isophotes that have 'disky' distortions from perfect ellipticity, while pure collisionless simulations tend to result in 'boxy' ellipticals. This confirms a long standing hypothesis in observational studies of the formation of elliptical galaxies. A potential problem of the current modeling is that the dissipative simulations might be too effective in producing a central starburst. The resulting increase of central phase-space density leads to a central 'spike' in luminosity which appears inconsistent with observed luminosity profiles of elliptical galaxies.

In a third line of research, I study the formation of galaxies in their proper cosmological environment. To this end, I have simulated the formation of a rich cluster of galaxies in a flat, low-density universe using very high numerical resolution. These simulations have been carried out on a parallel supercomputer and achieve the largest dynamic range ever reported for cosmological N-body simulations. By storing a large number of simulation outputs, these high-resolution simulations can be used to study the formation history of dark matter haloes in great detail. In particular, they give a proper account of the fate of the dark haloes that fall into a larger object. Such haloes are not dissolved immediately. Rather, their cores may survive and orbit as independent physical entities within the potential of the larger system. In this thesis, I develop new techniques to identify and track such substructure within dark matter haloes.

The cluster simulations only compute the evolution of the collisionless dark matter, which is the gravitationally dominant component of the universe. To study the formation of the luminous parts of galaxies, I employ so-called semi-analytic models of galaxy formation. In these models, processes like gas cooling, star formation, feedback, or metal enrichment, are treated with a set of simple, yet physically motivated assumptions. I extend previous models of semi-analytic galaxy formation to allow the inclusion of the substructure information extracted from the high-resolution simulations, and I show that this improves the results obtained considerably. For example, the luminosity function of cluster galaxies becomes well fit by a Schechter function. It develops a well defined 'knee', and a problem of excessively bright first ranked cluster galaxies that occurs in simpler models, is avoided. I show that these improvements primarily result from a much better estimate of the actual merger rate of galaxies within dark matter haloes, which is provided by high resolution simulations.

In their current form, the semi-analytic models provide a remarkably successful description of the population of cluster galaxies. With a relatively small set of free parameters, the luminosity function of cluster galaxies and the cluster mass-to-light ratio can be fit. A pronounced morphology-density relation is found, and the $B-K$ color distribution shows a dichotomy between spirals and ellipticals, much like that observed. Simultaneously, the Tully-Fisher relation of field galaxies is well fit, and the predicted global star formation history is in reasonable agreement with current observational results.

Much of the work described above has been based on the extensive use of computer simulations, both with collisionless dynamics and with hydrodynamics. These simulations have been carried out with a new simulation code that I have written for this thesis. The code combines a gravity-solver based on a hierarchical tree-algorithm with smoothed particle hydrodynamics. It implements a number of algorithmic and numerical improvements compared to previous methods. For example, all particles are integrated with individual, fully adaptive timesteps, and a new, more efficient cell-opening criterion for the tree-algorithm is introduced. I have also parallelized the code for massively parallel supercomputers with distributed memory. Here, an explicit communication approach was adopted, requiring substantial algorithmic changes compared to the serial version of the simulation code. Due to its Lagrangian nature, the code is not restricted by geometry or by the dynamic range of the problem under study. It has thus a wide range of applications. In a fourth line of my research, I describe the numerical algorithms employed in the code, and I present a number of tests showing the performance and accuracy of this cosmological tool.

Finally, in a fifth line of research, I study the large-scale density field of the Universe using the 1.2-Jy redshift survey. This catalogue of galaxies has been selected in the infrared, from a parent catalogue of galaxies detected by the *IRAS* satellite. The 1.2-Jy survey provides a nearly full sky map of the observed galaxy distribution in the Local Universe. Hierarchical galaxy formation according to CDM models starts from tiny primordial fluctuations in the matter density field. It is one of the great challenges of cosmology to determine the statistical nature of the initial perturbation field. Inflationary cosmology suggests that quantum fluctuations, amplified during the inflationary epoch, gave rise to the initial perturbations. In this case, the phases of different Fourier modes of the fluctuation spectrum should be uncorrelated, i.e. the perturbations should form a Gaussian random field. However, there are also alternative scenarios for the generation of a seed field for structure formation. For example, topological defects like cosmic strings or magnetic monopoles could provide perturbations as well, but such defect models generically predict phase correlations.

In this thesis, I use the topological genus-measure to characterize the geometry of isodensity surfaces of the large-scale density field. This statistical method can be used to test the Gaussian random phase hypothesis for the initial fluctuation field. I develop an extensive statistical methodology to compare theoretical N-body models for the clustering of dark matter with observed galaxy catalogues like the 1.2-Jy survey. This method works with large ensembles of mock galaxy catalogues that mimic the selection biases and the sampling noise inherent in observational data, and it employs a multivariate principal components analysis to compare the observed genus curves with theoretical models. I show that on large scales the 1.2-Jy density field is consistent with being a Gaussian random field. In the mildly non-linear regime of gravitational clustering, phase correlations develop which can be detected with the genus statistic. The genus statistic of the 1.2-Jy survey is shown to be inconsistent with the 'standard' cold dark matter model (SCDM) at a 99% confidence level. This is because the shape of the power spectrum of SCDM does not exhibit enough large-scale power. However, alternative variants of CDM models with more large-scale power are in good agreement with the 1.2-Jy survey. Unlike many other techniques to study large-scale clustering, the genus statistic is largely independent of galaxy bias. These results are thus unlikely to be changed if galaxies are biased tracers of the mass.

Parts of this thesis (Chapters 2 and 6) have already been published in the articles of Springel et al. (1998) and Springel & White (1999).

–I was sitting in a chair in the patent office at Bern when all of a sudden a thought occurred to me: "If a person falls freely he will not feel his own weight." I was startled. This simple thought made a deep impression on me. It impelled me toward a theory of gravitation.

Albert Einstein

1 Thesis objectives

1.1 Motivation

Now that the work on my PhD thesis is almost completed, I find it difficult to recount what originally motivated me to select the subject of galaxy formation for my dissertation. After all, the breadth of current research in physics offers such a multitude of possibilities that anyone's research career will be more like a random walk, constantly disturbed by interesting questions one happens to run into by chance, rather than being a straight path to some envisaged goal in the distance. So I would have to lie if I said that I had been determined already years ago to select the topic I have chosen. However, I still know well what originally drove me to study physics. It was the interest and curiosity in the way nature works, it was the admiration for the great discoveries of physicists, and it was the attraction by the intellectual challenges provided by natural sciences. Against the fundamental scope of physics, other academic disciplines always appeared to me less interesting.

Within theoretical physics, cosmology is a particularly fascinating subject. It offers large numbers of challenging problems, it stretches human imagination far beyond the realm of direct experience, and it is highly appealing because of its close ties to very fundamental questions about the world we live in. I find it satisfying that these questions are not only interesting for physicists, but that this interest is shared by the general public.

Galaxies are among the most beautiful objects on the sky, and I was always deeply struck by pictures of these magnificent systems. I'm thus pleased that I got the opportunity to work on these objects and to study processes related to their formation and evolution. I have also been lucky to live at the right time for this subject. Galaxy formation and cosmology might well be in their Golden Age right now. The last ten to twenty years have witnessed a dramatic progress towards an understanding of the formation of structure in the Universe. Satellite missions like *IRAS*, *COBE*, and *ROSAT* have revolutionized the field. The Hubble-Space Telescope has opened up the high-redshift Universe, and the new class of 8-10 meter telescopes are producing loads of fascinating data, just awaiting scientific exploitation. Simultaneously, the rapid advances of computer technology have fundamentally transformed the way research in theoretical astrophysics is done. I have always been fascinated by the growing complexity of information technology, and much of the work presented in this thesis tries to exploit computer simulations for the study of the highly non-linear processes occurring in galaxy formation.

1.2 Historical perspectives

Astronomy might well be the oldest scientific discipline of mankind. The stars on the night sky have always exerted a powerful influence on human culture and religion, and even today, many people entertain the belief that their fate is determined by stellar constellations. Many of the ancient cultures of the past also developed an advanced scientific understanding of astronomy. Unfortunately, most of this knowledge was lost when these civilizations were destroyed in the course of history. Only the astronomical knowledge of the Greek culture survived, and provided the foundation of modern astronomy. Every astronomer is reminded of this antique heritage in daily work when the arcane magnitude system is used to quantify the brightness of astrophysical objects.

The Greeks still believed in an eternal Universe. Plato and Aristoteles thought it to be ancient, without being subject to any evolution. Only towards the end of the Middle Ages, new views of the Universe started to emerge. Kopernikus discovered that the Earth revolves around the Sun, and Giordano Bruno realized in 1580 that stars are distant suns which might possibly have planets – he was burned on the stake for this revolutionary idea. Later, Galileo Galilei interpreted the Milky Way as an assembly of stars.

A big step forward was made, when Isaac Newton formulated the laws of gravity and mechanics, and hence laid out the foundation of classical physics. Using Newton's ideas, the philosopher Immanuel Kant explained the Milky Way as a rotating system of stars, where the centrifugal force balances the gravitational attraction. He also speculated about the existence of distant galaxies that are like our own.

These early steps in cosmology are very impressive, especially when one considers the limited observational techniques available at the time and the pre-enlightened society that could make scientific work even life-threatening. However, the most spectacular advances of modern cosmology really had to wait until this century.

It was in particular Albert Einstein, who laid out the very basis of our current view of an expanding Universe. With his general theory of relativity he revolutionized our ideas about space and time. This achievement Einstein's is also a marvelous lesson about the power of pure thought.

With the advent of general relativity it was soon realized that it permitted cosmological solutions describing expanding or contracting universes. Friedmann and Lemaitre were then the first who worked out the family of homogeneous and isotropic cosmologies. They thus may be viewed as fathers of the Big-Bang models, although it probably does injustice to the many other fathers that have contributed to this emerging theory. It is interesting to note that Einstein introduced the cosmological constant into his theory, because he tried to allow a static universe as a solution of the field equations. When it was realized that the Universe is really expanding, he dismissed it as the "greatest blunder of his life". However, the cosmological constant has been haunting astronomers again and again. Most recently, observations of distant supernovae have accumulated substantial evidence suggesting that the cosmological constant might actually be non-zero. In fact, theoretical models for structure formation had been favouring a cosmological constant even before these observations.

In 1929, Edwin Hubble published his famous 'Hubble-law', which is of fundamental importance for cosmology and describes the observational fact that distant galaxies are receding from us with velocities proportional to their distance. Hubble's finding immediately suggested that space itself is expanding, and that the Universe was a lot smaller and denser in the past. Based on this observation, Einstein and de Sitter proposed in 1932 a world model based on a pressure-less critical density cosmology. This model is still one of the standard background scenarios, and one of the most attractive cosmogonies on theoretical grounds. George Gamov then realized in 1946 that the Universe was dense and hot enough in its early phase to allow fast thermonuclear reactions. In particular, he realized that these reactions would synthesize light elements. Together with Alpher and Bethe, he then predicted in 1948 the cosmic background radiation.

Penzias and Wilson discovered this cosmic background radiation in 1965 more or less by chance. Dicke had been planning an experiment specifically designed to search for it, but he came too late. Although Dicke and Peebles immediately realized that the 'noise' in Penzias and Wilson's detector was the background radiation searched for so long, only Penzias and Wilson received the Nobel Prize. In the 1990s, the COBE satellite discovered tiny temperature fluctuations in the microwave background, and thus detected traces of the seeds of structure formation in the early universe. Nowadays, the expansion of the Universe, the cosmic microwave background radiation, and the synthesis of the light elements are the pillars of modern big-bang cosmologies.

Galaxy formation itself is really a very young subject. The first quantitative attempts to understand it date back just 20 years ago. In the late seventies, the pioneering work of Rees & Ostriker (1977) and White & Rees (1978) started up the modern picture of galaxy formation. In was in particular the study of White & Rees (1978) that laid out the main concepts of hierarchical galaxy formation. They for the first time computed the luminosity function of galaxies based on the assumption that galaxies form by cooling processes inside dark matter haloes. In the 1980s, these cold dark matter cosmogonies have started to flourish, and they have now become the most widely favoured model which is very actively investigated in current research.

1.3 Physical processes in galaxy formation

There are many excellent monographs on modern cosmology (e.g. Weinberg 1972; Peebles 1993; Padmanabhan 1993; Peacock 1999), so I will only give a very brief and limited account of the field in this introduction. The following sections are just meant to provide a framework for my further studies, and to put them into perspective within astrophysics. Note that a complete review of the current status of galaxy formation theory would easily fill several hundred pages and is therefore beyond the scope of this thesis.

Cosmology draws just about from all fields of physics, and the breadth it has thus acquired can be both a source of great stimulation and intimidation for a young researcher. A graduate student in cosmology has to catch up with the rapidly advancing theoretical understanding of the universe, and he has to struggle with observational astronomy, with its often obscure terminology and arcane conventions.

Within cosmology, galaxy formation is one of the most complicated phenomena. It is an inherently non-linear, three dimensional process, with a multitude of diverse physical processes operating on vastly different scales. Even viewed in isolation, some of the important processes in galaxy formation are only poorly understood, and often their relevance for the problem of galaxy formation can be judged only vaguely. Is seems clear however, that all of the processes in the following (incomplete) list are relevant to some degree: (1) Evolution of the background cosmology. (2) Nonlinear gravitational collapse. (3) Infall and shock heating of gas. (4) Radiative cooling and heating processes. (5) Star formation. (6) "Feedback" by supernova explosions and stellar winds. (7) Galaxy interactions and mergers. (8) Chemical enrichment and stellar evolution. (9) Galactic winds. (10) Radiation transfer.

Below I will briefly discuss some of the most basic aspects of these processes. Note that I will not talk about the early universe, and its crucial processes of big bang nucleosynthesis and inflationary cosmology, nor do I have room to give an account of the cosmic microwave background and the thermal history of the universe.

1.3.1 The homogeneous universe

I will assume that the cosmological principle holds, i.e. the Universe can be approximated as being homogeneous and isotropic, at least on large enough scales. Then the spacetime of the Universe can be described by the Robertson-Walker metric, viz.

$$\mathrm{d}s^2 = c^2\,\mathrm{d}t^2 - a^2(t)\left(\frac{\mathrm{d}r^2}{1-kr^2} + r^2\,\mathrm{d}\Omega^2\right). \tag{1.1}$$

Here, spherical coordinates (r, θ, ϕ) are used to describe spatial positions, and $\mathrm{d}\Omega^2 = \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2$ is a shorthand for the square of the solid angle. Note that the coordinates are comoving with the expansion of the Universe, whose spatial extent grows proportional to the dimensionless expansion parameter $a(t)$. By convention, $a(t_0) = 1$ at the present time t_0. Depending on the value of k, there are three different global geometries for the Universe, commonly refered to as closed ($k > 0$), flat ($k = 0$) and open ($k < 0$).

Einstein's field equations simplify for the Robertson-Walker metric and for the assumption that the contents of the Universe can be described as an ideal fluid. One then obtains the Friedmann equations

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda c^2}{3}, \tag{1.2}$$

and

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2} + \frac{\Lambda c^2}{3}. \tag{1.3}$$

Here I also allowed for a cosmological constant Λ in Einstein's equations. In the Newtonian limit of the field equations, it is easy to see that Λ effectively acts as an energy density $\rho_\Lambda c^2 = c^4\Lambda/8\pi G$ of the vacuum.

In order to solve the Friedman equations, an equation of state for the cosmic fluid needs to be specified. There are three main components of the material content of the Universe: baryons, dark matter and relativistic matter, like photons or neutrinos. Note that I will assume that the dark matter consists of some form of a weakly interacting massive particle (WIMP). In the early Universe, the energy density is dominated by radiation and relativistic particles, hence the equation of state is given by $p = \frac{1}{3}\rho c^2$. However, as the Universe expands, the energy density of radiation decays as a^{-4}, while that of non-relativistic matter just scales as a^{-3}. From the temperature of the cosmic microwave background we can then infer that ordinary matter dominated over radiation for most of the history of the Universe. In this phase, the pressure can be neglected altogether (dust universe). In this study I will only be concerned with lookback times that are always in the matter-dominated phase.

The Hubble constant is defined as the expansion rate at the present time t_0:

$$H_0 = \left.\frac{\dot{a}}{a}\right|_{t_0}. \tag{1.4}$$

Often, I will parameterize the value of H_0 by $H_0 = 100\,h\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, where h is a dimensionless fudge factor that will appear in many quantitative results in this work. The matter content of the universe is usually expressed in terms of the critical density

$$\rho_\mathrm{c} = \frac{3H_0^2}{8\pi G}, \tag{1.5}$$

allowing the definition of useful cosmological parameters:

$$\Omega_0 = \frac{\rho}{\rho_c}, \quad \Omega_\Lambda = \frac{\Lambda c^2}{3H_0^2}, \quad \Omega_c = -\frac{kc^2}{a_0^2 H_0^2}. \tag{1.6}$$

Note that these definitions refer to the values of the corresponding quantities at the present time. If desired, the definitions can also be used to define time dependent cosmological parameters. Upon division by H_0^2, the Friedmann equation (1.3) at the present time becomes $\Omega_0 + \Omega_c + \Omega_\Lambda = 1$, while (1.2) takes the form $q_0 = \frac{1}{2}\Omega_0 - \Omega_\Lambda$. Here, the deceleration parameter

$$q_0 \equiv - \left.\frac{\ddot{a}a}{\dot{a}^2}\right|_{t_0} \tag{1.7}$$

was introduced; it is sometimes used to parameterize cosmological background models. Using the above definitions, a simple matter-dominated universe may thus be completely specified by the cosmological parameters $(H_0, \Omega_0, \Omega_\Lambda)$ or (H_0, Ω_0, q_0). Once these parameters are known, the expansion rate as a function of size is given by

$$H(a) = \frac{\dot{a}}{a} = H_0 E(a), \tag{1.8}$$

where

$$E(a) = \sqrt{\Omega_0 a^{-3} + (1 - \Omega_0 - \Omega_\Lambda)a^{-2} + \Omega_\Lambda}\,. \tag{1.9}$$

These basic relations may also be used to compute the lookback time to a given size of the Universe in the past, i.e.

$$t(a) = \int_a^1 \frac{\mathrm{d}a}{aH_0E(a)}. \tag{1.10}$$

At present, it is still one of the major goals of observational cosmology to determine the numerical values of the cosmological parameters.

The perhaps simplest cosmological model is a flat universe with zero cosmological constant. It is commonly called Einstein-de-Sitter universe and is characterized by $\Omega_0 = 1$ and $\Omega_\Lambda = \Omega_c = 0$. Since the simplest inflationary theory predicts a critical density universe, this model is very attractive on theoretical grounds. However, in the last ten years evidence has been constantly growing that we probably live in a low-density universe with $\Omega_0 \sim 0.3$. Very recently, there have also been strong observational hints that the cosmological constant is non-zero. In fact, the observations of distant supernovae suggest that $\Omega_\Lambda \sim 0.7$, implying that the universe might be vacuum dominated at the present time, and that the rate of expansion is actually accelerating ($q_0 < 0$).

Objects at cosmological distances are seen redshifted by $z = (\lambda_{\rm obs} - \lambda_{\rm em})/\lambda_{\rm em}$, where $\lambda_{\rm obs}$ is the observed wavelength, and $\lambda_{\rm em}$ is the wavelength in the emitter's rest frame. This cosmological redshift z is of fundamental importance in the study of the Universe because it can be directly measured with observations of spectra. If one neglects peculiar motions and solves the geodesic equation of the observed light ray one finds that the redshift can be expressed as

$$z + 1 = \frac{1}{a(t_{\rm em})}. \tag{1.11}$$

This redshift arises because space itself expands, and it is therefore quite different from the ordinary Doppler effect. The above relation characterizes both the distance of the object and the epoch of the emission of the light. The redshift may therefore serve as a convenient time variable, and I will frequently employ it in that sense.

1.3.2 Gravity and the growth of structure

Above I have briefly discussed the dynamics of homogeneous universes. Clearly, the observed Universe is far from being homogeneous – there is an enormous richness of structure ranging from dwarf galaxies, to groups and clusters of galaxies, and finally to huge superclusters. For the observer, galaxies appear as islands of stars, with stellar densities 10^8 times larger than the mean stellar density. Their typical size is roughly 10 kpc, yet their average distance is of order 1 Mpc.

Nevertheless, the common assumption is that on scales somewhat larger than 100 Mpc homogeneity is approached. If the Universe is smoothed on larger scales, it will appear homogeneous. Despite its small-scale lumpiness the global dynamics of the Universe can then still be described by the Friedmann-Lemaitre models.

The currently favoured theories of structure formation in the Universe assume that the structure grows due to gravitational instability out of primordial fluctuations in the density. Note that these theories are in principle completely specified by their initial conditions. In the simplest case, these are already fully given by the form of the initial fluctuation spectrum, and by the values of the cosmological parameters. However, the problem encountered in practice is that it is highly non-trivial to compare the predictions of such a model (the clustering of dark matter) with the objects that are actually observed (the galaxies).

At this point, it is worth pointing out that the dominant interaction that governs the dynamics and evolution of galaxies, stellar systems in general, and of large-scale structure in the Universe is gravitation. Although dissipative gas physics is clearly important for galaxy formation, it likely plays only a minor role for the structure of elliptical galaxies, for clusters of galaxies, for the dynamics of large-scale structure, and for other gas-poor objects. The relative importance of the gravitational interaction is corroborated since it appears that the majority of the mass of the Universe consists of dark matter, which has been detected only by its gravitational interactions. Modern cosmology assumes that there is a class of yet unidentified particles contributing as dark matter to the mass density. While the dark matter has not been directly detected on Earth so far, there is a lot of indirect evidence that demonstrates its existence. This indirect evidence stems from the dark matter's gravitational effects on its surroundings, seen for example in the orbital motions of galaxies within clusters, or in the gravitational lensing of background galaxies by massive foreground objects, in the rotation curves of galaxies, or in the kinematics of their satellites.

The dark matter is believed to interact with other matter *only* by gravity. This idea brings a great simplification, because the dynamics of the dark matter can then be studied using the laws of gravity alone, without having to worry about the much more complicated physical processes related to the gas that ultimately forms the stars in the visible parts of galaxies.

The study of the growth of density perturbations under their self-gravity in an expanding universe is therefore one of the major topics in cosmology. As long as the fluctuations are small, the equations of motion can be linearized by expanding around the homogeneous solution. The resulting *linear theory* together with its higher-order extensions are among the best methods to study the formation of large-scale structure. However, especially in the context of galaxy formation, analytic techniques are severely limited by the inherently non-linear character of gravity. In galaxies, the mean density is more than $\sim 10^7$ times larger than the background density, making linear theory useless. There are a few special analytic solutions for the nonlinear collapse of structures, most notably the spherical collapse model. While these solutions offer invaluable insight into the relevant dynamics, they provide no adequate description for the highly asymmetric processes occurring in hierarchical galaxy formation.

Consequently, direct numerical simulation of the gravitational dynamics has become an indispensable tool in cosmology. At the heart of these methods lies the gravitational N-body problem, which describes the dynamics of a collection of N point particles of comparable mass under their mutual gravity. For example, the stars in a galaxy, or the dark matter 'particles' of the Universe are often modeled as N-body systems. Below I will briefly try to summarize the basics of linear theory and cosmological N-body simulations in turn.

1.3.2.1 Linear theory

Dark matter is assumed to behave as a collisionless fluid for most of the history of the Universe. Since the number of particles is believed to be very large, gravitational two-body scattering events occur at such a low rate, that the system can be accurately described in terms of a smooth one-particle distribution function $f(\mathbf{r}, \mathbf{v}, t)$ in phase space. This distribution function can be interpreted as a number or mass density, giving the probability of finding a particle near the phase-space point $(\mathbf{r}, \mathbf{v})$ at time t.

Assuming that there are no collisions between particles, the phase-space evolution is given by the collisionless Boltzmann equation (CBE), also known as Vlasov equation. It simply describes conservation of phase-space density along a particle trajectory, viz.

$$\frac{\mathrm{d}f}{\mathrm{d}t} \equiv \frac{\partial f}{\partial t} + \mathbf{v}\frac{\partial f}{\partial \mathbf{r}} - \frac{\partial \Phi}{\partial \mathbf{r}}\frac{\partial f}{\partial \mathbf{v}} = 0. \tag{1.12}$$

This states that the phase space fluid f streams freely in the self-consistent potential Φ, which is obtained as solution of Poisson's equation

$$\nabla^2\Phi(\mathbf{r}, t) = 4\pi G\rho(\mathbf{r}, t), \tag{1.13}$$

where the density

$$\rho(\mathbf{r}, t) = \int f(\mathbf{r}, \mathbf{v}, t)\mathrm{d}\mathbf{v} \tag{1.14}$$

is simply a moment of the distribution function. We may also define a mean velocity field as

$$\mathbf{v}(\mathbf{r}, t) = \frac{1}{\rho}\int \mathbf{v}f(\mathbf{r}, \mathbf{v}, t)\mathrm{d}\mathbf{v}\,. \tag{1.15}$$

Taking further moments with respect to velocity, it is easy to derive an equation for the mean streaming that corresponds to the familiar Euler equation of fluid dynamics (Binney & Tremaine 1987):

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{v} = -\nabla\Phi - \frac{1}{\rho}\left(\mathbf{P}\cdot \overleftarrow{\nabla}\right). \tag{1.16}$$

The last term on the right hand side describes an anisotropic kinetic pressure arising from random motions of the particles. This stress tensor $\mathbf{P}$ is defined as

$$P_{ij} = \rho\left(\langle v_i v_j\rangle - \langle v_i\rangle\langle v_j\rangle\right) \tag{1.17}$$

where

$$\langle v_i v_j\rangle = \frac{1}{\rho}\int v_i v_j f(\mathbf{r}, \mathbf{v}, t)\mathrm{d}\mathbf{v}, \tag{1.18}$$

and $\langle v_i\rangle$ is just a component of the mean velocity field defined in equation (1.15). For a collisionless gas, the velocity distribution may be far from isotropic, and the resulting shear stress can give rise to a collisionless damping process called free-streaming damping. In simple words, particles can damp perturbations by flying away from overdense or underdense regions at their thermal speeds. This process is relevant for so-called 'hot' dark matter. However, if one is dealing with 'cold' dark matter, the pressure term can be neglected in linear theory. Here, one is usually interested in a regime where the scale λ of a perturbation is larger than the Jeans length $\lambda_{\mathrm{J}} = (\pi\sigma^2/G\rho)^{1/2}$, i.e. gravity is more important than pressure. Note however that the pressure term is bound to become important for the structure of nonlinear collapsed objects.

We will now discuss the evolution of perturbations in an expanding universe. It is convenient to employ so-called comoving coordinates $\mathbf{x}$ defined by

$$\mathbf{r} = a\mathbf{x}, \tag{1.19}$$

where a is the scale factor. These coordinates take out the uniform expansion of the background metric. The physical velocity $\dot{\mathbf{r}}$ of a particle can then be written as

$$\dot{\mathbf{r}} = \dot{a}\mathbf{x} + a\dot{\mathbf{x}}, \tag{1.20}$$

where the motion relative to the comoving frame, $\mathbf{v} \equiv a\dot{\mathbf{x}}$, is called the peculiar velocity, and $\dot{a}\mathbf{x} = H\mathbf{r}$ is the Hubble flow. Note that the velocity $\mathbf{v}$ used in the discussion above (equations 1.12 to 1.18) is really $\dot{\mathbf{r}}$, and not the peculiar velocity.

In principle, the proper treatment of the growth of structure in the universe requires the use of general relativity. However, in the limit of non-relativistic motions, and for fluctuations well inside the horizon, the relevant equations reduce to the Newtonian limit. The only difference lies in the occurrence of the cosmological constant, describing effectively a pressure of the vacuum. In physical coordinates, the 'Euler' equation in the Newtonian limit then becomes

$$\frac{\mathrm{d}\dot{\mathbf{r}}}{\mathrm{d}t} = \frac{\partial \dot{\mathbf{r}}}{\partial t} + (\dot{\mathbf{r}} \cdot \nabla_{\mathbf{r}})\dot{\mathbf{r}} = -\nabla_{\mathbf{r}}\Phi - \frac{\nabla_{\mathbf{r}}P}{\rho} + \Omega_\Lambda H_0^2\mathbf{r}, \tag{1.21}$$

where I adopted an isotropic pressure $P = \rho\sigma^2$, for simplicity. In the following, I will drop the pressure entirely, which is appropriate for a cold collisionless fluid.

We know rewrite the Euler equation in comoving coordinates. Note that $\ddot{\mathbf{r}} = a\ddot{\mathbf{x}} + 2\dot{a}\dot{\mathbf{x}} + \ddot{a}\mathbf{x}$ and $\nabla_{\mathbf{r}} = \frac{1}{a}\nabla_{\mathbf{x}}$. The term involving $\ddot{a}$ can be eliminated using equation (1.2), which may be written as

$$\frac{\ddot{a}}{a} = \Omega_\Lambda H_0^2 - \frac{4\pi G}{3}\overline{\rho}. \tag{1.22}$$

One then obtains

$$\ddot{\mathbf{x}} + 2\frac{\dot{a}}{a}\dot{\mathbf{x}} = -\frac{1}{a}\nabla_{\mathbf{r}}\Phi + \frac{4\pi G}{3}\overline{\rho}\,\mathbf{x}. \tag{1.23}$$

Note that the Ω_Λ-term has cancelled. It is only implicitly present through the equation governing the expansion rate. The term linear in $\mathbf{x}$ describes the attractive force from the mass inside a sphere of radius $|\mathbf{x}|$ with background density. By defining the *peculiar gravitational potential* as

$$\phi(\mathbf{r}) = -G\int \frac{\rho(\mathbf{r}') - \overline{\rho}}{|\mathbf{r} - \mathbf{r}'|}\,\mathrm{d}\mathbf{r}' \tag{1.24}$$

this term can be incorporated into the potential gradient. One then obtains the equation of motion

$$\ddot{\mathbf{x}} + 2\frac{\dot{a}}{a}\dot{\mathbf{x}} = -\frac{1}{a^2}\nabla_{\mathbf{x}}\phi\,. \tag{1.25}$$

This shows that it is really the *density fluctuation field* $\delta\rho(\mathbf{r}) = \rho(\mathbf{r}) - \overline{\rho}$ that causes motions in the Newtonian limit. Note that the formulation of equation (1.25) also treats the boundary conditions in Newtonian theory properly.

In linear theory, we expand the equations of motion around the homogeneous universe. To this end, one commonly introduces a dimensionless density contrast

$$\delta(\mathbf{x}, t) = \frac{\rho(\mathbf{x}, t) - \overline{\rho}(t)}{\overline{\rho}(t)}. \tag{1.26}$$

The fluid equations can then be written as

$$\dot{\delta} + (1 + \delta)\nabla\dot{\mathbf{x}} = 0, \tag{1.27}$$

$$\ddot{\mathbf{x}} + 2\frac{\dot{a}}{a}\dot{\mathbf{x}} = -\frac{1}{a^2}\nabla\phi, \tag{1.28}$$

$$\nabla^2\phi = 4\pi G\,\overline{\rho}\,a^2\delta\,. \tag{1.29}$$

The first of these equations is just the continuity equation. Note that all spatial derivatives in these three equations now refer to the comoving coordinates $\mathbf{x}$. Combining these relations and assuming that the fluctuation field is small, one can drop all terms of higher than linear order in δ and $\dot{\mathbf{x}}$. Note that $\dot{\mathbf{x}}$ vanishes for the homogeneous universe. One then obtains

$$\ddot{\delta} + 2\frac{\dot{a}}{a}\,\dot{\delta} = 4\pi G\,\overline{\rho}\,\delta \tag{1.30}$$

as a single equation for the density contrast. This linear equation in δ looks just the same in Fourier space, i.e. all Fourier modes are decoupled and grow independently from each other. In general, equation (1.30) has two independent solutions, a growing and a decaying mode. For $\Omega_0 = 1$ the solutions are particularly simple. The growing mode is $\delta \propto t^{2/3} \propto a$, and the decaying one is $\delta \propto t^{-1}$. It is interesting to note that the Hubble parameter $H(t) = \dot{a}/a$ itself fulfills equation (1.30). Since $H(t)$ decreases with time, at least as long as the universe is not dominated by the cosmological constant, $H(t)$ is a decaying solution. The growing solution can then be obtained from $H(t)$ by observing that the Wronskian $W = |\dot{\delta}_a \delta_b - \delta_a \dot{\delta}_b|$ of two independent solutions δ_a, δ_b of equation (1.30) has the value a^{-2}. For a general cosmology, the growing solution is thus

$$\delta(t) \propto D(t) = H(t) \int_0^t \frac{\mathrm{d}t'}{a^2(t') H^2(t')} = H(t) \int_0^{a(t)} \frac{\mathrm{d}a}{\dot{a}^3} \,. \tag{1.31}$$

1.3.2.2 Zel'dovich approximation

Zel'dovich suggested another approach to linear theory. Using the kinematical ansatz

$$\mathbf{x}(t) = \mathbf{q} + b(t)\, \mathbf{s}(\mathbf{q}), \tag{1.32}$$

his method assumes that the particles continue to move in the direction of their initial displacement. Here, the comoving coordinates $\mathbf{q}$ are the original, unperturbed *Lagrangian* positions of the particles. Equation (1.32) provides a mapping from the homogeneous particle distribution described by $\mathbf{q}$ to the perturbed distribution described by the *Eulerian* coordinates $\mathbf{x}$. Linearizing the Jacobian of this mapping, it is easy to show that the density perturbation is given by

$$\delta = -b(t)\, \nabla \cdot \mathbf{s}. \tag{1.33}$$

Since the first order Lagrangian perturbation theory needs to agree with the first order Eulerian perturbation theory treated above, we can infer that $b(t)$ must be equal to the growing solution $D(t)$ – otherwise we do not obtain $\delta(t) \propto D(t)$. The Zel'dovich approximation is therefore given by

$$\mathbf{x}(t) = \mathbf{q} + D(t)\, \mathbf{s}(\mathbf{q}). \tag{1.34}$$

Comparison with equation (1.28) and (1.30) shows that the direction of the particle displacement $\mathbf{d}(t) \equiv \mathbf{x}(t) - \mathbf{q}$ is given by the gradient of the peculiar potential, viz.

$$\mathbf{d} = -\frac{\nabla \phi}{4\pi G\, \overline{\rho}\, a^2}. \tag{1.35}$$

Similarly, the peculiar velocity is proportional to the displacement field

$$\mathbf{v} = a\dot{\mathbf{x}} = a\frac{\dot{D}}{D}\, \mathbf{d} = aH(a) f(\Omega)\, \mathbf{d}. \tag{1.36}$$

The last term expresses $\mathbf{v}$ in terms of the logarithmic derivative

$$f(\Omega) \equiv \frac{\mathrm{d}\log D}{\mathrm{d}\log a} \simeq \Omega(a)^{0.6} = \left[\frac{\Omega_0}{a + \Omega_0(1-a) + \Omega_\Lambda(a^3 - a)}\right]^{0.6} \tag{1.37}$$

of the growth factor. The approximation $f(\Omega) \simeq \Omega(a)^{0.6}$ is quite accurate and therefore often used instead of a cumbersome exact computation of the growth factor with equation (1.31). In particular, one may now express the peculiar velocity as

$$\mathbf{v} = \frac{2}{3}\frac{f(\Omega)}{H\Omega}\mathbf{g}, \tag{1.38}$$

where $\mathbf{g} = -\nabla\phi/a$ is the peculiar acceleration.

The advantage of the Zel'dovich approximation lies in its ability to still provide a reasonable approximation when linear theory already starts to break down. In order to construct initial conditions for cosmological N-body simulation, one therefore usually employs the Zel'dovich approximation to initialize particle velocities via equation (1.36).

1.3.2.3 N-body simulations

Although linear theory is an extremely valuable tool to study the early growth of structure, and the growth on scales that are so large that they are still in the linear regime at the present time, it is clearly limited by its inability to describe the non-linear evolution of structure. Using higher order perturbation theory, the validity of the analytic method can be extended somewhat, albeit at the expense of highly increasing mathematical complexity. There are also a small number of tractable analytic models for non-linear gravitational collapse. However, these solutions are limited to very special initial conditions of high symmetry. Computer simulations that directly solve the full non-linear equations provide an alternative approach that has become increasingly popular with the advent of ever more powerful computers. In this thesis, I will make extensive use of such simulation techniques.

As was discussed above, dark matter is assumed to obey the collisionless Boltzmann equation (CBE). In principle, one may thus try to directly solve this six-dimensional partial differential equation. However, this is mathematically very difficult and computationally usually not competitive with a simple alternative approach: In the N-body method, one replaces the smooth function $f(\mathbf{r}, \mathbf{v})$ with a set of N particles, which essentially represent δ-functions in phase space. This system of particles is then evolved under its self-gravity. If enough particles are used, their collective potential will be a good approximation to the true potential of the underlying distribution function and the particles will approximately move along characteristic curves of the true solution of the CBE. The evolved particle system will then provide a fair sampling of the phase-space of the true distribution function $f(\mathbf{r}, \mathbf{v})$. In essence, the N-body method may thus be viewed as a Monte-Carlo technique to solve the CBE.

Large gravitational N-body systems like galaxies can also be described by the CBE (Binney & Tremaine 1987). In this case, the physical system that one tries to model is clearly represented by a large, but finite number of bodies - the stars of the galaxy. Since dark matter is important for the structure of galaxies, there are actually two separate collisionless components, the stars and the dark matter.

In the context of a many-particle phase-space description, the CBE amounts to neglecting particle-particle correlations altogether, i.e. many-particle distribution functions are assumed to simply factorize in products of one-particle distribution functions. This approximations holds as long as the relaxation timescale associated with gravitational 'Rutherford-scattering' is much larger than the natural dynamical time scale of the system. The larger the number N of bodies in an object of a given size, the longer the relaxation time. For galaxies with 10^8 stars it is already far longer than the Hubble time. Note however, that small star clusters with 10^4 stars are usually not well described by the CBE any more. As a note of caution one might also add that although the CBE is physically well motivated, there is often no rigorous proof of its validity (Kandrup 1997). Similarly, estimates of the time scale where it might be expected to fail have only approximative character.

In numerical N-body simulations one is usually forced to use a much smaller number of particles than are actually present in the modeled astrophysical system. For example, state-of-the-art simulations of galaxies use of the order 10^5 particles to represent the stars of an isolated galaxy, some five to six orders of magnitude less than the real number of stars. To prevent excessive two-body relaxation, one resorts to a numerical trick and introduces an explicit softening of the gravitational force, i.e. at small separations the gravitational force is smoothly reduced below the Newtonian value. One thus deliberately introduces a lower resolution cut-off. This reduces gravitational two-body scattering, and it also simplifies the numerical orbit integration by allowing larger timesteps. More technical details about our numerical techniques will be given in Chapter 5.

In this study, I will use two 'types' of N-body simulations. Chapters 2 and 3 are concerned with simulations of individual and interacting galaxies. In these simulations, the galaxies are simulated with vacuum boundary conditions, i.e. the background cosmology is neglected. Since the galaxies have already collapsed during their formation and decoupled from the expansion of the universe, this is a fair approximation. One needs to keep in mind though, that this approach neglects the possible further infall of material and the gravitational tidal fields of matter in the surroundings of the galaxies.

In the second class of simulations, I consider 'proper' cosmological simulations. These simulations start shortly after the Big Bang, with a mass distribution that is close to being perfectly homogeneous. The initial Lagrangian particle positions are perturbed according to the fluctuation spectrum predicted by inflationary cosmology (Guth 1981), and the initial particle velocities are set with the Zel'dovich approximation.

The initial fluctuation spectrum of cold dark matter cosmologies can be described by (Efstathiou et al. 1992)

$$P(k) = \frac{Bk}{\left\{1 + \left[ak + (bk)^{3/2} + (ck)^2\right]^\nu\right\}^{2/\nu}}, \tag{1.39}$$

where $a = 6.4\,\Gamma^{-1}\,h^{-1}$kpc, $b = 3.0\,\Gamma^{-1}\,h^{-1}$kpc, $c = 1.7\,\Gamma^{-1}\,h^{-1}$kpc and $\nu = 1.13$. This form is able to fit a wide class of theoretical models where Γ acts as a shape parameter of the power spectrum. The standard, flat, scale-invariant, adiabatic cold dark matter model (SCDM) is accurately fitted for $\Gamma = h$. Flat cold dark matter models with a cosmological constant are also reasonably well described if $\Gamma = \Omega_0 h$. Variants of CDM models that exhibit more large scale power due to decaying neutrinos are well approximated if $\Gamma \simeq \Omega_0 h[0.861 + 3.8(m\,\tau)^{2/3}]^{1/2}$, where m is the neutrino mass in units of 10 keV and τ is its lifetime in years.

The form of equation (1.39) results from a scale-invariant primordial power-spectrum $P(k) \propto k$, modified by the linear growth of fluctuations during the early phase of the universe. Here, the growth rate of individual modes differs depending on whether they are in or outside the horizon, leading to a bend in the power spectrum. A critical prediction of inflationary theory is that the initial fluctuations form a Gaussian random field. In this case, they are completely determined by their power spectrum alone, and all phase correlations vanish. Alternative models for seeds of cosmic structure, e.g. topological defects like cosmic strings, typically introduce such higher-order correlations. In Chapter 6 I will directly test the Gaussian random phase hypothesis by studying the topology of isodensity surfaces in the observed density field of the Universe.

After the start of a cosmological simulation, it first continues to follow the further linear growth of perturbations. However, unlike the analytic techniques, it is then capable of accurately following the evolution into the mildly non-linear, and finally into the highly non-linear regime. In this way, the simulation is able to correctly compute the hierarchical growth of structure.

In universes dominated by cold dark matter, primordial fluctuations survive on small scales. The first structures that undergo non-linear collapse are then small objects of sub-galactic mass. These objects then cluster together, and merge hierarchically into larger and larger objects. Galaxy formation thus proceeds in a *bottom-up* picture, where larger galaxies tend to form later than smaller ones. In this picture, huge clusters of galaxies, which are the most massive non-linear objects at the present time, form very recently.

In models with hot dark matter galaxy formation would proceed very differently. Due to the damping of fluctuations on small scales, the fluctuation spectrum would be dominated by scales that are close to the horizon at matter-radiation equality $z_{\rm eq}$. As a result, the first structures that form are clusters and superclusters, which then need to fragment into smaller objects like galaxies. This *top-down* picture for galaxy formation has been successfully ruled out, simply because it proved to be incompatible with observations. However, there is still the possibility that a small admixture of hot dark matter contributes to the density of universe.

In the cosmological simulations of this work, discussed in Chapter 4, I specifically study the formation of a rich cluster of galaxies. By using substantially more particles than in all so far published work on this problem, the mass resolution of these simulations has been greatly increased. The resulting highly resolved dynamics of the dark matter is then used to construct the galaxy population of the cluster using semi-analytic techniques. In this way, an explicit model for the formation of cluster galaxies is constructed.

1.3.3 Hydrodynamics

In the currently favoured class of cosmogonies it is assumed that most of the mass of the Universe is in the form of cold dark matter, which interacts only gravitationally. However, there is of course also a contribution to the matter density made up of ordinary baryonic material. After all, the baryons are responsible for the luminous stars we see, and for the life on this planet.

One can try to model the baryonic gas of the Universe along with the dark matter. The gas component is gravitationally coupled to the dark matter, but unlike collisionless material it also feels hydrodynamical forces arising from its thermal pressure. This apparently simple difference gives already rise to the rich phenomenology of gaseous shocks and turbulence. So even in the simplest case of a non-radiative gas, hydrodynamics adds substantial physical complexity. Note that the highly simplified assumption of a non-radiative gas is a reasonable approximation for the hot plasma in clusters of galaxies, and has hence immediate applications in this context.

At the next level of complexity, radiative heating and cooling processes can be taken into account. In contrast to the dark matter, the baryonic component can lose part of its internal energy by radiative processes. White & Rees (1978) were the first who suggested that galaxy formation essentially proceeds in two stages, for which these cooling processes are of critical importance. First, dark matter haloes form in a collisionless gravitational collapse, and then baryons sink to the centers of these haloes. The condensation of baryons in the halo centers is only possible because the gas can dissipate energy by emitting radiation. Radiative cooling thus plays a primary role in the formation of galaxies, but it is also required to understand properties of the intergalactic medium such as the absorption lines in QSO spectra generated by gas clouds along the line of sight.

The gasdynamical equations adopted in cosmological studies are usually those of an inviscid ideal gas, supplemented with optional heating and cooling terms in the energy equation. Since the densities in astrophysical plasmas are usually very low, this description is sufficient. The relevant equations (continuity equation, Euler equation, and energy equation) are then

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} + \rho\,\nabla\cdot\mathbf{v} = 0, \tag{1.40}$$

$$\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = -\frac{\nabla P}{\rho} - \nabla\Phi, \tag{1.41}$$

$$\frac{\mathrm{d}u}{\mathrm{d}t} = -\frac{P}{\rho}\nabla\cdot\mathbf{v} + \frac{\Gamma_{\mathrm{heat}} - \Lambda_{\mathrm{cool}}}{\rho}, \tag{1.42}$$

where u is the specific energy per unit mass, and heating and cooling terms were introduced in the energy equation. Usually, an adiabatic equation of state $P = (\gamma - 1)\rho u$ is assumed. For simplicity, I have written these equations in their Newtonian form. The adaptation to an expanding universe is discussed in Chapter 5.

Most of the relevant radiative heating and cooling processes in cosmology are well understood by atomic physics, and the rates of these processes can be computed precisely. However, in some astrophysical situations a multi-species chemistry together with proper radiative transfer calculations are required to accurately follow the radiation processes. In this study, I will not be concerned with situations that reach such a degree of complexity.

The astrophysical plasma encountered in studies of galaxy formation is an ionized mix of hydrogen and helium of primordial abundance, which may also be polluted by a small admixture of 'metals', i.e. with elements more massive than helium. Under the assumption of collisional ionization equilibrium, the cooling rate of such a plasma can be computed as a function of metallicity. Note that the cooling rate depends quite strongly on metallicity. The treatment of chemical enrichment in galaxy formation models can therefore be very important.

From a technical point of view, it is not very difficult to include radiative cooling into gasdynamical simulations. One only needs to look up tabulated results for the cooling function, and then use it in the thermal energy equation. However, the inclusion of cooling *alone* can lead to the problem of *over-cooling*. This describes the formation of extremely dense and cold gas clumps. In reality it is assumed that star formation and accompanying feedback processes prevent the formation of these knots of cold and extremely dense gas. Hence the inclusion of cooling more or less requires the inclusion of these processes as well.

1.3.4 Star formation, feedback and other trouble

In the first step of galaxy formation, dark matter forms virialized haloes in a process known as violent relaxation. Gas then falls into the potential wells provided by the dark matter and heats up by shocks. Inside the haloes, the baryonic gas becomes dense enough to participate in radiative cooling processes at an appreciable rate. It will thus lose pressure support and sink to the center of the halo, where it eventually will form luminous stars.

Up to this point, the general framework of galaxy formation is relatively well known, both analytically and numerically. However, the severe problem with actual star formation is that important parts of it are not understood in detail. What seems relatively clear is that the condensation of gas leads to the formation of a complex multi-phase interstellar medium. Among the phases one also expects a population of giant molecular clouds. These complexes may fragment or collide with each other. Eventually, the cores of molecular clouds may undergo collapse down to stellar densities, with the subsequent onset of thermonuclear reactions. It is not well understood, what triggers the collapse, nor what the mass spectrum of the formed population of stars is going to be.

Clearly, in an actively star forming region, violent things go on, and the gas physics is expected to be highly non-linear. When a new stellar association has formed, its

UV radiation can ionize gas and significantly alter the thermodynamic properties of the interstellar and intergalactic medium. The stars may also affect the surrounding gas by powerful mass-loaded winds. Furthermore, most massive stars will only live a short time, and then explode as supernovae. The powerful shock waves of supernovae can blow holes into the interstellar medium, or they might even blow it away entirely in small dwarf galaxies.

All these processes associated with star formation deposit energy back into the gaseous medium, and are therefore usually refered to as *feedback*. Unfortunately, the current understanding of the detailed mechanisms of feedback is probably even worse than that of star formation. However, feedback is thought to provide an important regulation mechanism for star formation. For example, the star formation rate in ordinary disk galaxies is relatively constant with time, presumably a result of self-regulation of star formation by feedback. The two problems of star formation and feedback are therefore intertwined and probably require a simultaneous solution.

Given these conceptual difficulties, it is clear that star formation cannot be treated from first principles at this point. It should also be noted that the numerical resolution of currently feasible simulations of whole galaxies is still by orders of magnitude too poor to resolve individual star formation events. Simulation strategies for star formation have therefore usually adopted one of two approaches. In one approach, one focuses on a very small patch of the interstellar medium, say a cold molecular cloud, and then tries to follow its evolution with very high numerical resolution. In this method one tries to incorporate the physics as realistically as possible. In a complementary approach, one attempts to simulate the formation or evolution of whole galaxies. Here, one would already be content with a good description of the global star formation rate and of the global dynamics of the interstellar medium. In this latter approach one therefore resorts to phenomenological models for star formation and feedback, which are motivated by observational data.

In Chapter 3 of this thesis, I will study the dynamics of the baryonic component in simulations of isolated and colliding galaxies. These relatively small simulations provide ideal laboratories to examine the physics of the gaseous phase, and to explore phenomenological models for star formation and feedback. The model I develop turns out to be quite successful for isolated and colliding galaxies, but the inclusion of dissipative gas physics is numerically very costly. In particular, a dissipative cosmological simulation with comparable resolution to the large cluster studied in this work is not yet possible (this will likely change in the near future with the advent of newer generations of computers). This highlights that 'semi-analytic' galaxy formation techniques may currently be the best tool to study galaxy formation using cosmological initial conditions. In such models, the gasdynamical equations are replaced by a set of simplified yet physically motivated approximations for the star formation rate, the feedback, merger timescales, etc. I will use such a method in Chapter 4 to study the formation of cluster galaxies.

In passing, I just want to briefly caution that galaxy formation offers a further variety of potentially important effects, all of them explored to some level, but mostly without final conclusions yet. For example, magnetic fields could be important for the dynamics of the ionized gas if the primordial seed fields are strong enough, or if dynamo processes are effective. Currently it appears that magnetic fields are probably unimportant for galaxy formation; considering the complexities of magneto-hydrodynamics this certainly comes as a relief. Another debated issued is the potential relevance of central, supermassive black holes for the formation of galaxies. Note that even if such black holes play no important role for the actual formation of the galaxy, they will probably strongly influence the structure and dynamics of the galaxies' innermost part.

1.3.5 Mergers of galaxies

As discussed above, mergers of galaxies are a common and inevitable process in hierarchical theories of galaxy formation. In these models, mergers are not only a mechanism for building up ever more massive stellar systems, but they also determine the morphology of galaxies to a large extent.

Observed galaxies exhibit a wide class of morphological types, ranging from almost featureless elliptical galaxies over disk-like objects to magnificent grand design spiral galaxies, which are probably among the most beautiful objects on the sky. When gas cools inside a dark matter halo, it first settles into a rotationally supported disk, because it still carries some of its initial angular momentum. Star formation inside this gaseous disk will then produce a stellar disk. It is thus thought that this generic mode of star formation produces the populous class of disk and spiral galaxies.

When two disk galaxies of comparable mass collide and merge, the disks are destroyed and a spheroidal remnant results, with structural properties resembling elliptical galaxies. According to the *merger hypothesis* it is therefore believed that most, if not all, elliptical galaxies form under the influence of galaxy mergers.

If one studies the morphologies of observed galaxies in detail, one finds that there also exists a large class of 'peculiar' galaxies, which cannot be classified as bona-fide elliptical or disk galaxies. These galaxies often feature long and extended tails of stars, protruding from their main body to distances reaching several hundred kiloparsec. Among the most famous systems of this type are the 'Antennae' galaxies (NGC4038/39), the 'Mice' (NGC4676), or Arp295, and many more of them may be found in Arp's *Atlas of Peculiar Galaxies*.

In a classic paper, Toomre & Toomre (1972) showed with simple numerical experiments that such tails can be ejected by tidal forces in close encounters of dynamically cold disk galaxies. It is now generally accepted that the stellar tails seen in many peculiar galaxies are indeed generated when two galaxies collide with each other. Usually, such encounters lead to a final coalescence of the galaxies, where they form a single pile of stars and lose their original identity.

Recall that the stars seen in a pair of colliding galaxies represent only the visible part of the galaxies. In all viable models of cold dark matter cosmologies, the mass of the dark matter halo belonging to a galaxy is expected to be much larger than the total baryonic mass of the disk itself. Also note that the dark haloes are much more extended than the disks. This points to a potential problem with such models, because disk galaxies embedded in massive and extended dark haloes can have difficulty to form tidal tails. This is because such galaxies experience higher encounter velocities, leading to a smaller overall strength of the perturbation of the disk. In addition, the perturbed material has difficulty to climb out of the deeper potential well. This therefore suggests a potential conflict of cold dark matter models with the observed existence of long tails. In Chapter 2, I will further examine this problem using collisionless N-body simulations of colliding disk galaxies.

When two disk galaxies of similar structure collide, strong gravitational tides during their first encounter can induce a bar instability that quickly transforms the disks into a pair of open bisymmetric spirals. The interstellar gas contained in the disks loses parts of its angular momentum in this process. As a result, the gas is driven towards the center of the galaxies, where the high inflow rate can start and feed a violent starburst. In such a burst, the total star formation rate can reach values of more than $100\mathrm{M}_\odot\,\mathrm{yr}^{-1}$, about two orders of magnitude larger than that in quiescently star forming spiral galaxies. Since the centers of galaxies are heavily obscured by dust, much of the optical and UV radiation of the newly formed stars is re-radiated in the infrared. As a result, such galaxies become extremely luminous in the infrared and often outshine optically bright galaxies. This class of ultraluminous infrared galaxies has been discovered by the *IRAS* satellite, and it is now generally assumed that many of these objects are undergoing an interaction-induced starburst. Some of my hydrodynamical simulations of colliding galaxies in Chapter 3 will examine this phenomenon in more detail.

1.4 Outline

The formation of galaxies is one of the most active fields of research in cosmology. This fact alone illustrates that as of yet no final theory for galaxy formation exists. Obviously, this work cannot provide one either – its aims have to be more modest.

In this thesis, I work mainly on two related aspects of galaxy formation. In one line of research, I study the evolution and interaction of individual disk galaxies. In particular, I will consider major mergers of spiral galaxies, and the formation of elliptical remnants. In a complementary line of research, I apply semi-analytic models for galaxy formation to high-resolution cosmological simulations of the formation of a rich cluster of galaxies. Here, the full galaxy population of a cluster is constructed in the framework of a hierarchical cold dark matter model. A common tool needed in both approaches is a powerful simulation code, which employs sophisticated numerical techniques. Its development represented a substantial part of this work, and may thus be viewed as representing the third line of research in this thesis. Finally, in a fifth line of research I test the Gaussian random phase hypothesis for the initial conditions of galaxy formation using the topological genus measure.

In Chapter 2, I discuss collisionless simulations of mergers between disk galaxies. These numerical experiments focus on the question, whether the structural properties expected for disk galaxies in currently viable cold dark matter models contradict the observed occurrence of long tidal tails in pairs of interacting galaxies.

I will then move on in Chapter 3 to simulations of galaxies that include a modeling of the gaseous interstellar medium and its associated star formation processes. Here, I develop a model for the self-regulation of the star formation rate using a phenomenological approach to feedback. I then apply this model to mergers between gas-rich spiral galaxies, allowing a study of the resulting starbursts, and of the structural properties of the merger remnants.

While these first two chapters deal with simulations of individual galaxies under vacuum boundary conditions, I carry out high-resolution cosmological simulations of the formation of a rich cluster of galaxies in Chapter 4. These simulations achieve a very high mass resolution (up to 69 million particles have been used) and have been run on a massively parallel supercomputer. I develop new techniques to identify and track substructure within the dark matter halo of the cluster and its progenitors, allowing a much more detailed analysis of the dark matter merging history than it has been possible before. I use this substructure information to implement a new semi-analytic scheme for the formation of cluster galaxies, and I show that it leads to remarkably successful results.

In Chapter 5, I will discuss the numerical techniques developed for the simulations presented in this work. The newly written code is a combined N-body gravity/hydrodynamics solver, where gasdynamical interactions are treated with so-called *smoothed particle hydrodynamics*, while the gravitational interaction is computed with a hierarchical tree algorithm. The code employs individual timesteps for all particles, and it has been developed in a serial and a parallel version. The latter was designed to run on massively parallel supercomputers with distributed memory, using explicit message passing routines. The serial version of the code has also been adapted to the special-purpose hardware GRAPE , which may be used instead of the tree-algorithm to compute the gravitational forces.

A critical assumption in current cold dark matter theories is the hypothesis that the primordial fluctuation spectrum was a Gaussian random field. In Chapter 6, I will directly test this hypothesis by examining the topology of the large-scale density field, as measured with the 1.2-Jy redshift survey of *IRAS* galaxies. I compare the genus statistic of the 1.2-Jy survey with numerical N-body models for the clustering of dark matter, and I invest particular care in the statistical methodology of this comparison.

Finally, I will give in Chapter 7 a brief summary of the findings of this thesis, and an outlook for future work.

–Man braucht sich bloß die Mühe zu machen, fünf Minuten nachzudenken, dann begreift man, daß einem das Leben nichts Aufregendes bietet außer ein paar Dingen, die man nicht kaufen kann. Ich machte mein Bier auf und dachte an Betty.

Philippe Djian, Betty Blue

2
Tidal tails in CDM cosmologies

Abstract

We study the formation of tidal tails in pairs of merging disk galaxies with structural properties motivated by current theories of cold dark matter (CDM) cosmologies. In a recent study, Dubinski, Mihos & Hernquist (1996) showed that the formation of prominent tidal tails can be strongly suppressed by massive and extended dark haloes. For the large halo-to-disk mass ratio expected in CDM cosmologies their sequence of models failed to produce strong tails like those observed in many well-known pairs of interacting galaxies. In order to test whether this effect can constrain the viability of CDM cosmologies, we construct N-body models of disk galaxies with structural properties derived in analogy to the analytical work of Mo, Mao & White (1998). With a series of self-consistent collisionless simulations of galaxy-galaxy mergers we demonstrate that even the disks of very massive dark haloes have no problems developing long tidal tails, provided the halo spin parameter is large enough. For our class of models, the halo-to-disk mass ratio is not a good indicator of the ability to produce tails. Instead, the relative size of disk and halo, or alternatively, the ratio of circular velocity to local escape speed at the half mass radius of the disk are more useful criteria. This result holds in all CDM models. While tidal tails can provide useful information on the structure of galaxies, it thus appears unlikely that they are able to constrain the values of the cosmological parameters within these models.

2.1 Introduction

In standard hierarchical scenarios for galaxy formation, mergers of galaxies are common events that lead to the build-up of ever more massive galaxies. In fact, such mergers have been observed for a long time. There is now a large database of well studied examples of merging or strongly interacting disk galaxies, among the most prominent of them are NGC4038/39 (the Antennae), NGC4676 (the Mice), and NGC7252. Many of these pairs feature extended *tidal tails*, with a length that can reach more than $100\,h^{-1}$kpc in projection, or in the extreme case of IRAS19254-7245 (the Superantennae) even $\sim 305\,h^{-1}$kpc from tip to tip (Colina et al. 1991).

The tails originate in close encounters of disk galaxies, when the mutual tidal field ejects disk stars into arcing trajectories that lead to the formation of long tails pointing way from the galaxies, and of bridges connecting them. This process was first demonstrated convincingly in a classic paper by Toomre & Toomre (1972). Later White (1978, 1979) computed the first fully self-consistent 3-dimensional simulations of merging galaxies and established the rapidity of the orbital decay, and the structural resemblence of the merger remnants to elliptical galaxies. This work has been confirmed and extended over the years by simulations with increasingly realistic initial conditions and ever better numerical resolution (Farouki & Shapiro 1982; Farouki et al. 1983; Negroponte & White 1983; Barnes 1988, 1989, 1992; Hernquist 1992, 1993b; Barnes & Hernquist 1996). There have also been quite successful attempts to model particular interacting systems in detail, for example NGC7252 (Hibbard & Mihos 1995) and NGC2442 (Mihos & Bothun 1997).

Recently, Dubinski, Mihos & Hernquist (1996, hereafter DMH) studied the morphology of tidal tails in a series of merging models of disk galaxies with varying halo-to-disk mass ratio. In their sequence of four models, they kept the inner rotation curve very nearly constant and surrounded the disk and the bulge with ever more extended and massive dark haloes. They found that with increasing mass of the dark halo, the resulting tidal tails became shorter and less massive. Their explanation for this effect is simple. For a fixed structure of the disk, a more massive halo leads to a deeper potential well and a higher encounter velocity. As a consequence, the duration and overall strength of the perturbation to the disk is smaller, and the perturbed material cannot as easily climb out of the deeper potential well.

Dubinski, Mihos, and Hernquist have followed up this study with an analysis of NGC7252 (Mihos et al. 1998), and an investigation of different dark matter profiles with a restricted 3-body code (Dubinski et al. 1997). Again they found that disk models with large halo-to-disk mass ratio were not able to produce prominent tidal tails. In particular, they concluded that for mass ratios above 10:1 it should be exceedingly difficult to make tails as long as those observed in systems like NGC7252 or NGC4038/39. Since the currently favoured theoretical values of halo-to-disk mass are considerably higher than this, they speculated that there might be a conflict with cold dark matter (CDM) cosmologies.

In this work we examine the tail-forming ability of realistic models of disk galaxies, where 'realistic' means that their structural properties are motivated to a large degree by current theories of CDM cosmologies. We derive the structural properties of our disk galaxies according to the analytic model of Mo, Mao & White (1998, hereafter MMW), and we collide pairs of these galaxies in self-consistent N-body simulations. We adopt initial conditions for these merger simulations that are favourable for tail formation.

We will demonstrate that the halo-to-disk mass ratio is not sufficient for characterizing ability to make tidal tails. We find that it is the relative distribution of disk and halo material that is relevant, not the mass ratio itself. This conclusion was reached earlier by Barnes (1997, private communication) through analysis of a series of simulations with halo/disk models differing both from those of DMH and from the CDM-based models we use here.

We will show that realistic disk models in CDM cosmologies can produce long and massive tidal tails, provided the spin parameter of their dark halo is large enough. This result is practically independent of the cosmological parameters.

This chapter is organized as follows. In Section 2 we describe the structural properties of our disk models, and in Section 3 we discuss our techniques for setting up N-body representations of these models. A description of the different simulations we have performed is given in Section 4, while Section 5 presents the results. Finally, we summarize and discuss our findings in Section 6.

2.2 Models of disk galaxies

MMW have developed an analytical model for the structure of disk galaxies embedded in cold dark matter haloes. Their model rests on a number of simple yet plausible assumptions, and it is very successful in reproducing the observed properties of disk galaxies. In particular, the predicted population can match the slope and scatter of the Tully-Fisher relation as well as the properties of damped Lyα absorbers in QSO spectra. We take their model as basis to derive the structural properties of our N-body models of disk galaxies. For definiteness, we briefly summarize the relevant assumptions and equations.

2.2.1 Dark haloes

Using high-resolution N-body simulations, Navarro, Frenk & White (1996, 1997, hereafter NFW) established that haloes formed by the gravitational clustering of cold dark matter exhibit a universal structure. Suitably scaled, the density distribution of these dark matter haloes does not depend on cosmology. The NFW-profile is given by

$$\rho(r) = \rho_{\rm crit.} \frac{\delta_{\rm c}}{(r/r_{\rm s})\,(1+r/r_{\rm s})^2}, \tag{2.1}$$

where $\rho_{\rm crit.}$ is the background density at the time of the halo formation, $r_{\rm s}$ is a scale radius, and $\delta_{\rm c}$ is a characteristic overdensity. Note that the slope of this profile is shallower than isothermal at the center, and it gradually steepens outward to an asymptotic slope of -3. Following NFW, we define the *virial radius* r_{200} as the radius with mean overdensity 200, i.e. it contains the *virial mass*

$$M_{200} \equiv 200\rho_{\rm crit.} \frac{4\pi}{3} r_{200}^3\,, \tag{2.2}$$

and we define the *concentration*

$$c \equiv \frac{r_{200}}{r_{\mathrm{s}}} \tag{2.3}$$

of the halo. With these definitions, the characteristic overdensity is given by

$$\delta_{\mathrm{c}} = \frac{200}{3} \frac{c^3}{\ln(1+c) - \frac{c}{1+c}}. \tag{2.4}$$

Further, let

$$v_{200}^2 \equiv \frac{GM_{200}}{r_{200}} \tag{2.5}$$

be the circular velocity at the virial radius. Given the concentration c and the Hubble constant $H(z)$, the radial density profile of a halo may then be specified by anyone of the parameters v_{200}, r_{200}, or M_{200}. In particular, we have

$$M_{200} = \frac{v_{200}^3}{10GH(z)}, \quad \text{and} \quad r_{200} = \frac{v_{200}}{10H(z)}. \tag{2.6}$$

2.2.2 Putting a disk into the halo

We now put a stellar disk into an NFW halo according to the model of MMW. This rests on four key assumptions:

1. The mass M_{d} of the disk is a given fraction m_{d} of the halo mass.
2. The spin J_{d} of the disk is a given fraction j_{d} of the angular momentum J of the halo.
3. The disk has the structure of a thin exponential disk, and it is cold and centrifugally supported.
4. Only disks that are dynamically stable against bar formation correspond to observable disk galaxies.

The angular momentum J of a halo with total energy E is often characterized by the dimensionless spin parameter

$$\lambda = \frac{J|E|^{\frac{1}{2}}}{GM^{\frac{5}{2}}}. \tag{2.7}$$

According to N-body results (Warren et al. 1992; Lemson & Kauffmann 1999), the distribution of λ is well approximated by

$$p(\lambda)\,\mathrm{d}\lambda = \frac{1}{(2\pi)^{1/2}\sigma} \exp\left[-\frac{(\ln\lambda - \ln\overline{\lambda})^2}{2\sigma^2}\right] \frac{\mathrm{d}\lambda}{\lambda} \tag{2.8}$$

with $\sigma = 0.5$ and a typical value $\overline{\lambda} = 0.05$. This distribution is practically independent of cosmology, and of the mass and environment of the haloes (Lemson & Kauffmann

1999). The initial kinetic energy of the spherically symmetric halo may be computed by assuming that all particles move around the center on circular orbits, with speed equal to the circular velocity. This 'trick' results in

$$E_{\rm kin} = \frac{GM^2}{2r_{200}} f_c, \tag{2.9}$$

where

$$f_c = \frac{c\left[\frac{1}{2} - \frac{1}{2(1+c)^2} - \frac{\ln(1+c)}{1+c}\right]}{\left[\ln(1+c) - \frac{c}{1+c}\right]^2}. \tag{2.10}$$

Using the virial relation $E = -E_{\rm kin}$, the angular momentum of the halo then becomes

$$J = \lambda G^{\frac{1}{2}} M^{\frac{3}{2}} \left(\frac{2r_{200}}{f_c}\right)^{\frac{1}{2}}. \tag{2.11}$$

We now put a fraction $m_{\rm d}$ of the initial halo mass into a thin stellar disk with an exponential surface density, viz.

$$\Sigma(R) = \Sigma_0 \exp\left(-\frac{R}{R_{\rm d}}\right) \tag{2.12}$$

with $\Sigma_0 = M_{\rm d}/(2\pi R_{\rm d}^2)$. Here $M_{\rm d} = m_{\rm d} M_{200}$ is the total mass of the disk and $R_{\rm d}$ is its scale radius. The condition

$$J_{\rm d} = j_{\rm d} J \tag{2.13}$$

will then determine the scale radius of the disk, because its spin is given by

$$J_{\rm d} = M_{\rm d} \int_0^{r_{200}} \left(\frac{R}{R_{\rm d}}\right)^2 v_{\rm c}(R) \exp\left(-\frac{R}{R_{\rm d}}\right) {\rm d}R, \tag{2.14}$$

where the circular velocity $v_{\rm c}$ is the sum of two contributions, namely

$$v_{\rm c}^2(R) \equiv R\frac{\partial \Phi}{\partial R} = v_{\rm c,disk}^2(R) + v_{\rm c,dm}^2(R). \tag{2.15}$$

2.2.3 Response of the dark matter profile

We take the gas, that later forms the disk, to be initially distributed just like the dark matter. However, the structure of the dark halo will be changed when the disk forms in its center. We again follow MMW and assume that the dark matter reacts adiabatically to the disk formation. In particular, we assume that the spherical symmetry of the halo is retained, and that the angular momentum of individual dark matter orbits is conserved. This latter condition may be formulated as

$$r_{\rm i} M(r_{\rm i}) = r_{\rm f} M_{\rm f}(r_{\rm f}). \tag{2.16}$$

Here $r_{\rm i}$ and $r_{\rm f}$ are the initial and final radii of some dark matter mass shell, $M(r)$ gives the initial NFW mass profile, and $M_{\rm f}(r)$ is the final cumulative mass profile after the disk is formed. $M_{\rm f}(r)$ is the sum of the cumulative mass of the disk and the dark mass inside the initial radius, i.e.

$$M_{\rm f}(r_{\rm f}) = M_{\rm d}(r_{\rm f}) + (1 - m_{\rm d})M(r_{\rm i}). \tag{2.17}$$

The final profile $M_{\rm h}(r)$ of the dark matter halo is then given by

$$M_{\rm h}(r) = M_{\rm f}(r) - M_{\rm d}(r). \tag{2.18}$$

For a given set of parameters v_{200}, c, $m_{\rm d}$, $j_{\rm d}$, λ, and a formation redshift z, the above equations uniquely determine a disk model. Note that in practice the scale length $R_{\rm d}$ of a disk needs to be determined iteratively in order to satisfy equations (2.13) and (2.17).

2.2.4 Including a bulge

In many galaxies, including the Milky Way, a central bulge population of stars is observed. For spirals like the Milky Way or of later type, the bulge mass is less than 20% of the disk mass. For this reason, the dynamical importance of the bulge in these systems should be small. However, there are also systems with a higher mass fraction in the bulge. While most of the models in this study do not have a bulge, we still want to investigate its possible influence on our results. Hence we here generalize the above model to allow the option of a bulge.

Bulges appear to be flattened triaxial systems, that may be partly supported by rotation. However, Hernquist (1993b) found that it hardly matters for the density and velocity structure of merger remnants whether bulges are spinning or not. For simplicity, we therefore neglect a possible flattening of the bulges and model them as non-rotating spheroids with a spherical Hernquist profile of the form

$$\rho_{\rm b}(r) = \frac{M_{\rm b}}{2\pi}\frac{r_{\rm b}}{r(r_{\rm b}+r)^3}. \tag{2.19}$$

In analogy to the treatment of the disk, we assume that the bulge mass is a fraction $m_{\rm b}$ of the halo mass. Since we take the bulge to be non-rotating it has lost its specific angular momentum either to the halo, or to the disk. We will assume that there is no angular momentum transport between the disk and the dark halo, and none between the disk and the bulge. In this case $j_{\rm d} = m_{\rm d}$.

For simplicity, we further assume that the bulge scale radius $r_{\rm b}$ is a fraction $f_{\rm b}$ of that of the disk, i.e. $r_{\rm b} = f_{\rm b}R_{\rm d}$. Note that the disk half mass radius is $1.678\,R_{\rm d}$, while that of the bulge is $2.414\,r_{\rm b}$.

It is then straight forward to generalize the above disk model to accomodate the bulge. The circular velocity of equation (2.15) gets an additional contribution from the bulge, i.e.

$$v_{\rm c}^2(R) = v_{\rm c,disk}^2(R) + v_{\rm c,dm}^2(R) + v_{\rm c,b}^2(R), \tag{2.20}$$

with $v^2_{\mathrm{c,b}}(R) = GM_{\mathrm{b}}(R)/R$. Further, equation (2.17) needs to be replaced by

$$M_{\mathrm{f}}(r_{\mathrm{f}}) = M_{\mathrm{d}}(r_{\mathrm{f}}) + M_{\mathrm{b}}(r_{\mathrm{f}}) + (1 - m_{\mathrm{d}} - m_{\mathrm{b}})M(r_{\mathrm{i}}), \tag{2.21}$$

and the dark mass profile of equation (2.18) now becomes

$$M_{\mathrm{h}}(r) = M_{\mathrm{f}}(r) - M_{\mathrm{d}}(r) - M_{\mathrm{b}}(r). \tag{2.22}$$

2.3 N-body realizations of model galaxies

2.3.1 Introduction

In order to construct near-equilibrium N-body realizations of our disk models, we need to initialize both positions and velocities of particles according to the solution of the collisionless Boltzmann equation (CBE)[1]. While the first can easily be done according to the derived mass distributions for halo, bulge, and disk, the latter is considerably more complicated.

Instead of attempting to solve the CBE directly, we follow Hernquist (1993a) and assume that the velocity distribution at a given point in space can be sufficiently well approximated by a multivariate Gaussian. In this case, only the first two moments of the velocity distribution are needed. They can be obtained by taking moments of the CBE, a process that leads to a hierarchy of generalized Jeans equations (Magorrian & Binney 1994).

For a static, axisymmetric system, the energy E and the angular momentum component L_z are conserved along orbits. With the assumption that the distribution function depends only on E and L_z one can show (Magorrian & Binney 1994) that the first velocity moments are given by

$$\overline{v_R} = \overline{v_z} = \overline{v_R v_z} = \overline{v_z v_\phi} = \overline{v_R v_\phi} = 0, \tag{2.23}$$

$$\overline{v_R^2} = \overline{v_z^2}, \tag{2.24}$$

$$\overline{v_z^2} = \frac{1}{\rho}\int_z^\infty \mathrm{d}z'\,\rho(R,z')\frac{\partial\Phi}{\partial z'}(R,z'), \tag{2.25}$$

$$\overline{v_\phi^2} = \overline{v_R^2} + \frac{R}{\rho}\frac{\partial}{\partial R}(\rho\overline{v_R^2}) + v_{\mathrm{c}}^2, \tag{2.26}$$

where the *azimuthal circular velocity* is defined as

$$v_{\mathrm{c}}^2 \equiv R\frac{\partial\Phi}{\partial R}. \tag{2.27}$$

Not specified by the Jeans equations is the azimuthal streaming $\overline{v_\phi}$, which can essentially be freely chosen in the context of the above approximations. This reflects the fact that

[1] also known as Vlasov equation.

the distribution function is even in L_z; the relative contribution of the parts with positive and negative L_z can be arbitrarily chosen.

We employ the assumption $f = f(E, L_z)$ for the dark matter halo and the optional bulge. However, a realistic distribution function for the disk has a more complicated structure, and we will treat it slightly differently, as described below.

2.3.2 Structure of the disk

Real stellar disks have a finite thickness. For their vertical structure we adopt the common choice of Spitzer's isothermal sheet, viz.

$$\rho_{\rm d}(R,z) = \frac{\Sigma(R)}{2z_0}\mathrm{sech}^2\left(\frac{z}{z_0}\right). \tag{2.28}$$

Here the thickness z_0 of the disk sets its 'temperature'. Most spiral galaxies seem to be consistent with a constant vertical scale length with a value of $z_0 \simeq 0.2R_{\rm d}$, which we will adopt in the following.

The distribution function of the disk depends on more than just two conserved quantities, hence it is unrealistic to assume an isotropic velocity dispersion. However, we will keep the assumption that the velocity ellipsoid is aligned with the coordinate axes. Then equation (2.25) remains valid, and we use it to compute $\sigma_z^2(R,z)$. Note that due to the radial variation of $\Sigma(R)$ and the presence of the halo, the vertical velocity structure of the disk will not be exactly isothermal.

We further employ the epicycle approximation (Binney & Tremaine 1987, chapter 4) to relate the radial and azimuthal velocity dispersions by

$$\sigma_\phi^2 = \frac{\sigma_R^2}{\gamma^2}. \tag{2.29}$$

Here we have defined

$$\gamma^2 \equiv \frac{4}{\kappa^2 R}\frac{\partial\Phi}{\partial R}, \tag{2.30}$$

and the epicyclic frequency κ as

$$\kappa^2 \equiv \frac{3}{R}\frac{\partial\Phi}{\partial R} + \frac{\partial^2\Phi}{\partial R^2}. \tag{2.31}$$

The epicycle approximation also implies that the asymmetric drift is small. We neglect it altogether and set the streaming velocity to be equal to the circular velocity, i.e. $\overline{v_\phi} = v_{\rm c}$. For simplicity, we also continue to assume that $\sigma_R^2 = \sigma_z^2$. With these assumptions the velocity structure of the disk is fully specified.

Note that the thickness of the disk must be chosen large enough to fulfil Toomre's stability criterion, which requires

$$Q \equiv \frac{\sigma_R\kappa}{3.36G\Sigma} > 1 \tag{2.32}$$

to ensure local stability in differentially rotating disks.

For our models, the minimum value of Q is about 1.4. Hence, we could have taken somewhat colder disks, which might produce sharper tidal tails.

2.3.3 Rotation of the halo

For consistency, we want to properly represent the angular momentum carried by the dark matter, although we do not expect that halo rotation will have a strong influence on tail formation.

We model the streaming velocity of the dark halo as some fixed fraction $f_{\rm s}$ of the local *azimuthal* circular velocity, i.e.

$$\overline{v_\phi} = f_{\rm s} v_{\rm c}. \tag{2.33}$$

If the specific angular momentum of the dark matter is conserved during disk formation, the factor $f_{\rm s}$ stays fixed as well. Hence it can be computed for the initial NFW-profile. For the streaming of equation (2.33) the initial angular momentum is

$$J = \frac{2}{3} f_{\rm s} g_c \frac{G^{\frac{1}{2}} M^{\frac{3}{2}} r_{\rm s}^{\frac{1}{2}}}{\left[\ln(1+c) - \frac{c}{1+c}\right]^{\frac{3}{2}}}, \tag{2.34}$$

where g_c is the integral

$$g_c = \int_0^c \left[\ln(1+x) - \frac{x}{1+x}\right]^{\frac{1}{2}} \frac{x^{\frac{3}{2}}}{(1+x)^2} {\rm d}x. \tag{2.35}$$

Comparing this with equation (2.11) we see that $f_{\rm s}$ is given by

$$f_{\rm s} = \frac{3}{2}\lambda \left(\frac{2c}{f_c}\right)^{\frac{1}{2}} g_c^{-1} \left[\ln(1+c) - \frac{c}{1+c}\right]^{\frac{3}{2}}. \tag{2.36}$$

The quantity $f_{\rm s}/\lambda$ varies only weakly, e.g. it takes the value 4.5 for $c = 1$, and reduces to 3.3 for $c = 100$.

2.3.4 Halo truncation

For the halo we encounter a slight technical problem, since the cumulative mass distribution of the NFW profile actually diverges for large r. This is simply due to the fact that the NFW-profile in the form of equation (2.1) is not valid out to arbitrarily large distances; it just provides a good fit to the profile up to about the virial radius.

Instead of truncating the profile sharply at the virial radius we rather want to derive N-body models where the density fades out smoothly. For simplicity, we have chosen an exponential cut-off that sets in at the virial radius and turns off the profile on a scale $r_{\rm s}$, viz.

$$\rho(r) = \frac{\rho_{\rm crit.}\delta_{\rm c}}{c(1+c)^2} \left(\frac{r}{r_{200}}\right)^a \exp\left(-\frac{r - r_{200}}{r_{\rm s}}\right) \tag{2.37}$$

for $r > r_{200}$. The power law exponent a allows a smooth transition of the profile at r_{200}. We select a such that the logarithmic slope

$$n = r \frac{\rm d}{{\rm d}r} \ln \rho \bigg|_{r_{200}} = -\frac{1+3c}{1+c} \tag{2.38}$$

of the profile at the virial radius is continuous. This implies $a = c + n$.

Note that this truncation results in some additional halo mass beyond the virial radius. The total mass $M_{\rm tot}$ is roughly 10% larger than M_{200}. However, we want to keep our definitions of disk and bulge masses in terms of the virial mass. As a consequence we need to slightly modify equation (2.17). It becomes

$$M_{\rm f}(r_{\rm f}) = M_{\rm d}(r_{\rm f}) + M_{\rm b}(r_{\rm f}) + \left[1 - \frac{M_{200}}{M_{\rm tot}}(m_{\rm d} + m_{\rm b})\right] M(r_{\rm i}).$$

2.3.5 Numerical procedure

Finally, we briefly describe our computer code to set up a galaxy according to the above model. The following steps are followed:

1. Particle positions are initialized according to the density profiles for halo, disk, and bulge.

2. We then compute the velocity dispersions on a fine logarithmic mesh in the (R, z)-plane. For this purpose, we compute the integrals

$$\overline{v_z^2} = \frac{1}{\rho}\int_z^\infty {\rm d}z'\, \rho \frac{\partial \Phi}{\partial z'} \tag{2.39}$$

numerically at the grid points. This also determines $\overline{v_R^2} = \overline{v_z^2}$. Note that the density refers only to the component under consideration, while the potential is given by all the material. For the halo and the bulge, we use equation (2.26) and find $\overline{v_\phi^2}$ by numerically differentiating $\rho\overline{v_R^2}$ in this plane. For the disk, we use the epicycle approximation to determine $\overline{\sigma_\phi^2}$, and we set the azimuthal streaming equal to the local circular velocity. The streaming of the halo is given by equation (2.33), while that of the bulge (if present) is set to zero.

   It should be noted that the integrals of equation (2.39) require elaborate numerical techniques, since the computation of the combined force field is nontrivial.

3. Finally, particle velocities are initialized by drawing random numbers from multivariate Gaussians with dispersions interpolated from the (R, z)-grid to the particle positions.

This scheme is similar to that of Hernquist (1993a), although our numerical procedure and treatment of the disk is somewhat different.

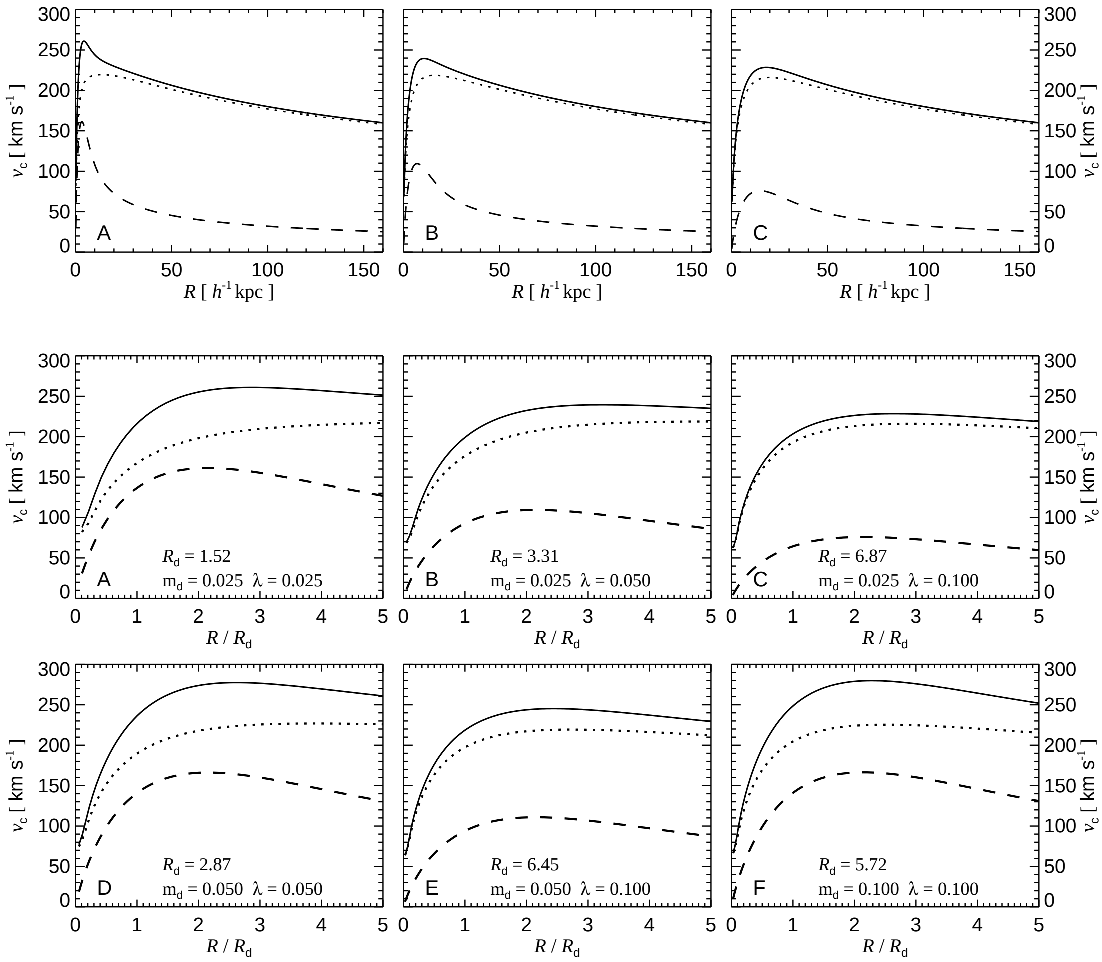


Figure 2.1: Rotation curves of disk models A to F. The top three panels show the full rotation curve of models A-C out to the virial radius. In the other six panels, the radial coordinate is normalized to the disk scale length, and we plot the inner rotation curve out to 5 disk scale lengths. In each panel, the dotted curve gives the contribution of the dark matter, the dashed that of the disk, and the solid line is the total rotation curve. In all the models, the total mass of the galaxies is equal, and corresponds to an initial NFW dark matter profile with $v_{200} = 160\,\mathrm{km\ s^{-1}}$ and $c = 15$.

2.4 Simulations

2.4.1 Models

2.4.1.1 The basic disk models

We have constructed a basic set of six disk models with a constant total mass corresponding to $V_{200} = 160\,\mathrm{km\ s^{-1}}$ and a concentration of $c = 15$. These models are those

Table 2.1: Parameters of our basic set of six disk models. All of them have the same total mass corresponding to $V_{200} = 160\,\mathrm{km\,s^{-1}}$, and an initial halo concentration of $c = 15$. From the possible combinations of $\lambda, m_\mathrm{d} \in \{0.025, 0.5, 0.1\}$ we consider only those models that are stable according to the criterion $\lambda \geq m_\mathrm{d}$. We assume that the disk material conserves its specific angular momentum, i.e. $j_\mathrm{d} = m_\mathrm{d}$. The two tables illustrate our labeling of the models and the resulting disk scale lengths. Note that these models have no bulge.

		λ		
		0.025	0.05	0.1
	0.025	A	B	C
m_d	0.05		D	E
	0.1			F

Model	A	B	C	D	E	F
R_d [h^{-1}kpc]	1.52	3.31	6.87	2.87	6.45	5.72

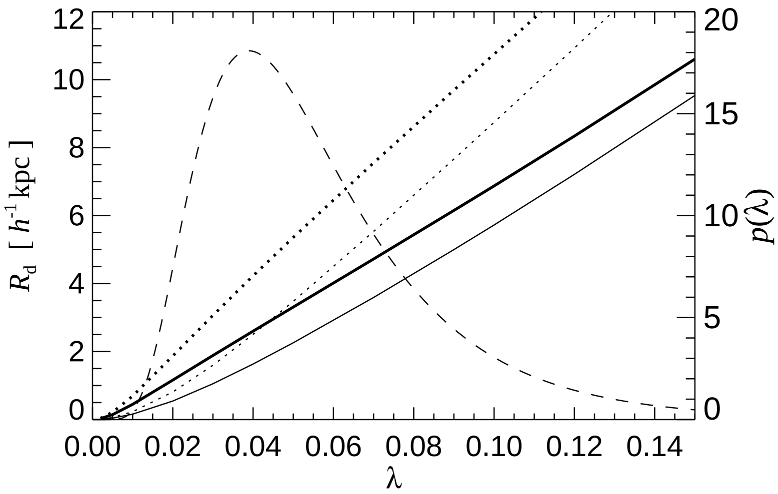


Figure 2.2: Sizes of disks. The two solid lines show the scale length R_d of disks as a function of the spin parameter λ. The solid lines are for haloes with $c = 15$, while the dotted curves refer to $c = 5$. In both cases, the heavy curves are for $m_\mathrm{d} = 0.025$, and the thin ones for $m_\mathrm{d} = 0.1$. The dashed line shows the distribution $p(\lambda)$ of λ expected in CDM cosmologies.

Table 2.2: List of runs. The table gives the orbital angular momentum of the different runs in terms of the minimum Keplerian separation $R_{\rm kep}$. All the runs had an initial galaxy separation of $R_{\rm start} = 320\,h^{-1}$kpc, and were set-up on a parabolic encounter with zero total energy. Each of the runs is a collision between identical disk models. The latter is specified by the initial character of the labels. The middle table gives the actual separation $R_{\rm min}$ of the disks in their first encounter, while the bottom table lists the total spin parameters $\lambda_{\rm tot}$ of the colliding binary systems.

$R_{\rm kep}$ $[h^{-1}$kpc]					
3.5	7.0	14.0	28.0	56.0	112.0
A0	A1	A2			
B0	B1	B2			
C0	C1	C2	C3	C4	C5
		C2r			
		C2i			
	D1				
		E2			
		F2			
	T1				
	U1				
	V1				
	W1				

$R_{\rm min}$ $[h^{-1}$kpc]					
A0	6.4	B2	23.7	D1	10.8
B0	8.3	C2	22.4	E2	18.2
C0	7.4	C2r	19.4	F2	22.4
A1	10.5	C2i	21.2	W1	10.9
B1	14.0	C3	37.4	T1	8.3
C1	13.6	C4	58.2	U1	9.4
A2	21.3	C5	109.6	V1	10.8

$\lambda_{\rm tot}$					
A0	0.035	B2	0.069	D1	0.056
B0	0.047	C2	0.094	E2	0.094
C0	0.072	C2r	0.006	F2	0.094
A1	0.044	C2i	0.067	W1	0.056
B1	0.056	C3	0.113	T1	0.056
C1	0.081	C4	0.139	U1	0.071
A2	0.057	C5	0.176	V1	0.051

combinations of $\lambda, m_{\rm d} \in \{0.025, 0.5, 0.1\}$ that result in stable cold disks, i.e. that have $\lambda \geq m_{\rm d}$ (the stability criterion is discussed in more detail below). We label these models A to F, as outlined in Table 2.1. Also given in this Table are the resulting disk scale lengths. Note that we always assume $j_{\rm d} = m_{\rm d}$, i.e. the specific angular momentum content of the disk material is exactly conserved during disk formation. In contrast to this, gas-dynamical simulations of disk formation (Navarro & White 1994; Navarro & Steinmetz 1997; Weil et al. 1998) have typically led to a loss of angular momentum from the gas to the halo. As a result, the disks formed in these simulations were much too small to be identified with real spiral galaxies. However, as all of the above authors note, this angular momentum problem may well be due to an insufficient treatment of feedback processes.

Note that in the model of MMW, the structure of the disk galaxies depends only on $\lambda' = (j_\mathrm{d}/m_\mathrm{d})\lambda$. Hence, angular momentum loss from the disk ($j_\mathrm{d} < m_\mathrm{d}$) has the same effect as lowering the value of λ.

2.4.1.2 Rotation curves

In Figure 2.1 we show the rotation curves of our six primary models described in Table 2.1. For each model, we give the inner rotation curve out to 5 disk scale lengths, which is about the accessible regime in most disk galaxies. For models A to C, we also show the full rotation curve out to the virial radius of $160\,h^{-1}$kpc.

Several interesting trends may be observed. In the models A, B, and C, only the spin parameter is increased. This leads to larger disks with roughly $R_\mathrm{d} \propto \lambda$. The dependence of R_d on the spin parameter λ is shown in Figure 2.2. However, the smaller disks pull in the dark matter more strongly, leading to a larger concentration of the dark matter for smaller disks. This effect reduces the differences between the rotation curves when their radial coordinate is normalized to the disk scale length.

On the other hand, for very small disk mass, the dark matter profile will be nearly unaffected by the disk formation. In this limit, the disk stars behave more or less like test particles in the dark matter potential, yet the size of the disk is still determined by the halo spin parameter. Also, in this limiting case of a massless disk it is quite clear that the mass ratio between halo and disk must be irrelevant for the formation of stellar tidal tails.

2.4.1.3 A model with a massive bulge

We also consider a model with a massive bulge, designed to have a similar rotation curve to DMH's models, and hence being more directly comparable to them than our standard models. We also adopt their relatively high disk-to-bulge mass ratio of 2:1. In detail, our parameters for this model, which we call 'W', are $c = 15$, $v_{200} = 160\,\mathrm{km\ s^{-1}}$, $\lambda = 0.05$, $m_\mathrm{d} = 2/3 \times 0.05$, $m_\mathrm{b} = 1/3 \times 0.05$, and $f_\mathrm{b} = 0.1$. This results in the rotation curves shown in Figure 2.3. While the inner rotation curve is very similar to that of DMH, the contribution of the dark matter is much more important in our model, even at small radii. In particular, it is always much larger than the contribution of the disk.

2.4.1.4 The amount of dark mass in the disks

As the rotation curves in Fig. 2.1 show, all our 'basic' disk models A-F are gravitationally dominated by dark matter, even in the innermost regions of the disks. While the presence of a dark matter halo has been convincingly demonstrated by the flatness of observed

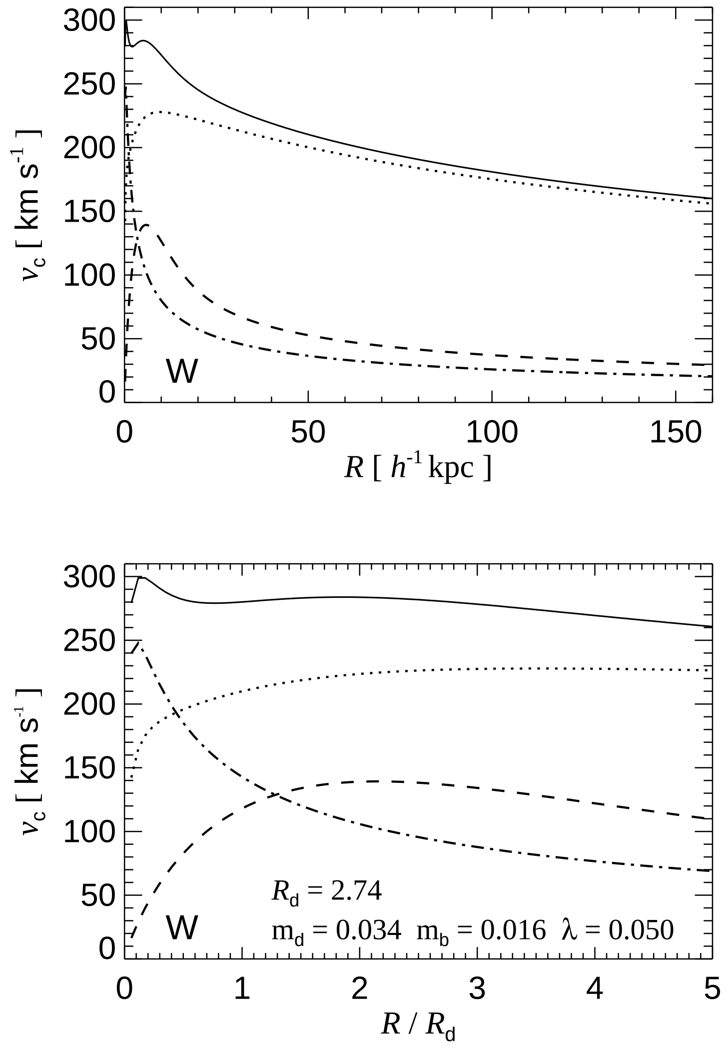


Figure 2.3: Rotation curves of the bulge model W. The top panel shows the full rotation curve, and the lower panel displays the inner rotation curve out to 5 disk scale lengths. The dashed line is the contribution of the disk, the dot-dashed that of the bulge, the dotted that of the dark halo, and the solid line gives the total curve.

rotation curves, there is still a controversy about the amount of dark matter in the inner regions of disk galaxies.

This controversy has arisen, because the decomposition of an observed rotation curve into a stellar and a dark matter component is rather ambiguous, since the result depends strongly on assumptions about the dark matter profile and the mass-to-light

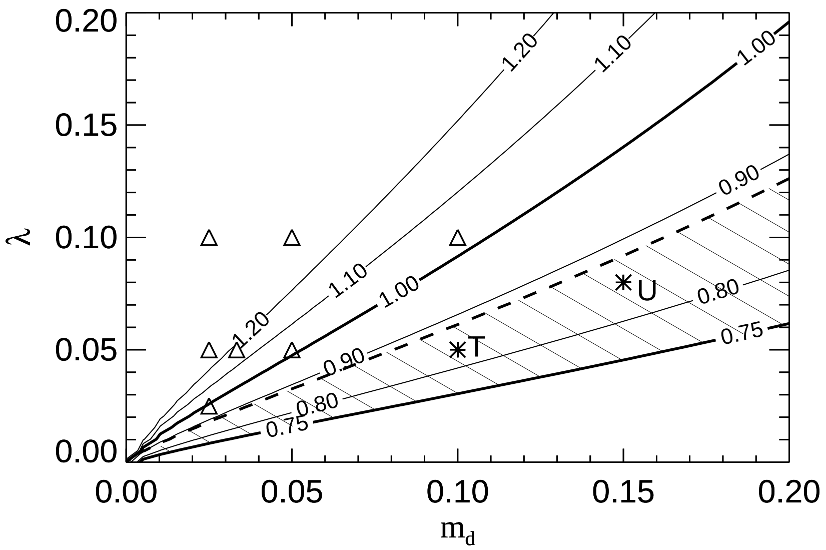


Figure 2.4: Contours of the disk stability parameter $\epsilon_{\rm m}$ (solid lines) for $c = 15$. According to Syer, Mao & Mo Syer et al. (1997) disks with $\epsilon_{\rm m} \geq 0.75$ should be stable against bar formation, while the earlier work of Efstathiou, Lake & Negroponte Efstathiou et al. (1982) gives the condition $\epsilon_{\rm m} \geq 1.1$. Disk models lying below the thick dashed line are dominated by the disk gravity at the maximum of the disk rotation curve, thus the hatched region shows the parameter space that may contain stable disks which are not everywhere dominated by dark matter. Two of our models, 'T' and 'U' (stars), lie in this region. The other models are indicated as triangles.

ratio (Navarro 1998). Traditionally, rotation curves have therefore been fitted using the 'maximum-disk' hypothesis, i.e. the largest mass possible is assigned to the disk consistent with the rotation curve. The recent work of Debattista & Sellwood (1998) suggests that the central density of dark matter in barred galaxies should be low, thus supporting the maximum-disk hypothesis. However, others (e.g. van der Kruit 1995) maintain that the contribution of dark matter in the inner regions of disk galaxies must be substantial. There is also recent observational work that supports this assertion (Quillen & Sarajedini 1998). We also note that even if the inner rotation curve of a galaxy can be well accounted for by a disk component alone, this does not provide evidence for the absence of dark matter in the inner parts of the disk.

The theoretical results employed in this work predict dark matter profiles with a central density cusp, and a resulting strong contribution of dark matter in the inner disks. We now briefly discuss, to what extent the model of MMW may also accommodate galaxies that are maximum-disk, or at least somewhat closer to it. In order to make the self-gravity of the disk more important, we can either reduce λ (making the disk

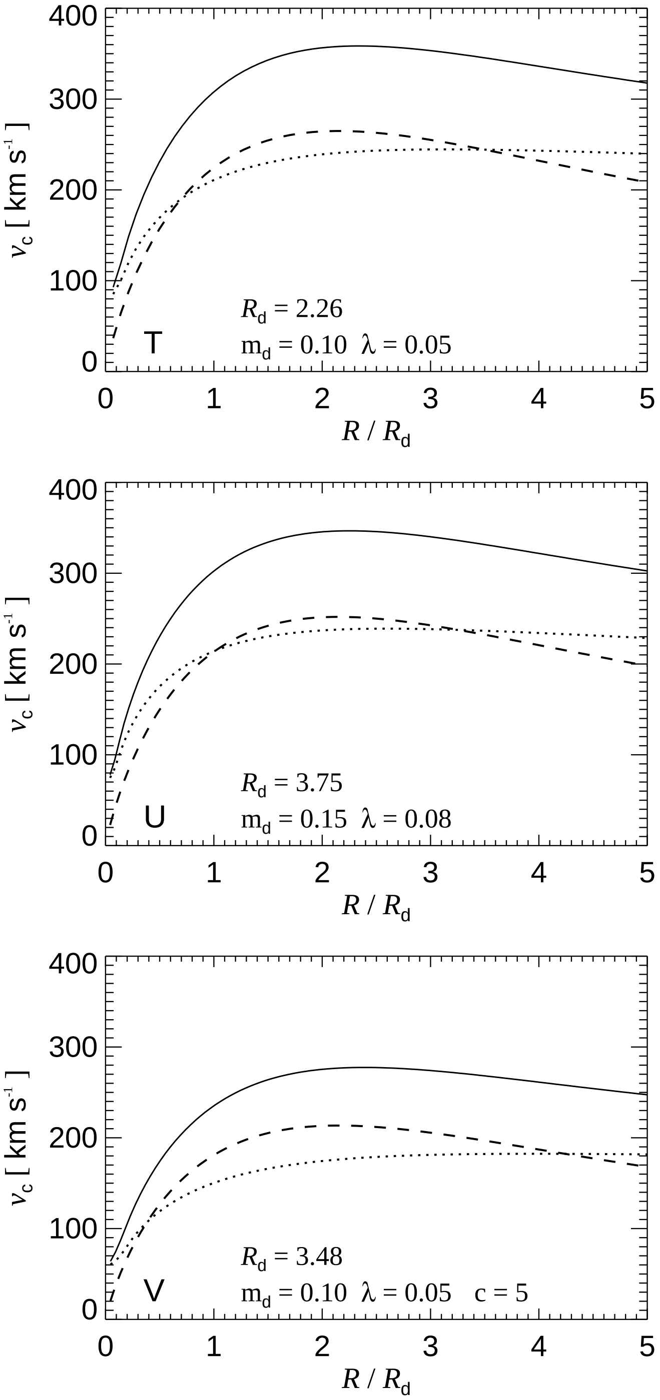


Figure 2.5: Inner rotation curves of the models T, U, and V out to 5 disk scale lengths. The dashed line is the contribution of the disk, the dotted that of the dark halo, and the solid line gives the total rotation curve.

smaller), increase $m_{\rm d}$ (making the disk heavier), or lower c (reducing the concentration of the halo). However, for fixed $m_{\rm d}$ and fixed c, disk-stability poses a lower limit on λ. Thin, fully self-gravitating disks have been known to be violently bar-unstable for a long time, a fact that suggested to Ostriker & Peebles (1973) that there must be dark matter that stabilizes the disks. Later, Efstathiou, Lake & Negroponte (1982) used N-body simulations to derive the stability criterion $\epsilon_{\rm m} \geq 1.1$ for the disk, where

$$\epsilon_{\rm m} \equiv \frac{v_{\rm max}}{(GM_{\rm d}/R_{\rm d})^{1/2}}, \tag{2.40}$$

and $v_{\rm max}$ is the maximum rotation velocity. Recently, Syer, Mao & Mo (1997) confirmed that $\epsilon_{\rm m}$ is a good diagnostic for bar-instability, although they found a somewhat weaker stability criterion, $\epsilon_{\rm m} \geq 0.75$.

In Figure 2.4 we show contours of $\epsilon_{\rm m}$ in the $m_{\rm d}$-λ plane. Also shown is the region, where the disk gravity at the maximum of the disk rotation curve is larger than the contribution by the dark matter. There is thus a small region of parameter space (hatched) where the galaxies are disk-dominated, but where they should be still stable against bar formation according to Syer et al. (1997). Incidentally, $\epsilon_{\rm m} \geq 1.0$ corresponds very closely to the condition $\lambda \geq m_{\rm d}$, the choice we employed so far.

Note that lowering λ may not only render a disk unstable, it will also make it substantially smaller. Hence, the observed sizes of disk galaxies can also provide a lower bound on λ. Indeed, the results of MMW suggest that disks become too small if they lose a substantial part of their angular momentum to the dark halo. Since these arguments disfavour low λ, one may rather try to increase $m_{\rm d}$ to make the gravity of the disk more important. In principle, we expect that the universal cosmic baryon fraction poses an upper limit on $m_{\rm d}$, while the actual value of $m_{\rm d}$ could be a lot smaller if the efficiency of disk formation is low. Taking a big bang nucleosynthesis value of $\Omega_{\rm B} \simeq 0.015h^{-2}$ (Copi et al. 1995) for the baryon density, $m_{\rm d}$ should be smaller than 0.06 in a critical density universe with a Hubble constant of $h = 0.5$. However, clusters of galaxies suggest that the baryon fraction is larger by at least a factor of three (White et al. 1993). Note that in a low density universe this can be reconciled with cosmic nucleosynthesis. If Ω_0 is as low as 0.2, the limit on $m_{\rm d}$ goes up to about 0.15-0.2.

We examine these possibilities to a limited extent with three additional models which are disk-dominated in the inner regions. It is interesting to see how their tidal tails fit into the systematic properties of the other models. We label one of these models 'T', and give it the parameters $m_{\rm d} = 0.1$, $\lambda = 0.05$, $c = 15$, $v_{200} = 160\,{\rm km\,s^{-1}}$, and $m_{\rm b} = 0$. For a further model, called 'U', we instead adopt $m_{\rm d} = 0.15$, $\lambda = 0.08$, i.e. here we make the disk substantially more massive. Finally, we consider a model 'V' with a smaller concentration of the halo. Here we use $c = 5$, $m_{\rm d} = 0.1$, and $\lambda = 0.05$. Hence this model is consistent with the value of $c \simeq 5$ favoured by Navarro (1998) in a recent analysis of a sample of spiral galaxies. Note that such low concentrations are theoretically expected for flat, low-density universes.

The inner rotation curves of these models are shown in Fig. 2.5. The stability parameter for them is $\epsilon_{\rm m} \simeq 0.84$. Hence they lie in the hatched region of Figure 2.4. In contrast to the other models, we here chose $\sigma_R = 2.0\,\sigma_z$ for the velocity structure of the disk to prevent it developing a bar before the galaxies collide. This raises Q to about 2.0, while it would have been $Q \simeq 1.0$ for our conventional choice for σ_R.

2.4.2 Collision simulations

The most favourable condition for making tidal tails are prograde encounters where the spin vectors of the disks are aligned with the orbital angular momentum. In this situation, the approximate resonance between the disk rotation and the orbital angular frequency amplifies the perturbation of particle orbits on the far sides of the disks, since they stay for a longer time in the region of the strongest tidal field.

Since we here try to achieve as prominent tails as possible, we usually set up our disk-disk collisions on prograde parabolic orbits. For simplicity we run only symmetric encounters between pairs of identical models; that is we collide model A with A, B with B, and so on.

We always chose the initial separation of the galaxies to be twice the virial radius. i.e. $R_{\rm start} = 320\,h^{-1}$kpc. The remaining undetermined orbital parameter is the orbital angular momentum. We specify it in terms of the minimum separation $R_{\rm kep}$ the galaxies would reach if they were point masses moving on the corresponding Keplerian orbit. In reality, once the galaxies overlap they will start to deviate from this trajectory due to dynamical friction. As a consequence, the actual separation $R_{\rm min}$ of the galaxies in their first encounter will generally be larger than $R_{\rm kep}$.

Note that most of the encounter velocity is generated by the potential gradient inside the virial radii of the two galaxies, and the tails themselves also move primarily within this region. Hence it seems unlikely that material 'outside' the haloes strongly affects tail formation. In principle, such material will be present in simulations that also follow the cosmological context of the collision.

We examined three main choices for $R_{\rm kep}$, 3.5, 7 and $14\,h^{-1}$kpc. For each of the models A, B, and C, we have run all three of these combinations, while we restricted ourselves to just one 'impact parameter' for the other models. Note that these orbital parameters are consistent with typical energies and eccentricities found in cosmological N-body simulations (Navarro et al. 1995; Tormen 1997). Additionally, we simulated a set of wider encounters with $R_{\rm kep} = 28$, 56, $112\,h^{-1}$kpc for the C model. Runs labeled 'A0', 'B0', etc. refer to $R_{\rm kep} = 3.5\,h^{-1}$kpc, those containing the digits 1 or 2 to $7\,h^{-1}$kpc and $14\,h^{-1}$kpc, respectively.

We also simulated two additional versions of run C2 where the disks do not have a prograde orientation. In the collision 'C2r' both disks are retrograde, i.e. their spins are just flipped, while in the model 'C2i' they are inclined by $90^{\rm o}$ relative to the orbital plane.

With respect to the bulge model and the disk-dominated models, we have only run one simulation in each case ('W1', 'T1', 'U1', 'V1'), an encounter at $R_{\rm kep} = 7\,h^{-1}$kpc. Table 2.2 gives an overview of all these runs. In this table, we also list the total spin parameter $\lambda_{\rm tot}$ for the colliding binary systems. Apart from the extreme collisions C4 and C5, these values are well consistent with the log-normal distribution (2.8) for λ.

Also shown in Table 2.2 are the actual minimum separations $R_{\rm min}$ of the centers of the disks in their first encounter. We here defined the center of a disk as its densest point, and used a kernel interpolation like in smoothed particle hydrodynamics (SPH) to estimate the density of particles. However, using simply the center-of-mass of the individual disks gives similar results.

When the galaxies start to overlap, the interaction potential between them becomes shallower than that of the corresponding point masses. This effect will make the orbits wider than the Keplerian expectation. However, the galaxies are also slowed down by dynamical friction, an effect that brings the galaxies closer together. Both mechanisms compete with each other, and their relative strength depends on the distribution of mass inside the galaxy. As Table 2.2 shows, the minimum separations $R_{\rm min}$ are usually somewhat larger than the corresponding Keplerian value $R_{\rm kep}$. Note however, that the measurement of $R_{\rm min}$ has an uncertainty of order $\pm 1\,h^{-1}$kpc, because we stored only a limited number of output times.

2.4.3 Numerical techniques

All the simulations in this work have been run with our GADGET-code (**GA**laxies with **D**ark matter and **G**as int**E**rac**T**). It is a newly written N-body/SPH-code in C, specifically designed for the simulation of galaxy formation and interaction problems. The gravitational interaction is either computed with the special-purpose hardware GRAPE (if available) or with a TREE code. In this work, the SPH part of GADGET is not used; we just treat dark matter and stellar material as collisionless particles. Futher details of GADGET will be described in Chapter 5.

For all of the basic models A to F, and for T to V, we used 20000 particles to represent each disk, and 30000 particles for each halo, hence each simulation had a total of 100000 particles. We chose a gravitational softening length of $0.4\,h^{-1}$kpc for the dark matter, and $0.1\,h^{-1}$kpc for the disk. Time integration was performed with high enough accuracy, such that the total energy was conserved to better than 0.8% in all runs.

For the bulge model W, we used an additional 10000 particles for each bulge, and we employed a softening of $0.1\,h^{-1}$kpc for the bulge as well.

Some of the models have been integrated using GRAPE (A0, B0, B2, E2, F2, T1, U1, W1), the others with the TREE code. For the latter we used the cell opening criterion of Dubinski (1996) with $\theta = 1.0$, we included quadrupole moments, and we matched the spline softening of the TREE code to the Plummer softening of GRAPE cited above.

Each simulation was run for 2.6 internal time units, or 0.26 Hubble times, corresponding to $2.54 \times 10^9\,h^{-1}$yr. At this point of time, the merger remnants are not yet fully relaxed, but the tidal tails have already largely decayed.

2.5 Results

2.5.1 Dynamical evolution of the models

Figure 2.6 shows a representative example of the time evolution of one of our runs (B1). Overall, the models follow the well-known behaviour of close encounters of pairs of disk galaxies. When the galaxies reach orbital centre, violent tidal forces induce a bar instability that quickly transforms the disks into a pair of open bisymmetric spirals. Simultaneously, disk material from the far side of the encounter is ejected by the tidal field into arcing trajectories that later form tidal arms. Material from the near side is drawn towards the companion, giving rise to bridges between the galaxies as they temporarily separate again. While the bridges are destroyed when the galaxies come back together for a second time, the tails can survive and grow for a longer time in the relatively quiet regions of the outer potential.

Nevertheless, the dynamical evolution of the tidal tails is quite rapid. After their initial phase of expansion, the most strongly bound material in the inner region of the tail quickly starts to rain back onto the merging pair. Eventually, this also happens to material progressively further out, such that the surface density and prominence of the tidal tails quickly decrease with time.

2.5.2 Comparison of tidal tails

Depending on the strength of the tidal response, the tidal tails can contain a varying amount of mass, and reach different lengths. By comparing the time evolution of the models A to C, such differences are readily apparent. For example, the tails of the larger disks of run C1 are much more massive and prominent than those of run B1, while the tails of the small disks of simulation A1 are rather thin and anemic. However, the spatial extent of the tails is quite comparable in the models. When normalized to the initial disk scale length the anemic tails of the A-models are even longer than those of the C-models.

These trends are clearly visible in Figures 2.7 and 2.8, where we compare different runs at the same time, approximately corresponding to the moment when the tails are most impressive. Note however, that due to the rapid evolution of the morphology of the tidal tails it is not easy to compare different models at exactly equivalent times of dynamical evolution.

When models with different impact parameters are compared, some finer trends in the tail morphology may be observed. With growing impact parameter (labels $0 \rightarrow 2$), the bridges between the galaxies become more pronounced, and the tails are slightly more curved. This is related to the larger orbital angular momentum of these encounters.

Note that the runs A, B, and C of Figure 2.7 all have a large halo-to-disk mass ratio of more than 40:1. Nevertheless, the spatial extent of the tails is often quite large. For example, the B-models with a disk scale length of $R_\mathrm{d} \approx 3.5\,h^{-1}$kpc produce tails reaching $200\,h^{-1}$kpc in length, i.e. about 60 times the original disk scale length.

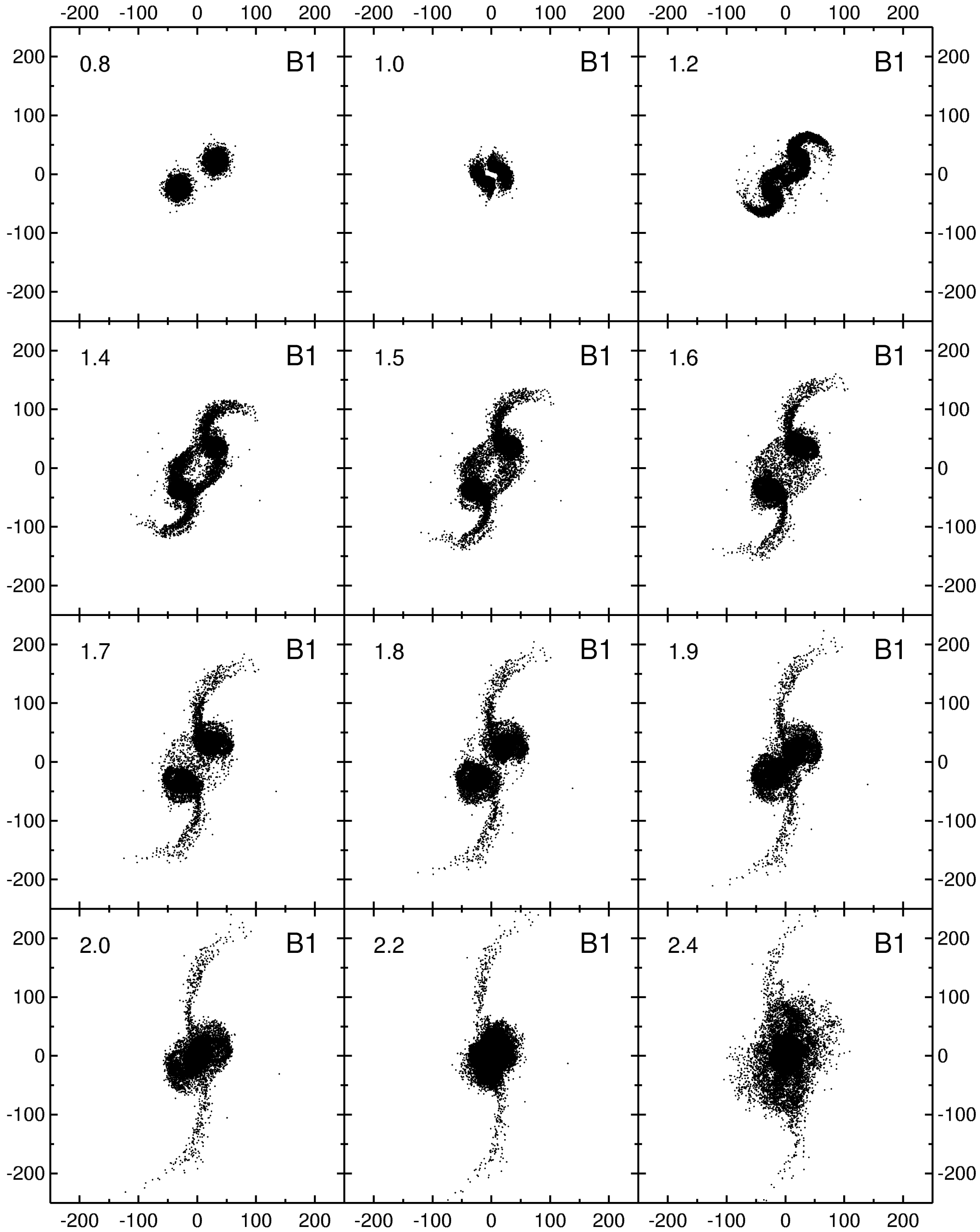


Figure 2.6: Time evolution of run B1. The panels show the disk particles projected onto the orbital plane. The length units labeling the axes are given in h^{-1}kpc. The elapsed time since the start of the simulation (upper left corners) is given in units of 0.1 Hubble times, or $9.8 \times 10^8\, h^{-1}$yr.

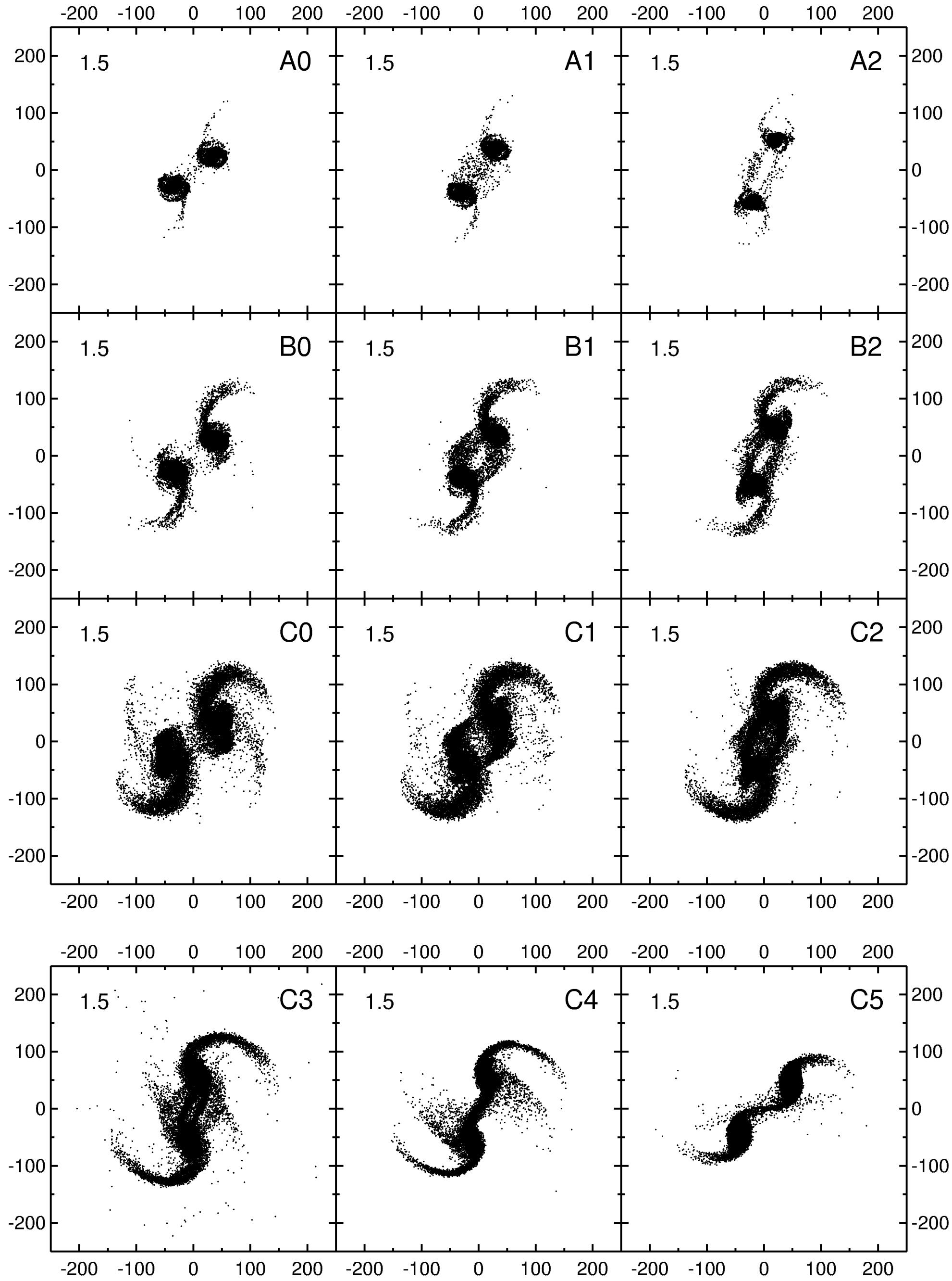


Figure 2.7: Comparison of tidal tails between different runs. All the panels show the disk particles projected onto the orbital plane at the same time since the start of the simulation. The top 9 panels display runs that involve the disk models A-C for the set of impact parameters $R_{\rm kep} = 3.5$, 7, and $14\,h^{-1}$kpc (indicated by the digits 0, 1, and 2, respectively, in the labels of the runs). From A to C, the spin parameter λ increases in the sequence 0.025, 0.05, and 0.1, but all three models have an equal disk mass given by $m_{\rm d} = 0.025$. In the lower three panels we show additional collisions of model C with wider impact parameters in the sequence $R_{\rm kep} = 28$, 56, and $112\,h^{-1}$kpc.

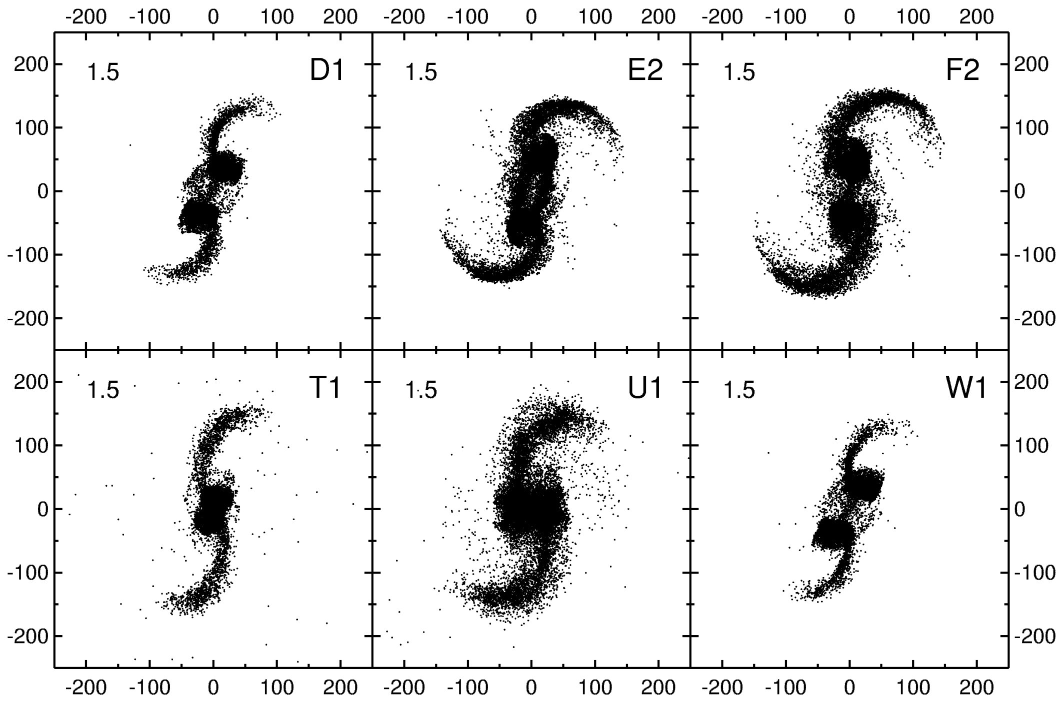


Figure 2.8: Comparison of tidal tails between different runs. The panels show the disk particles projected onto the orbital plane at the same time since the start of the simulation. Runs D1 and E2 involve models with twice as heavy disks ($m_\mathrm{d} = 0.05$) than A-C, while F2 has $m_\mathrm{d} = 0.1$. The D-model has a spin parameter of $\lambda = 0.05$, while the E- and F-disks have $\lambda = 0.1$. The T ($m_\mathrm{d} = 0.1$, $\lambda = 0.05$) and U ($m_\mathrm{d} = 0.15$, $\lambda = 0.08$) models are disk-dominated in their inner regions, while model W ($m_\mathrm{d} = 0.034$, $m_\mathrm{b} = 0.016$ $\lambda = 0.05$) also contains a bulge. The digit in the label of the runs parameterizes the impact parameter; 0, 1, and 2 are for $R_\mathrm{kep} = 3.5$, 7, or $14\,h^{-1}$kpc, respectively.

When tails of models with different halo-to-disk mass ratio but equal spin parameter are compared, it becomes clear that the mass ratio is not the relevant parameter that decides whether tails form or not. For example, the tails of runs E2 and F2 in Figure 2.7 may be compared to the ones of run C2. Despite a variation of the mass ratio by a factor 4 or so, the tails are almost equally strong in these three simulations. This is clearly due to the approximately equal size of the disks in these models. Because the dark matter is gravitationally dominant even in the regions of the disks, the disk stars behave almost like test particles in the gravitational potential of the dark halo. In this limiting case it is clear, that only the *location* of the disk material inside the dark halo determines the disk response, i.e. it is the relative size of disk and dark halo that matters.

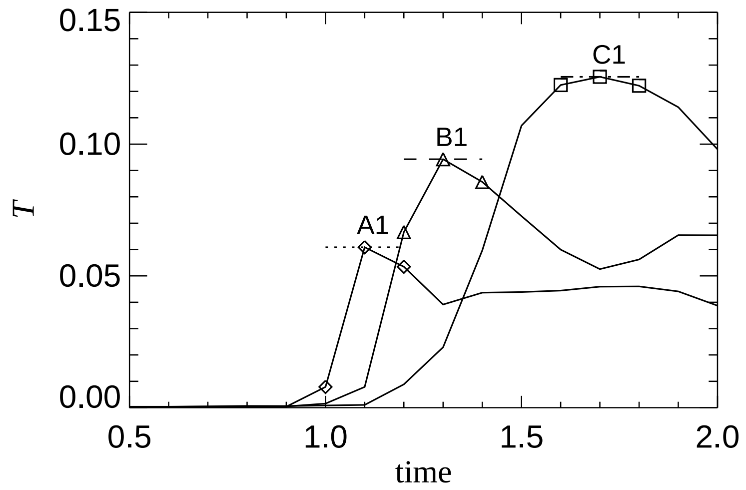


Figure 2.9: Tidal response T as a function of time for three selected runs. We define T as the mass fraction of each disk that reaches a distance of more than $10R_\mathrm{d}$ to its center-of-mass. Also indicated as horizontal lines are the values of T_eff that we take as measure for the tidal response of the disk. Since we only stored simulation outputs with a spacing of 0.1 time units, the measurement of T_eff is slightly uncertain.

2.5.3 An indicator for tidal response

As we have seen above, knowledge of the halo-to-disk mass ratio is not sufficient to predict how prone a particular galaxy model is to tail formation. MMW suggested using the quantity

$$\mathcal{E} = \left[\frac{v_\mathrm{e}(R)}{v_\mathrm{c}(R)}\right]^2 \tag{2.41}$$

as a more suitable indicator. $\mathcal{E}$ compares the depth of the potential well with the specific kinetic energy of the disk material. The quantity $\mathcal{E}$ also arises, when one tries to estimate the relative increase ΔE of specific kinetic energy of disk stars in a nearly head-on encounter of identical disk models. In this situation one finds (Binney & Tremaine 1987; Mo, Mao & White 1998)

$$\Delta E/v_\mathrm{c}^2 \simeq \left(\frac{v_\mathrm{c}}{v_\mathrm{e}}\right)^2 = 1/\mathcal{E}. \tag{2.42}$$

This suggests the use of $\mathcal{E}$ as an indicator for the susceptibility of a disk model to tidal perturbations. In order to obtain a typical value for $\mathcal{E}$ we evaluate it at $R = 2R_\mathrm{d}$, which is about the half mass radius of the disk.

We now want to test how well $\mathcal{E}$ works as an indicator for the ability of a particular disk-halo model to develop massive tails. Here one encounters two immediate problems.

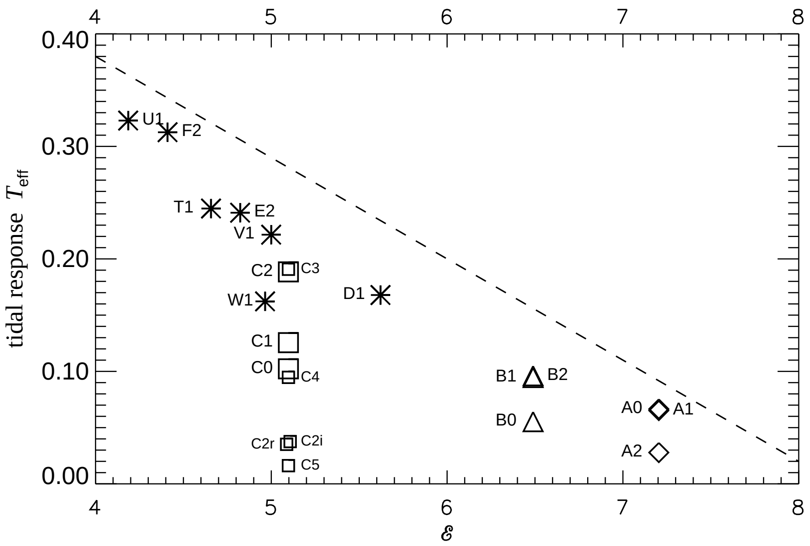


Figure 2.10: Tidal response $T_{\rm eff}$ of the different runs versus the value of $\mathcal{E}$ for the corresponding disk model, where $\mathcal{E}$ is evaluated at $R = 2R_{\rm d}$. The measurement of $T_{\rm eff}$ is slightly uncertain, and $T_{\rm eff}$ also depends on details of the orbital parameters of the encounter. However, our runs are designed to produce very nearly the strongest tails possible for a given disk model. At fixed $\mathcal{E}$, it should therefore be hard to find a model that gives a higher value for $T_{\rm eff}$ than the maximum of our runs. In other words, there should be no model in the region above the indicated dashed line.

First, the tidal response of a disk depends on the orbital parameters of the encounter with its companion. For example, if a disk is tilted against the orbital plane, the tidal forces felt by the bulk of the disk material will generally be smaller, resulting in a less prominent tidal tail. Similarly, a change of the impact parameter or the orbital energy can affect the tidal response. We here do not intend to investigate the complete parameter space. Rather, we focus on collisions that produce the strongest tails possible for mergers of a given disk model.

In this spirit, prograde encounters are an obvious choice for the orientations of the disks, since this configuration has repeatedly been shown to produce the strongest tails, and we here confirm this with the runs C2r and C2i. We use parabolic orbits, because they are plausible candidates for real interacting galaxies if one assumes that they are coming together for their first time. However, even when this assumption is dropped, DMH showed that moderately bound orbits are no more effective in producing tails than zero energy orbits. With respect to the 'impact' parameter $R_{\rm kep}$, we have examined a range of different choices and found that the tails appear to be of maximum strength for $R_{\rm min} \approx 2 - 6R_{\rm d}$, i.e. in collisions where the disks pass each other at a distance of a few disk scale lengths. Hence, our merger simulations of equal disk galaxies have been

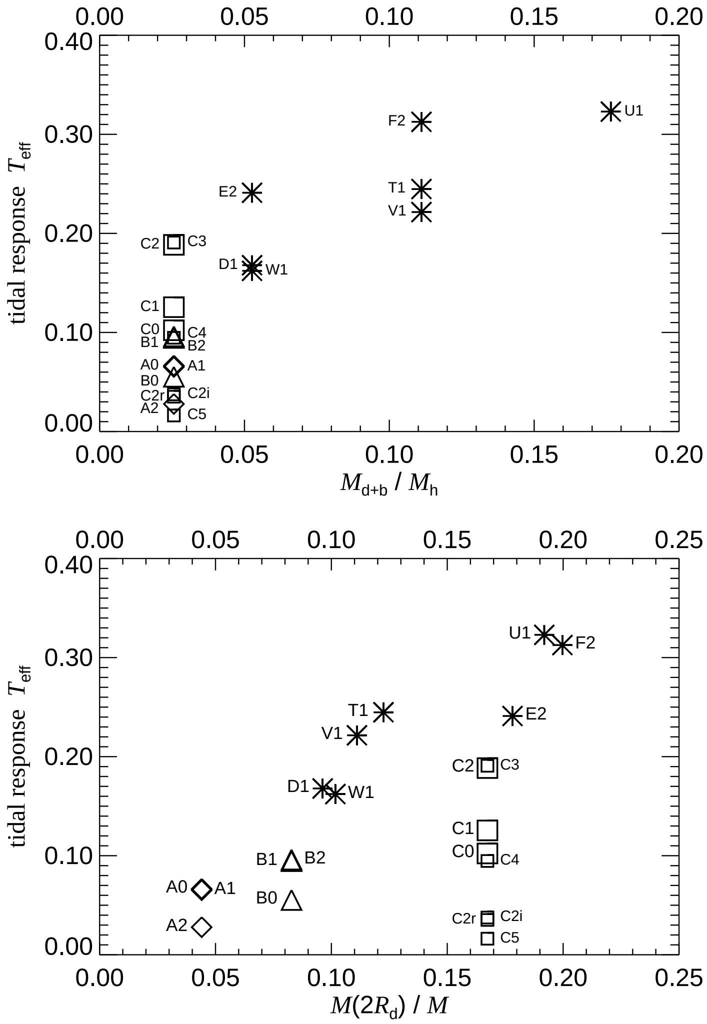


Figure 2.11: Tidal response $T_{\rm eff}$ of the different runs versus the disk-to-halo mass ratio (left panel), and versus $M(2R_{\rm d})/M$ (right panel). The quantity $M(2R_{\rm d})/M$ is a measure of the fraction of the total mass in the disk region.

set up to exhibit the most favourable conditions for tail formation, and should indeed produce the strongest tails possible for these disk models.

Second, it is not obvious how to define the mass or extent of the tidal tails in an objective way. This is further complicated by the rapid dynamical evolution of the tails, which makes it difficult to compare simulations that may form their tails at different times.

In order to solve this problem and measure the strength of the tidal response, we have come up with the following scheme. We start by defining the quantity T to be the mass

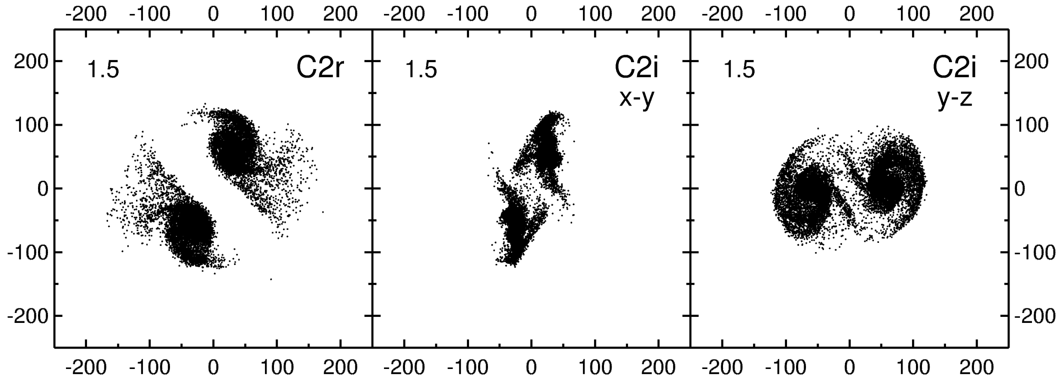


Figure 2.12: Tidal tails of simulations C2r and C2i. The panels show the disk particles, projected either onto the xy-plane (orbital plane), or the yz-plane. Apart from a different orientation of the disks, runs C2r and C2i are identical to C2. In C2r, both disks are retrograde, while in C2i the spins of the disks lie in the orbital plane, pointing along the x-axis. These runs may be compared with C2 shown in Figure 2.7.

fraction of each disk that reaches a distance of more than $10R_\mathrm{d}$ from its center-of-mass, where R_d is the original scale length of the unperturbed disk.

In Figure 2.9 we show examples for the time evolution of T. Shortly after the disks come together for the first time, T jumps up, reaches a maximum, and slowly decays, until the disks are scrambled up in their second encounter and T loses its initial meaning. Note that the different runs reach their maximum of T at different times. In order to compare them on an equal footing, we therefore define an effective response T_eff as the peak value reached by T.

In Figure 2.10 we plot the tidal response T_eff of our runs versus the value of $\mathcal{E}$, evaluated at $R = 2R_\mathrm{d}$. Although our coverage of parameter space is limited, it is nevertheless clear that there is a correlation between T_eff and $\mathcal{E}$. Of course, there is some uncertainty in the measurement of T_eff and this introduces some scatter. Also, for fixed $\mathcal{E}$, the tidal response T_eff depends somewhat on the impact parameter R_kep, and it might also have a slight dependence on m_d. However, to the extent that our simulations really produce the strongest tails possible for our disk models, Fig. 2.10 shows that it will be exceedingly hard to find a model that produces tails that *lie above* the dashed diagonal line. This establishes that $\mathcal{E}$ is a good indicator for the *maximum tidal response* T_eff, that may be obtained for a given class of disk models. In particular, models with $\mathcal{E} \geq 8$ should be unable to produce strong tails. This is in excellent agreement with the analysis of MMW, who estimated $\mathcal{E} = 4.2$, 5.5, 7.2, and 9.3 (in order of increasing halo mass) for the sequence of four models of DMH; in agreement with Figure 2.10, the last two of these models failed to produce prominent tails.

Figure 2.10 may also be compared to the two panels of Figure 2.11. In the left panel, we plot the tidal response versus the disk-to-halo mass ratio. This again shows, that the disk-to-halo mass ratio is not a good indicator for the ability to form tidal tails. For example, the models A, B, and C differ subtantially in the mass of their tails despite their equal disk-to-halo mass ratio. However, if we use instead the ratio of the *total* amount of mass in the region of the disk to the total mass of the galaxy, the mass ratio criterion can be partly resurrected. This is shown in the right panel of Fig. 2.11, where we plot $T_{\rm eff}$ versus $M(2R_{\rm d})/M$. Here $M(2R_{\rm d})$ is the total mass inside two disk scale lengths, and M is the total mass of the galaxy and its halo. In this formulation, the mass ratio measures the relative distribution of mass within the system, and is a fair indicator of the ability of a galaxy model to form tidal tails. However, a detailed comparison with Fig. 2.10 shows, that $\mathcal{E}$ does a better job than $M(2R_{\rm d})/M$. For example, the model T1 appears as an outlyer in Fig. 2.11, failing to fit the monotonic trend of larger $T_{\rm eff}$ with increasing $M(2R_{\rm d})/M$, but fits within the general distribution in Fig. 2.10.

Starting from a head on collision, Figure 2.10 also shows that the tidal response $T_{\rm eff}$ becomes larger as the impact parameter $R_{\rm kep}$ is increased. However, for very wide encounters one expects only a small distortion of the disks. As the sequence of models C0-C5 demonstrates, there is indeed a maximum response for an intermediate impact parameter $R_{\rm kep}$ of order a few disk scale lengths.

Also, the orientation of the disks is an important factor in determining the strength of the disk response. In the retrograde collision C2r, the spins of the galaxies are just reversed compared to C2, yet this already makes the tidal tails much weaker, as seen in Figure 2.12. Similarly, the inclined galaxies of simulation C2i produce tails that are less extended than those of run C2.

We also note that the simulations T1, U1, and V1, in which our 'disk-dominated' models collide, produce tidal tails well in line with the general trend of Figure 2.10.

2.5.4 A model with a bulge

In Figure 2.13 we show the time evolution of run W1, which collides two disk + bulge galaxies. These disk galaxies (our model W) are descendents of model D, but one third of the stellar mass has been put into a centrally concentrated bulge. This results in a rotation curve (Figure 2.3) that is practically flat up to the very center of the disk, with a shape quite similar to the models of DMH.

Due to their strong central concentration, the bulges survive largely unaffected until their final coalescence. However, the disks develop prominent tails that appear to be similar in strength to the other models with specific angular momentum corresponding to $\lambda = 0.05$. A measurement of $T_{\rm eff}$ shows that the strength of the tails is in fact quite similar to the directly comparable simulation D1. Also, the run W1 fits well into the plot of Figure 2.10, although the inner structure of model W is very different from that of the other models.

This suggests that bulges are not effective in preventing tail formation, at least as long as they primarily affect the inner rotation curve.

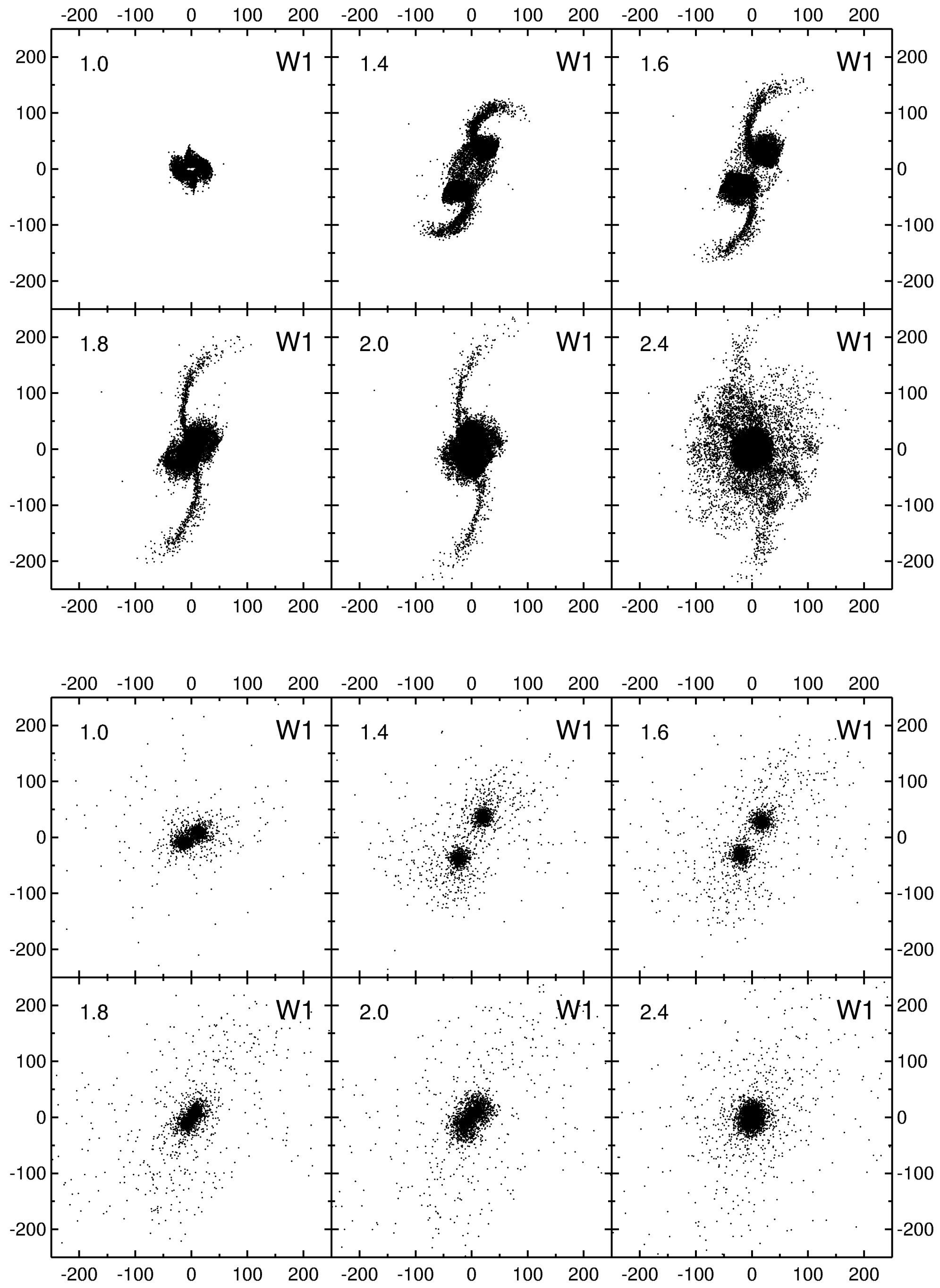


Figure 2.13: Time evolution of run W1, which collides our disk+bulge model with itself. The top series of panels displays the disk particles projected onto the orbital plane, while the bottom panels show the bulge particles. The length units refer to h^{-1}kpc, and the elapsed time since the start of the simulation (upper left corners) is given in units of 0.1 Hubble times, or $9.8 \times 10^8\, h^{-1}$yr.

2.6 Discussion

In this study we constructed N-body models of disk galaxies with structural properties directly motivated by current theories of hierarchical galaxy formation. In particular, the mass of the dark haloes in these models is much larger than that of the stellar disks. In the most extreme models we consider, the halo-to-disk mass ratio is larger than 40:1, yet these models can produce long and massive tidal tails, provided the spin parameter is not too small. The halo-to-disk mass ratio is thus not critical for tail-making ability.

Instead, the size of the disk compared to that of the halo seems to be the critical factor. In our approach, the size of the disk is tied to the spin of the dark halo. A larger spin parameter λ leads to larger disks. The bulk of the disk material is then more loosely bound in the dark matter potential well and can be more easily induced to form long tidal tails. This effect can be quantified in terms of the ratio $\mathcal{E}$ of the circular velocity to the escape speed at a radius $R = 2R_{\rm d}$. We have shown that $\mathcal{E}$ correlates well with the tidal response of the disk models. For models with $\mathcal{E} \geq 8$ we do not expect significant tails, while models with $\mathcal{E} \leq 6.5$ can produce substantial tails.

When $\mathcal{E}$ is used to characterize the tail-making ability, the results of DMH agree with our own. In their sequence of four models not only the mass of the halo changes, but also its spatial extent. We think the latter effect is critical in defining tail-making ability, because the relative size of disk and halo affects the value of $\mathcal{E}$ strongly. This also agrees with the earlier conclusion of Barnes (1997, private communication). For given *inner* structure and given mass ratio $M_{\rm d}/M_{\rm h}$, less extensive halos have larger $\mathcal{E}$ and so make weaker tails.

We focused in this work on just one halo mass. Note however, that the shape of the rotation curves does not depend on our particular choice for v_{200}. Rotation curves with other peak velocities may be realized by an appropriate scaling of v_{200}.

According to the work of NFW, the shape of the dark matter profile is insensitive to cosmology. Also, the distribution of λ is universal and independent of the initial power spectrum. Furthermore, the average value of the concentration c does not vary with halo mass, and it depends only weakly on cosmology. For low-density flat universes a smaller value, $c \simeq 5$, is probably more appropriate than the value $c = 15$ employed in most of the models in this work. Such a smaller concentration is also supported by observational data (Navarro 1998). However, our model 'V1', which has $c = 5$, produces tails well in line with the other models in Figure 2.10. Also note that according to Figure 2.2, smaller c gives rise to larger disks, in principle favouring even stronger tails.

This suggests that all 'reasonable' CDM cosmologies can produce disk galaxies with $\lambda \geq 0.05$ that are roughly equally capable of producing tidal tails when they collide and merge with similar objects. We conclude that the observed lengths of tidal tails in interacting galaxies are consistent with current CDM cosmologies, and that tidal tails are not useful to discriminate between different flavours of these scenarios.

However, tidal tails remain a useful tool to probe the structure of individual galaxies, since the length and mass of tails formed in an interaction is a strong function of the rotation curve and the potential in the region of the galactic disk. The presence of tidal tails can thus provide important constraints on models of disk galaxies. In fact, these constraints are potentially very powerful if dynamical modeling of tail formation is combined with detailed observations of the velocity field in merging pairs of galaxies.

We hope that the N-body representations of disk models constructed in this work may be more realistic caricatures of real spiral galaxies than those of previous work. In particular, the structural properties of our models are motivated by hierarchical structure formation and are less ad hoc than in previous simulations of this kind. These models should be useful for future work on galaxy evolution.

–Q: Where's the rest of the moon when it's not a full moon?
–A: When they landed on the moon in 1969, the astronauts shoveled most of the moon's surface into special containers and took it home. They would have taken the whole thing, but they needed to keep some dirt there to hold the flag up. If you see something that looks like a full moon, that's either false memory or someone playing a practical joke on you.

Scott Adams, Journey to Cubeville

3

Modeling star formation and feedback in simulations of interacting galaxies

Abstract

We discuss a heuristic model to implement star formation and feedback in hydrodynamical simulations of galaxy formation and evolution. In this model, gas is allowed to cool radiatively and to form stars with a rate given by a simple Schmidt-type law. We assume that supernovae feedback results in turbulent motions of gas below resolved scales, a process that can effectively pressurize the diffuse gaseous medium, even if it lacks substantial thermal support. Ignoring the complicated detailed physics of the feedback processes, we try to describe their net effect on the interstellar medium with a fiducial second reservoir of internal energy, which accounts for the kinetic energy content of the gas on unresolved scales. Applying the model to three-dimensional, fully self-consistent models of isolated disk galaxies, we show that the resulting feedback loop can be modeled with smoothed particle hydrodynamics such that converged results can be reached with moderate numerical resolution. With an appropriate choice of the free parameters, Kennicutt's phenomenological star formation law can be reproduced over many orders of magnitude in gas surface density. We also apply the model to mergers of equal mass disk galaxies, typically resulting in strong nuclear starbursts. Confirming previous findings, the presence of a bulge can delay the onset of the starburst from the first encounter of the galaxies until their final coalescence. The final density profiles of the merger remnants are consistent with de Vaucouleurs profiles, except for the innermost region, where the newly created stars give rise to a luminous core with stellar densities that may be in excess of those observed in the cores of most elliptical galaxies. By comparing the isophotal shapes of collisionless and dissipative merger simulations we show that dissipation leads to isophotes that are more disky than those of corresponding collisionless simulations.

3.1 Introduction

Numerical studies of large-scale structure formation have greatly helped to gain insight into the collisionless dynamics of dark matter in cold dark matter (CDM) universes. In contrast to the rapid progress in this field, modeling the formation of the luminous parts of galaxies has proven to be far more difficult. Here, a large range of complicated physical processes have to be considered that go far beyond the regime of 'simple' gravitational dynamics. In order to properly model the formation processes of luminous stars, the hydrodynamics of the interstellar medium (ISM) has to be followed, including gas shocks, radiative heating and cooling processes, and the actual formation and fragmentation of cold molecular clouds. In addition, newly formed stars will influence the surrounding ISM by energy input in the form of supernova explosions, stellar winds, or UV radiation. As a result, the ISM is believed to exhibit a complicated multi-phase structure (McKee & Ostriker 1977; Shu et al. 1987). Further complications arise from magnetic fields and gas turbulence (Norman & Ferrara 1996; Mac Low et al. 1998), which may play an important role in stabilizing molecular clouds against collapse. Also, the gas dynamics will remain tightly coupled to the gravitational dynamics of the whole galaxy at all times, e.g. star formation in the disk may be triggered and regulated by large-scale gravitational instabilities in the disk (Wyse & Silk 1989; Wang & Silk 1994; Silk 1997).

While it is obvious that one would like to include star formation in cosmological simulations of galaxy formation, it is also clear that these physical processes are too complicated to be modeled from first principles. And even if this were possible, a three-dimensional simulation that tries to address all the relevant physics would be not feasible with current technology, just because the dynamic range of the problem is so large. For example, a typical size of a cold gas cloud is about $10 - 100\,\mathrm{pc}$, while that of the galaxy is of order $100\,\mathrm{kpc}$, and scales above $1\,\mathrm{Mpc}$ are important for its formation. Given that the cloud has a complicated internal structure, one faces a problem with spatial dynamic range of at least 10^7. As a consequence, only two-dimensional simulations have so far been able to model the turbulent ISM on a global scale (Rosen & Bregman 1995; Wada & Norman 1999). However, in three dimensions the resolution problem will stay for some time until it might eventually be overcome with more sophisticated simulation codes, and with faster computer hardware. Adaptive mesh refinement codes (Norman & Bryan 1998; Klein et al. 1998) or parallel tree-SPH codes (Davé et al. 1997) are tools that potentially will be able to achieve sufficient dynamic range, at least for a small localized volume.

In the meantime, the problem might be amenable to heuristic solutions that avoid the brute-force approach required by attempts to correctly model all the relevant physics. Observationally, the star formation rate follows surprisingly well a global Schmidt-law (Kennicutt 1983, 1989, 1998), i.e. the star formation properties correlate well with the local gas density. This suggests that a simple phenomenological approach to star formation can be employed, in which the properties of the ISM are averaged over scales of order $100\,\mathrm{pc}$. In such an approach, we gloss over the details of the physics of star formation, but we might still get important clues about the spatial and temporal evolution of the star formation rate. For example, such a model may be applied to interactions of galaxies and to the study of the starbursts triggered by them.

In recent years, a number of authors have worked along such lines, coupling smoothed particle hydrodynamics (SPH) codes with simple star formation and feedback prescriptions (Katz 1992; Navarro & White 1993; Mihos & Hernquist 1994b; Steinmetz & Müller 1995; Tissera et al. 1997; Gerritsen & Icke 1997; Hultman & Källander 1997), or they developed such models for hydrodynamical mesh codes (Yepes et al. 1997).

A robust result of these studies has been that "thermal" supernova feedback is not effective in regulating star formation, i.e. if the energy released by supernova was assumed to simply heat the ISM, its dynamics was hardly affected. This is because the dense gas in star forming regions is able to radiate away this heat input very quickly. However, an introduction of "kinetic" feedback made it possible to obtain quasi-stationary models of disk galaxies with star formation rates resembling normal spiral galaxies. In most SPH simulations, this kinetic feedback has been implemented by imparting radial momentum kicks on neighboring SPH particles. In this study, we try to improve on such a feedback implementation, which is plagued by numerical intricacies, and directly resorts to the particle formalism of SPH. We instead investigate a model of the star formation/feedback loop that is based on the notion of an effective equation of state for the ISM, averaged over scales of order 100 pc. In essence, we augment the thermal pressure of the gas with a turbulent pressure, while not resolving the turbulence itself. This "sub-grid" model does not directly resort to the particle formalism of SPH, and has advantageous numerical properties. It leads to a smooth description of star formation, which can be shown to result in quantitatively converged results even with moderate numerical resolution. Our model is able to quantitatively reproduce the star formation law of Kennicutt (1998).

Once isolated disk galaxies can be modeled as quiescently star forming galaxies, one can try to explore more extreme environments by hoping that the model carries over to such regimes. For example, one may ask how the star formation rate evolves in major mergers of gas rich spirals. In previous simulations of this kind (Barnes & Hernquist 1991; Mihos & Hernquist 1994a, 1996; Barnes & Hernquist 1996) it was shown that major mergers can produce strong nuclear starbursts with sufficient strength to plausibly account for ultraluminous infrared galaxies (ULIRGS). However, a problem of the simulations has been that the nuclear starbursts led to the formation of extremely dense and concentrated cores of young stars in the center of the merger remnants (Mihos & Hernquist 1994a). The resulting break in the stellar profile seems not to be observed in this form in real elliptical galaxies, so one aim of this work is to examine to what extent such stellar cores can be circumvented by our new prescription to implement feedback. To this end, we simulate a number of major mergers, and we compare the structure of the merger remnants with those obtained in corresponding collisionless simulations. While the surface brightness profiles of our remnants follow $r^{1/4}$-laws for most of their light, the dissipative simulations exhibit a central luminosity excess, quite similar to the results of Mihos & Hernquist (1994a).

Our set of merger simulations also allows us to study the effect of dissipation on the isophotal shapes of the merger remnants. We demonstrate that dissipation produces "disky" isophots, confirming a long-standing hypothesis in studies of the systematic properties of elliptical galaxies.

This chapter is organized as follows. In Section 2 we describe in detail the assumptions of our modeling, while we briefly describe its numerical implementation in Section 3. We then consider simulations of isolated disk galaxies in Section 4, and we examine major mergers in Section 5. Finally, we discuss our findings in Section 6.

3.2 Model ingredients

3.2.1 Dark matter, stars, and gas

Initially, our galaxy models contain a dark matter halo, a stellar disk, an optional stellar bulge, and a gaseous disk capable of forming further stellar material. The structural properties of our model galaxies are derived according to the analytic work of Mo, Mao & White (1998), and the details of how we construct N-body representations of these galaxies are described in Chapter 2. Note that our initial conditions do not explicitely contain a hot gaseous phase in the galactic halo, although such a medium does form by shock heating in our merger simulations. In studies of galaxy formation, feedback is often invoked as a mechanism to regulate *further gas inflow*. In this study we mostly deal with another aspect of feedback, the one that refers to the regulation of star formation of the gas that has already cooled and settled to a disk.

In the collisionless limit, the dark matter and the stars follow the coupled Poisson-Vlasov equations, and we solve this system with the usual N-body approach. In contrast to that, the interstellar gas is more difficult to model, since it is collisional and may develop shocks, and it can participate in radiative processes and the poorly understood physics of star formation. To first approximation we treat the interstellar medium (ISM) as an ideal inviscid gas, i.e. it follows the Euler equation

$$\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = -\frac{\nabla P}{\rho} - \nabla\Phi, \tag{3.1}$$

and the energy equation

$$\frac{\mathrm{d}u}{\mathrm{d}t} = -\frac{P}{\rho}\nabla\cdot\mathbf{v}. \tag{3.2}$$

Here $\mathbf{v}$ and ρ describe the velocity field and the the density of the gas, and u is its internal energy per unit mass, while Φ is the total gravitational potential of the system. We take the thermal pressure to be $P = (\gamma - 1)\rho u$, with $\gamma = 5/3$, appropriate for a monoatomic ideal gas. Below we will supplement this simple gas dynamics with additional processes in order to model star formation and feedback.

3.2.2 Cooling

In contrast to dark matter, baryonic gas can lose energy by radiative cooling. This is an important process for the formation of the luminous part of galaxies, since it allows the

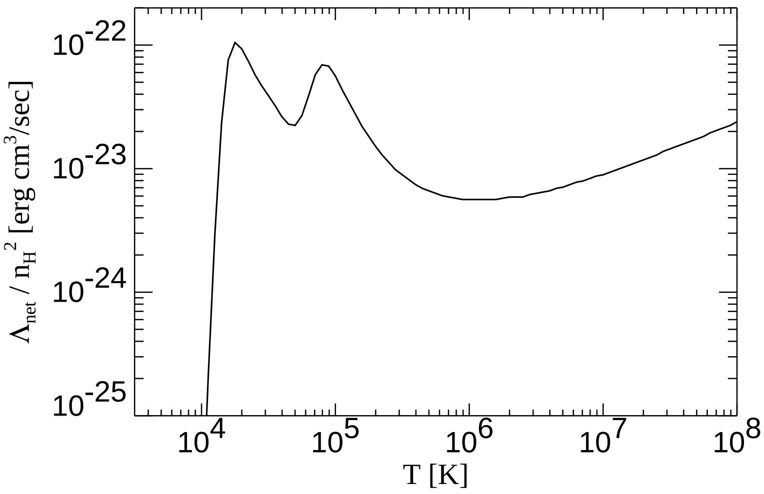


Figure 3.1: Cooling rate for a primordial mix of H and He according to Sutherland & Dopita (1993).

gas to sink to the centers of the dark matter potential wells, and to ultimately form the stars we see today.

We adopt cooling functions computed for collisional ionisation equilibrium (CIE) by Sutherland & Dopita (1993). The loss of energy per unit mass of gas is given by

$$\left(\frac{\mathrm{d}u}{\mathrm{d}t}\right)_{\mathrm{cool}} = -\frac{\Lambda_{\mathrm{net}}(\rho, T)}{\rho}. \tag{3.3}$$

For simplicity, we assume a primordial composition of hydrogen and helium, and we neglect changes of the metallicity of the gas in this work. In Figure 3.1, we show the adopted net cooling function for a primordial H/He mix, computed for collisional ionization equilibrium by Sutherland & Dopita (1993). Note that the temperature of the gas is given by

$$T = \frac{\overline{\mu} P}{k\rho} = \frac{(\gamma - 1)\overline{\mu}}{k}\, u, \tag{3.4}$$

where $\overline{\mu}$ is the mean particle mass. While the gas is fully ionized at temperatures above $\simeq 10^{5.2}$ K, it starts to recombine at lower temperatures. In this transitional region, $\overline{\mu}$ changes due to the varying number density of electrons, which needs to be taken into account in equation (3.4).

Below 10^4 K the gas is basically neutral and does not cool any further by simple line cooling or Compton cooling. In principle, there are less efficient molecular cooling processes that can eventually further reduce the gas temperature (Abel et al. 1997). These processes are highly relevant for the formation and fragmentation of dense molecular clouds (Abel et al. 1998), and for the generation of the cold, dense filaments that thread the surrounding warm gas in the ISM (Rosen & Bregman 1995). However, we do not attempt to resolve the true multi-phase structure of the ISM in this work, since we clearly lack resolution to do that. Rather we implement an effective star formation model that keeps the gas in a relatively smooth phase, which is probably best identified with the layer of warm, neutral gas observed in the Galaxy.

Note that the gas may also be heated by an ambient ultraviolet (UV) background or by UV radiation of newly formed stars. In this work, we do not consider such sources of heating. While an ambient UV field has had very little effect on star formation in the simulations of Navarro & Steinmetz (1997) and Weinberg et al. (1997), Gerritsen & Icke (1997) successfully used feedback by local UV sources to self-regulate their star formation model.

In regions of high density, the cooling time of the gas can become very short. In this regime the gas may become thermally unstable, and it then collapses essentially on a free fall time. If no further physics is included, this *catastrophic cooling* results in knots of cold, dense gas, with density and size set by the resolution of the gasdynamical simulation. This situation is not only numerically unpleasant, it also appears to be unphysical. In reality, star formation and its accompanying feedback processes are thought to alter the dynamics of the gas phase in the high density regime. These processes will prevent a collapse of all of the gas into high density clumps. Hence the inclusion of cooling requires the consideration of such additional processes.

3.2.3 Star formation

We take the star formation rate per unit volume to be proportional to the local gas density, and inversely proportional to the local dynamical time, viz.

$$\frac{\mathrm{d}\rho_\star}{\mathrm{d}t} = \alpha \frac{\rho}{t_{\mathrm{dyn}}}, \tag{3.5}$$

where $\rho_\star$ denotes the density of stars, and the dynamical time of the gas is given by $t_{\mathrm{dyn}} = (G\rho)^{-0.5}$. This parameterization leads to a Schmidt-type law $\dot{\rho_\star} \propto \rho^n$ with $n = 1.5$, which empirically has been used for a long time to describe the gross star formation properties of galaxies. A parameterization of the form (3.5) has also been employed in many numerical simulations that included star formation, and also in semi-analytic studies of galaxy formation (Kauffmann et al. 1993). Of course, the star formation also depletes the gas according to

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = -\alpha\, G^{0.5} \rho^{1.5}. \tag{3.6}$$

3.2.4 Feedback model

One of the robust result of previous simulations has been that thermal feedback alone has hardly an effect on the evolution of the ISM and on the star formation rate. This is caused by the short cooling times in dense star forming regions; the heat input by supernovae is just radiated away nearly instantly. To remedy this problem, Navarro &

White (1993) proposed to put some of the feedback energy directly into kinetic energy of the gas. In their SPH implementation of this idea, they imparted small radial momentum kicks to the particles around a star-forming SPH particle. While this scheme proved to be able to regulate the star formation rate, it is numerically and physically not without problems. First, the feedback description is formulated directly in terms of the SPH particle formalism, and its free parameters are thus bound to depend on details of numerical resolution and implementation of SPH. Also, the radial motions triggered by the star-forming particle seem unphysical, since the real expanding shells of supernova remnants are several orders of magnitude smaller than the typical resolution limit of current cosmological SPH simulations.

We here want to improve on the numerical properties of such a kinetic feedback scheme. In particular, the model should achieve converged numerical results in hydrodynamical simulations using moderate numerical resolution, and the small number of tunable parameters of the model should be independent of resolution. Only in this situation we can expect to obtain meaningful results if the model is applied to cosmological simulations of galaxy formation, where objects of a large range in masses form, and the first objects are inevitably small. At this stage of the modeling we therefore give well controlled numerical properties precedence over the implementation of more detailed physics.

Star-forming molecular clouds are known to have lifetimes much larger than their free-fall time under gravitational collapse. However, since the density in these clouds is high enough to allow efficient cooling, it can not be ordinary thermal pressure that supports them. Instead, turbulent gas motions caused by supernova explosions and stellar winds have been suggested as a mechanism of support for the gas. This turbulence could be of a supersonic hydrodynamical nature, or it could involve magnetic fields leading to supersonic or super-Alfvenic magnetic turbulence (Mac Low et al. 1998). By the same token, it is clear that the temperature of the interstellar HI in the Galaxy is far too low to provide enough thermal pressure to support the HI layer against collapse under its own weight. The support is likely provided by turbulent bulk motions of the gas (Lockman & Gehman 1991). This turbulence may be fed by supernovae explosions, superbubbles, stellar winds, HII regions, or differential sheer in the disk. By analysing the grand source function of this turbulence, Norman & Ferrara (1996) showed that the turbulent pressure is given by $p_{\rm turb} \sim 10-100\, p_{\rm th}$, i.e. it typically dominates the thermal pressure.

Our new approach to feedback therefore rests on the assumption that the ISM is primarily supported by turbulent pressure, and that this turbulence is sustained by kinetic energy feedback of exploding supernovae. We model this feedback by a second reservoir of internal energy, q, of the gas. This reservoir is introduced to describe the energy content per unit mass due to turbulent motion of gas at scales well below the spatial resolution limit. We further assume that the energy content q effectively pressurizes the gas, and we take the pressure of this medium to be

$$P = (\gamma - 1)(u + q)\rho. \tag{3.7}$$

Note that the temperature of the gas is still computed based on u alone. Hence, for $q \gg u$ the gas may be pressure supported even if it is cold.

Following Navarro & White (1993), we adopt a power-law initial mass function (IMF) with a slope of 1.5 between 0.1 and $40\,M_\odot$, and we assume that all stars above $8\,\mathrm{M}_\odot$ go supernova instantly with the release of 10^{51} ergs of energy. The local energy input due to supernovae is then $\epsilon_{\rm fb}\mathrm{d}\rho_\star/\mathrm{d}t$, with $\epsilon_{\rm fb} = 4\times10^{48}\,\mathrm{ergs\,M_\odot^{-1}}$. We assume that this energy is first released into the turbulent reservoir q, i.e.

$$\left(\frac{\mathrm{d}q}{\mathrm{d}t}\right)_{\rm feedback} = \epsilon_{\rm fb}\frac{1}{\rho}\frac{\mathrm{d}\rho_\star}{\mathrm{d}t}. \tag{3.8}$$

We further invoke a process that thermalizes this kinetic energy and transforms it into 'ordinary' thermal energy. We parameterize this process as

$$\left(\frac{\mathrm{d}q}{\mathrm{d}t}\right)_{\rm therm} = -\left(\frac{\mathrm{d}u}{\mathrm{d}t}\right)_{\rm therm} = -f(\rho)\,q, \tag{3.9}$$

where $f(\rho)$ is some function of density. For reasons that we will describe in more detail below, we choose $f(\rho) = \beta/\sqrt{\rho}$ for the functional dependence of $f(\rho)$, i.e. the thermalization proceeds more slowly in high-density regions.

There are then two free parameters in the model, α and β. Note that for $\beta \to \infty$ we have ordinary thermal feedback, while $\beta = 0$ corresponds to maximum kinetic feedback, where all the feedback energy is collected in q and star formation will eventually be quenched.

In summary, the equations governing the internal energy of the gas phase look like

$$\frac{\mathrm{d}u}{\mathrm{d}t} = -(\gamma-1)\,u\,\nabla\cdot\mathbf{v} - \frac{\Lambda(u,\rho)}{\rho} + \beta\frac{q}{\rho^{0.5}}, \tag{3.10}$$

$$\frac{\mathrm{d}q}{\mathrm{d}t} = -(\gamma-1)\,q\,\nabla\cdot\mathbf{v} + \alpha\epsilon_{\rm sfb}(G\rho)^{0.5} - \beta\frac{q}{\rho^{0.5}}, \tag{3.11}$$

where the adiabatic changes are described by the terms involving the velocity divergence.

To elucidate the essential properties of this model, consider a portion of gas with $\nabla\cdot\mathbf{v} \simeq 0$, but which is dense enough to form stars at appreciable rate. Then the gas will also be able to cool efficiently, i.e. the energy input by the feedback into the thermal reservoir will not be able to raise the temperature of the gas above $\sim 10^4$ K. On the other hand, the reservoir q will take on an equilibrium value of

$$q = \epsilon_{\rm fb}G^{0.5}\alpha\beta^{-1}\rho. \tag{3.12}$$

For high densities, efficient feedback will lead to $q \gg u$, resulting in an effective pressure of the gas of

$$P \simeq (\gamma-1)\epsilon_{\rm fb}G^{0.5}\alpha\beta^{-1}\rho^2. \tag{3.13}$$

In star forming regions, we therefore obtain an effective equation of state in the form

$$P = c\rho^2, \tag{3.14}$$

with $c \equiv (\gamma - 1)\epsilon_{\rm fb} G^{0.5} \alpha \beta^{-1}$. In principle, we can adjust the polytropic index of the gas by changing our model assumptions about the function $f(\rho)$. However, after some experimentation, we think that the choice leading to Eqn. (3.14) has two desirable properties, making it superior to other possibilities in the framework of the present model.

First, a self-gravitating sheet of gas with an equation of state of the form (3.14), and a surface mass density σ, has the density distribution

$$\rho(z) = \frac{k\sigma}{2} \cos(kz), \quad \text{for } |kz| \leq \frac{\pi}{2}, \tag{3.15}$$

with

$$k^2 = \frac{2\pi G}{c}. \tag{3.16}$$

Remarkable about this solution is the lack of dependence of the vertical scale height

$$z_0^2 \equiv \frac{\int z^2 \rho \,\mathrm{d}z}{\int \rho \,\mathrm{d}z} = \left(\frac{\pi^2}{4} - 4\right) \frac{c}{4\pi G} \tag{3.17}$$

on the surface mass density. This suggests that it should be comparatively easy to reconcile this model with the observed constancy of the stellar and gaseous scale heights with radius in many disk galaxies. Note that for the stiff equation of state (3.14), the volume polytropic index is equal to the surface polytropic index (Hunter 1972; Laughlin et al. 1998).

In conjunction with our star formation law, a further consequence of equation (3.15) is that the star formation rate per unit area of a sheet of gas is given by

$$\Sigma_{\rm SFR} = \frac{\sqrt{\pi}\,\Gamma(1.25)}{\Gamma(1.75)} \alpha (Gk)^{0.5} \left(\frac{\sigma}{2}\right)^{1.5}, \tag{3.18}$$

i.e. we have a Schmidt-law $\Sigma_{\rm SFR} \propto \sigma^{1.5}$ not only for the volume density, but also for the surface mass density of the gas. Such a law has indeed been found by Kennicutt (1998). Therefore, we now further investigate the proposed model by trying to model the gaseous disks in galaxies as self-gravitating sheets of gas. We note that for the Galaxy, a stable hydrostatic equilibrium configuration of the gas has been shown to provide a good description of the data, at least on large scales (Kalberla & Kerp 1998; Bloemen 1987).

3.2.5 Designing Kennicutt's law

By studying the star formation rates and gas densities of a large sample of normal disk galaxies, and of star-bursting systems, Kennicutt (1998) has found the remarkable result, that on average the star formation rate per unit area and the surface gas density are related by[1]

$$\Sigma_{\rm SFR} = (2.5 \pm 0.7) \times 10^{-4} \left(\frac{\Sigma_{\rm gas}}{\rm M_\odot pc^{-2}}\right)^{1.4\pm0.15} \frac{\rm M_\odot}{\rm yr\,kpc^2}. \tag{3.19}$$

[1]for $H_0 = 75\,\mathrm{km\,s^{-1}}$.

This 'global' Schmidt-law holds over an exceptionally wide range of scales, bridging at least 5 decades in density, and 7 decades in star formation rate.

The composite Schmidt-law (3.19) has been obtained based on suitably defined averages of gas and star formation densities of whole galaxies, but Kennicutt also studied the azimuthally averaged star formation rate per unit area *as a function of radius* in normal disk galaxies. Above $\Sigma_{\rm gas} \simeq 10\,{\rm M}_\odot {\rm pc}^{-2}$ the local star formation rate scales with gas density just in the way implied by equation (3.19). However, below this gas density, the star formation rate drops very steeply. Thus, there also exists a rather sharp threshold in gas density; below it, star formation is strongly suppressed.

We take Kennicutt's findings as a phenomenological basis to set the free parameters of our star-formation/feedback model. Looking at equations (3.17) and (3.18), we can express the free parameters α and β as

$$\alpha = 1.96 G^{-0.5} \frac{\Sigma_{\rm SFR}}{\Sigma_{\rm gas}^{1.5}}\, z_0^{0.5}, \tag{3.20}$$

$$\beta = 0.146(\gamma - 1)\epsilon_{\rm fb} G^{-1} \frac{\Sigma_{\rm SFR}}{\Sigma_{\rm gas}^{1.5}}\, z_0^{-1.5}. \tag{3.21}$$

Based on Kennicutt's law for the azimuthally averaged star formation rate we take a value of

$$\frac{\Sigma_{\rm SFR}}{\Sigma_{\rm gas}^{1.5}} = \frac{10^{-1.5} h^2 {\rm M}_\odot {\rm yr}^{-1}\,{\rm kpc}^{-2}}{(100\,h\,{\rm M}_\odot {\rm pc}^{-2})^{1.5}} \tag{3.22}$$

for the coefficient of the star formation law. We then need to assume a thickness z_0 for the gaseous sheet to determine α and β. In this work, we take $z_0 = 0.2\,h^{-1}{\rm kpc}$. This results in

$$\alpha = 0.0131 \left(\frac{z_0}{0.2\,h^{-1}{\rm kpc}}\right)^{0.5}, \tag{3.23}$$

and

$$\beta = 0.0051\, h^2 {\rm M}_\odot^{0.5} {\rm yr}^{-1} {\rm kpc}^{-1.5} \left(\frac{z_0}{0.2\,h^{-1}{\rm kpc}}\right)^{-\frac{3}{2}}. \tag{3.24}$$

We also introduce an explicit cut-off in the star-formation law by requiring the star formation rate to be strongly suppressed below $\Sigma_{\rm cut} = 10\,h\,{\rm M}_\odot {\rm pc}^{-2}$. We relate this cut-off surface density to a volume density by means of $\rho_{\rm cut} = k\Sigma_{\rm cut}/2$. This results in

$$\rho_{\rm cut} = 0.342 \frac{\Sigma_{\rm gas}}{z_0} = 0.0171 \left(\frac{z_0}{0.2\,h^{-1}{\rm kpc}}\right)^{-1} h^2 {\rm M}_\odot {\rm pc}^{-3} \tag{3.25}$$

Finally, we alter the star formation rate of equation (3.5) by setting $\dot\rho_\star = 0$ for $\rho < \rho_{\rm cut}$.

We have now fixed the free parameters of our star formation model *before* we have actually done a numerical simulation. In the following sections we will investigate how well this model performs in self-consistent SPH-simulations of individual and colliding disk galaxies.

3.3 Numerical techniques

In summary, the gasdynamical model outlined above is described by equations (3.1), (3.7), (3.10), and (3.11). We supplement this system with the collisionless dynamics of dark matter and stellar material, and evolve it under self-gravity with our new SPH code GADGET. Algorithmic and numerical details of this code will be described in Chapter 5. We here just discuss our methods to implement the actual formation of collisionless stellar material, and the depletion of the gas content.

We have formulated star formation in an entirely continuous fashion, and we would like to obtain an equally continuous description of these processes in the particle-based SPH formalism. However, it is not feasible to spawn new independent star particles at every timestep, for every gas particle, just because the resulting number of stellar particles would quickly grow prohibitively large, and the small mass of these particles would lead to excessive two-body heating in encounters with the heavy dark matter particles. To avoid this, we treat the gas particles as hybrid gas/star particles in the way suggested by Mihos & Hernquist (1994b). Each SPH particle carries a gas mass, and a collisionless mass corresponding to the stellar material formed by this particle. The gasdynamical interactions are computed with the gas mass only, while both mass contributions participate in the gravitational interaction. Star formation is then simply a conversion of some of the gas mass into collisionless stellar mass, with the total mass of the hybrid particle being constant.

While this scheme allows a smooth implementation of star formation, it has the disadvantage that the formed collisionless matter remains coupled to the gas phase. To cure this problem, we *also* spawn independent stellar particles. To this end, each SPH-particle computes the sum of the stellar mass that has been formed in the region of its smoothing neighbours. If this exceeds some predefined mass $m_\star$ (typically half the initial gas particle mass), a new stellar particle of mass $m_\star$ is formed, where the stellar mass of this particle is collected by kernel weighting from the SPH neighbours. The particle is created at the center-of-mass position and with the center-of-mass velocity of the various mass parts taken from surrounding gas particles. In this way, the fraction of newly created stellar material that is still bound to SPH particles falls below 30 per cent soon after the simulation is started.

In strongly star forming regions, the gas mass can be heavily depleted. In this case, SPH particles would ultimately turn into very light gas particles. To maintain a roughly constant gas mass resolution, we also dissolve gas particles completely once their mass drops below a prescribed fraction of their initial mass. In this case, the mass and energy of the particle is distributed kernel-weighted among its SPH-neighbours.

3.4 Isolated disk galaxies

In the following we consider models of disk galaxies with structural properties set according to the analytical study of Mo et al. (1998). The structure of the dark matter halo is assumed to follow the NFW-profile (Navarro et al. 1996, 1997), adiabatically modified by the formation of a stellar disk. The dark halo is described by its circular velocity v_{200}, its spin parameter λ, and its concentration c. The mass and virial radius of the halo are then given by

$$M_{200} = \frac{v_{200}^3}{10GH(z)}, \quad \text{and} \quad r_{200} = \frac{v_{200}}{10H(z)}. \tag{3.26}$$

We assume that a centrifugally supported stellar disk of mass $M_\mathrm{d} = m_\mathrm{d} M_{200}$ has formed inside the halo, and that the specific angular momentum of the disk material is equal to that of the halos. Adopting an exponential disk for the radial structure of the stellar disk, the disk scale length R_d can then be computed. Vertically, we model the stellar disk as an isothermal sheet with scale-length $R_z = 0.2 R_\mathrm{d}$.

Fully self-consistent N-body representations of such models can be constructed in the way outlined by Springel & White (1999). In this study, we supplement the disk models by an additional gaseous phase in the disk. For this purpose, we assume that a fraction f_gas of the disk mass is in the form of gas. Initially, we distribute the gas just like the stellar material in the disk. However, we only assign an azimuthal velocity component equal to the local circular velocity to the gas, and we set its initial internal energy per unit mass equal to $u = \overline{v_z^2}/(\gamma - 1)$. If the gas is allowed to cool radiatively, it will then quickly settle into a thin gaseous disk.

3.4.1 Cooling and star formation without feedback

We start by considering a simulation where the gas is allowed to cool radiatively and to form stars, but where no feedback effects are taken into account. For definiteness, we select an isolated disk galaxy with parameters $v_{200} = 160\,\mathrm{km\,s^{-1}}$, $c = 5$, $\lambda = 0.05$, $m_\mathrm{d} = 0.05$, and $f_\mathrm{gas} = 0.2$, and we label this model "I1". These parameters result in a disk scale-length of $R_\mathrm{d} = 4.5\,h^{-1}\mathrm{kpc}$. We employ 20000 gas, 20000 disk, and 30000 dark matter particles for this simulation, and we use a gravitational softening of $0.4\,h^{-1}\mathrm{kpc}$ for the dark halo, and $0.1\,h^{-1}\mathrm{kpc}$ for the gas, disk, and newly formed stellar particles.

In Figure 3.2 we show the time evolution of this model. After the start of the simulation, most of the gas quickly settles into a very thin disk at a temperature of $T \simeq 10^4\,\mathrm{K}$ due to efficient cooling. In Figure 3.3 we show a phase diagram of the SPH particles in the temperature-density plane for the initial gas distribution, and for the gas after 0.1 internal time units[2]. At that point of time, most of the gas has already cooled to $10^4\,\mathrm{K}$ due to the short cooling times in the initial gaseous disk.

Note that an isothermal sheet at a temperature of $10^4\,\mathrm{K}$, and gas surface density similar to our model galaxy, would have a thickness of just $\simeq 0.04\,h^{-1}\mathrm{kpc}$ at the center. This is already below the gravitational softening length of the current simulation, so a simulation with very high resolution would be necessary to accurately follow the self-gravity of the gas in such a situation. On the other hand, the observed thickness of the HI

[2]The internal unit time is $9.8 \times 10^8 h^{-1}\mathrm{yr}$.

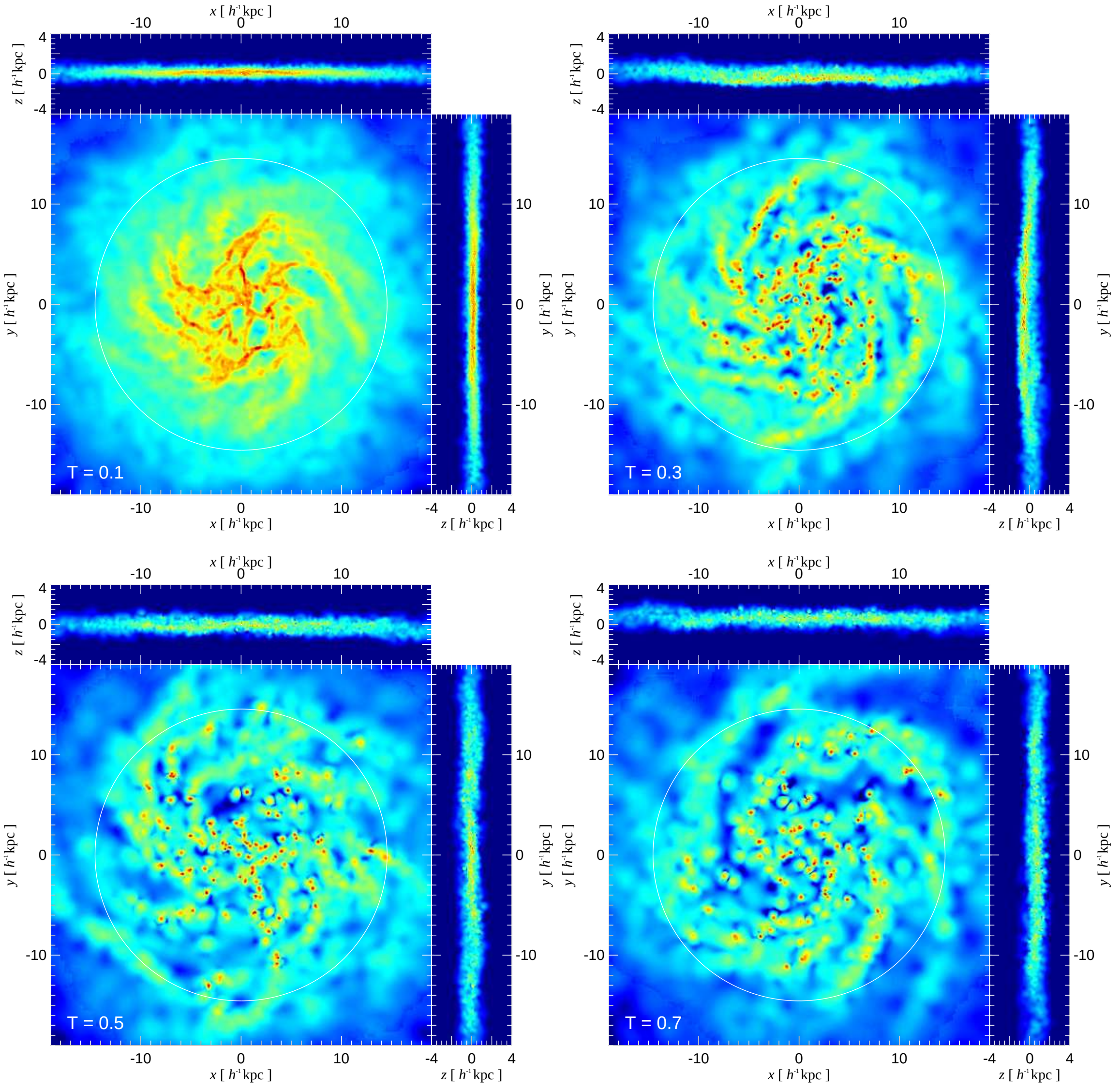


Figure 3.2: Time evolution of the projected gas density (color-coded) in simulation I1, where only cooling and star formation are enabled, but no feedback is considered. The gaseous disk is seen face-on in the xy-projection, and edge-on in the xz- and zy-projections. The circle marks the optical radius $R_{\rm opt} = 3.2 R_{\rm d}$. In each panel, the time T since the start of the simulation is given in units of $9.8 \times 10^8 h^{-1}$yr.

distribution in the Galaxy is several hundred pc, again showing that the warm gaseous layer cannot be supported by thermal pressure alone (Lockman & Gehman 1991).

The extreme thinness of the gaseous disk, and its lack of pressure support, make it

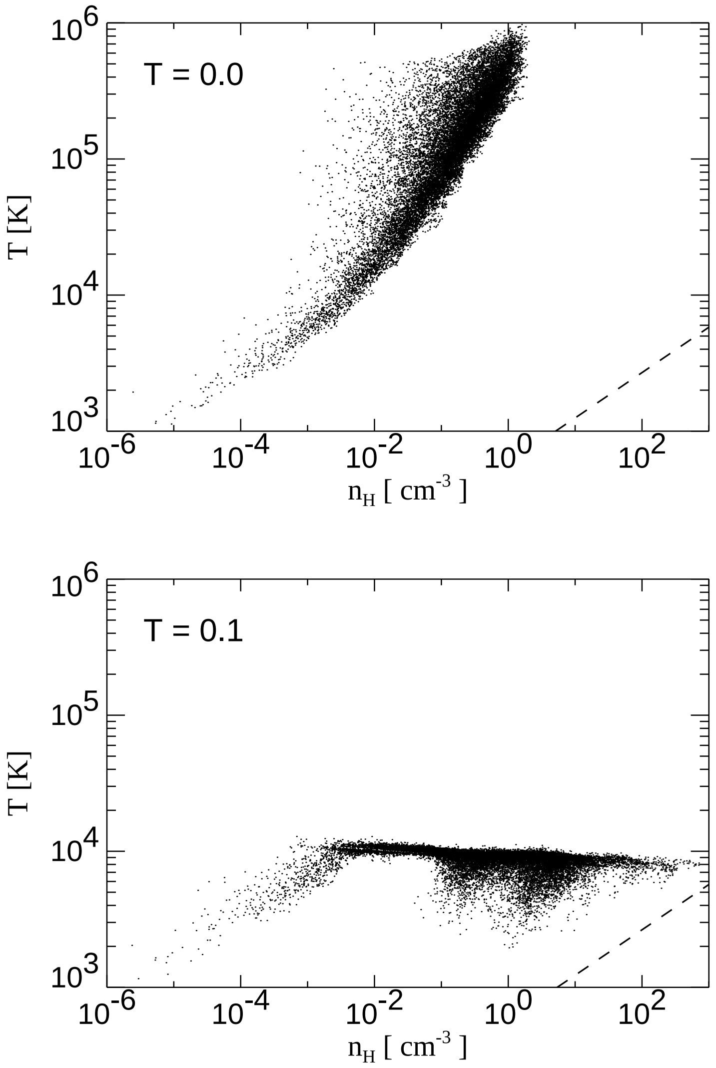


Figure 3.3: Temperature versus gas density of the SPH particles in simulation I1 for the initial time (top), and for the evolved simulation after 0.1 time units (bottom). The density is given in terms of n_{H}, the number density of hydrogen nuclei.

highly susceptible to axisymmetric and local perturbations. As a result, the gas disk soon breaks up into clumps, that sweep up more and more of the gaseous disk to form a scattered distribution of very dense and 'cold' gas lumps. Note that what we have here called 'cold' still has a temperature of $\sim 10^4$ K. In studies of the ISM, such temperatures are usually refered to as 'warm', since the gas in dense molecular clouds is much colder, and is then refered to as the cold component.

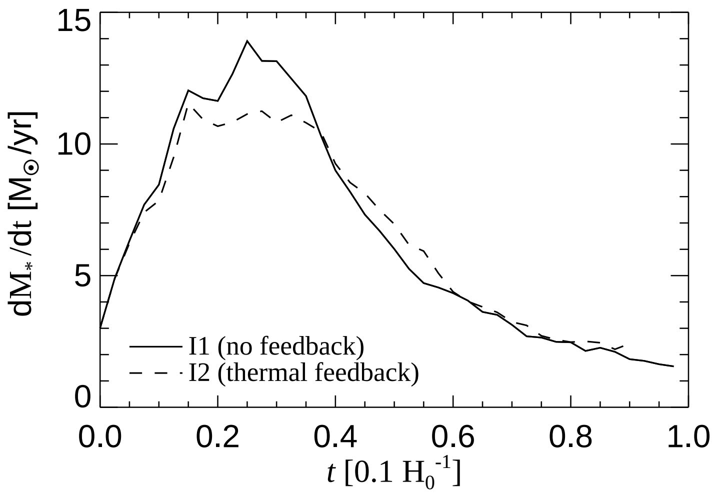


Figure 3.4: Star formation rate of models I1 and I2 as a function of time. Simulation I1 involves only cooling and star formation, while run I2 has also thermal feedback. However, the latter has little effect on the dynamics of the gas, and the evolution of the star formation rate.

The lumps of gas collapse under their self-gravity until they are either stopped by their residual thermal support, or by the gravitational softening length. The latter will often be the case in simulations with a minimum resolution of $0.1\,h^{-1}$kpc or larger. Note that in the real multi-phase interstellar medium very high densities are indeed reached. For example, the hydrogen number density in protostellar clouds may well be above $100\,\mathrm{cm}^{-3}$. However, these densities are many orders of magnitudes larger than the densities we can simulate in the present three-dimensional simulations of whole galaxies. This again illustrates that we cannot resolve the details of the star forming processes in detail. Being far less ambitious, we instead try to model the gross properties of star formation averaged over scales $0.1\,h^{-1}$kpc.

Since we have coupled the star formation rate (SFR) to the gas density, the SFR will grow strongly when the dense gas lumps form. This is seen in Figure 3.4, where we plot the star formation rate as a function of time for this simulation. After the start of the simulation, the star formation rate grows strongly, developing a starburst that has consumed 40% of the gas mass by time $t = 0.4$. Because the gas is rapidly depleted in this burst, the star formation rate declines again after that time. Note that quiescently star forming disk galaxies are believed to have a rather constant star formation rate with time. Clearly, this cannot be reconciled with the present simulation, which therefore does not represent a plausible physical scenario for the gas dynamics in disk galaxies. Perhaps its most severe shortcoming is the absence of a mechanism that can regulate star formation. Such a mechanism is thought to be provided from feedback

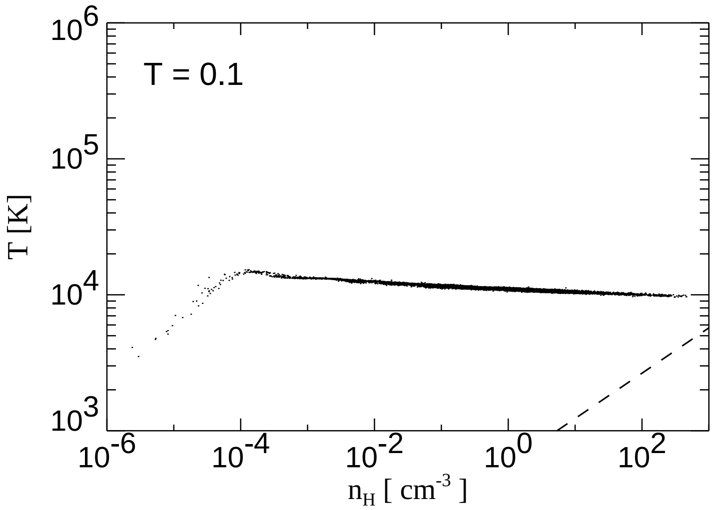


Figure 3.5: Phase diagram of the gas particles of model I2 at time $t = 0.1$. In contrast to simulation I1, we here included thermal feedback by supernovae. Comparing with the lower panel of Figure 3.3, one sees that this energy input quickly re-heats all adiabatically cooled gas back to 10^4 K. However, further heating is prevented by the efficient radiative cooling mechanism that sets in at that temperature. This thermometer confines the gas to a nearly isothermal equation of state.

of the newly formed stars on the surrounding ISM. We now investigate the changes in the gas dynamics and the star formation rate induced by the inclusion of feedback.

3.4.2 Self-regulation of star formation

A main motivation for the inclusion of feedback is to try to prevent the unphysical runaway of the SFR seen in simulation I1. However, if we include only thermal feedback ($\beta = \infty$), the situation hardly changes. This is seen in simulation I2, where we put the feedback energy, 4×10^{48} ergs per formed solar mass of stars, into the ordinary thermal reservoir u of the gas. As seen in the phase diagram of Figure 3.5, this energy can easily heat all adiabatically cooled gas back to a temperature of 10^4 K, but no temperature increase beyond 10^4 K is obtained, because the gas can cool very efficiently in the high-density knots where stars form. This again confirms previous studies; the injected heat energy is radiated away quickly without significantly affecting the dynamics of the gas. Consequently, the star formation rate is not changed very much either (see Figure 3.4).

The ineffectiveness of purely thermal feedback has previously been noted by several authors, all invoking slightly different mechanisms to remedy this problem. Navarro & White (1993) and Mihos & Hernquist (1994b) proposed a kinetic energy feedback, Gerritsen & Icke (1997) invoked UV radiation by young stars, and Yepes et al. (1997)

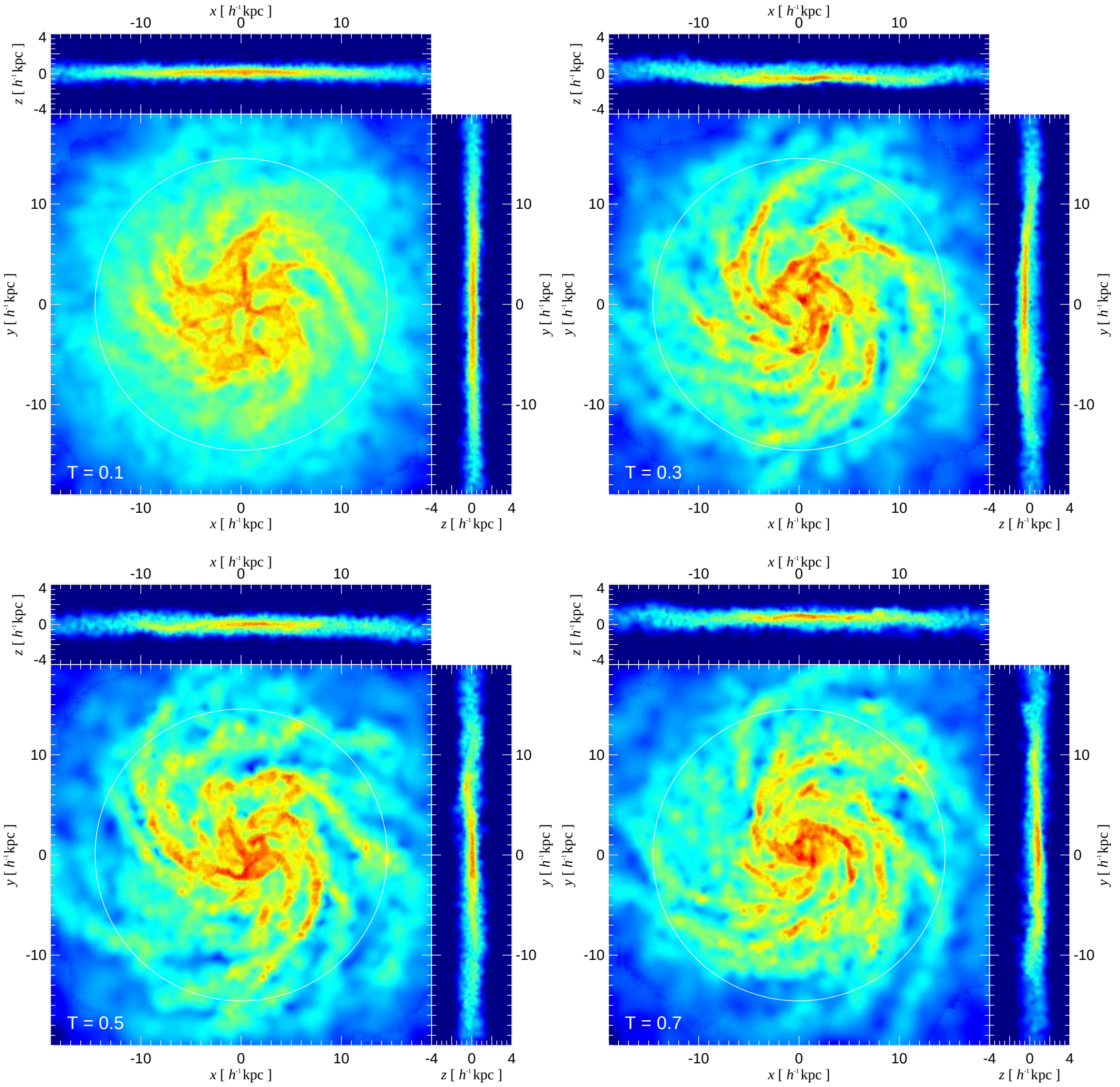


Figure 3.6: Time evolution of the projected gas density (color-coded) in simulation I3, which uses our new feedback model. The gaseous disk is seen face-on in the xy-projection, and edge-on in the xz- and zy-projections. The circle marks the optical radius $R_{\rm opt} = 3.2R_{\rm d}$.

tried to solve the problem by explicitly introducing a two-phase description of the ISM. We will now examine the properties of our new model. As outlined in Section 2, we fix the free parameters with the phenomenological star formation law of Kennicutt (1998), and by assuming a thickness of $z_0 = 0.2\,h^{-1}$kpc for the gaseous disk. This results in

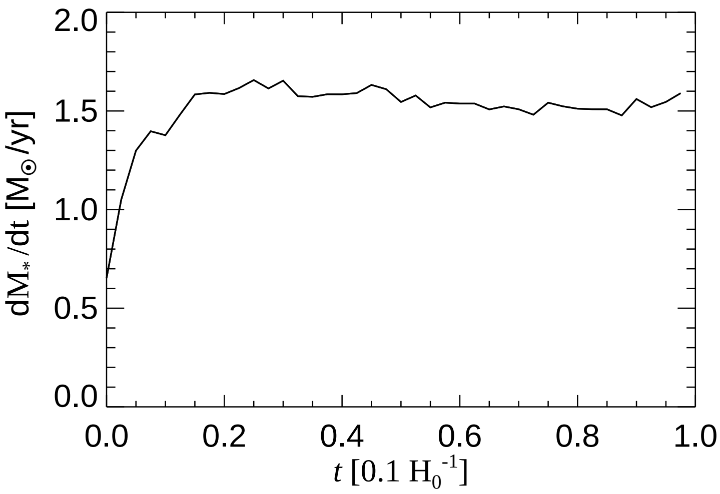


Figure 3.7: Star formation rate of model I3 as a function of time. This simulation uses our new prescription for implementing feedback. As can be seen by the near constancy of the star formation rate, the feedback leads to a quasi-static disk galaxy that quiescently forms stars at an almost constant rate.

$\alpha = 0.0131$ and $\beta = 0.0051\, h^2 \mathrm{M}_\odot^{0.5} \mathrm{yr}^{-1} \mathrm{kpc}^{-1.5}$, and we will employ these choices in all further simulations.

Simulation I3 is the model galaxy I1 yet again, but this time simulated with our feedback model enabled. In Figure 3.6, we show the time evolution of the gas density distribution of this model. The dynamics of the gas in the disk is very different from that shown in Figure 3.2. Instead of breaking up into small dense lumps, the gas now remains in a relatively smooth disk, which shows signs of transient spiral structure. The disk as a whole appears stable and does not show any strong signs of secular evolution.

In Figure 3.7, we show the evolution of the star formation rate as a function of time. After an initial relaxation phase, the SFR remains effectively constant. This demonstrates that the feedback model does indeed establish a self-regulation process for the star formation rate. The nature of this self-regulation process can be understood with the phase diagram of Figure 3.9. Here we plot the total gas pressure (in the form of an 'effective' temperature $T^\star$, defined as $T^\star \equiv \overline{\mu} P/k\rho$) versus the density of the gas particles in simulation I3. Above a density of $\simeq 1\,\mathrm{cm}^{-3}$, the effective equation of state of the gas changes to $P \propto \rho^2$. In this regime, the turbulent pressure dominates the thermal pressure, i.e. the gas is supported by the small-scale motions induced by supernovae explosions.

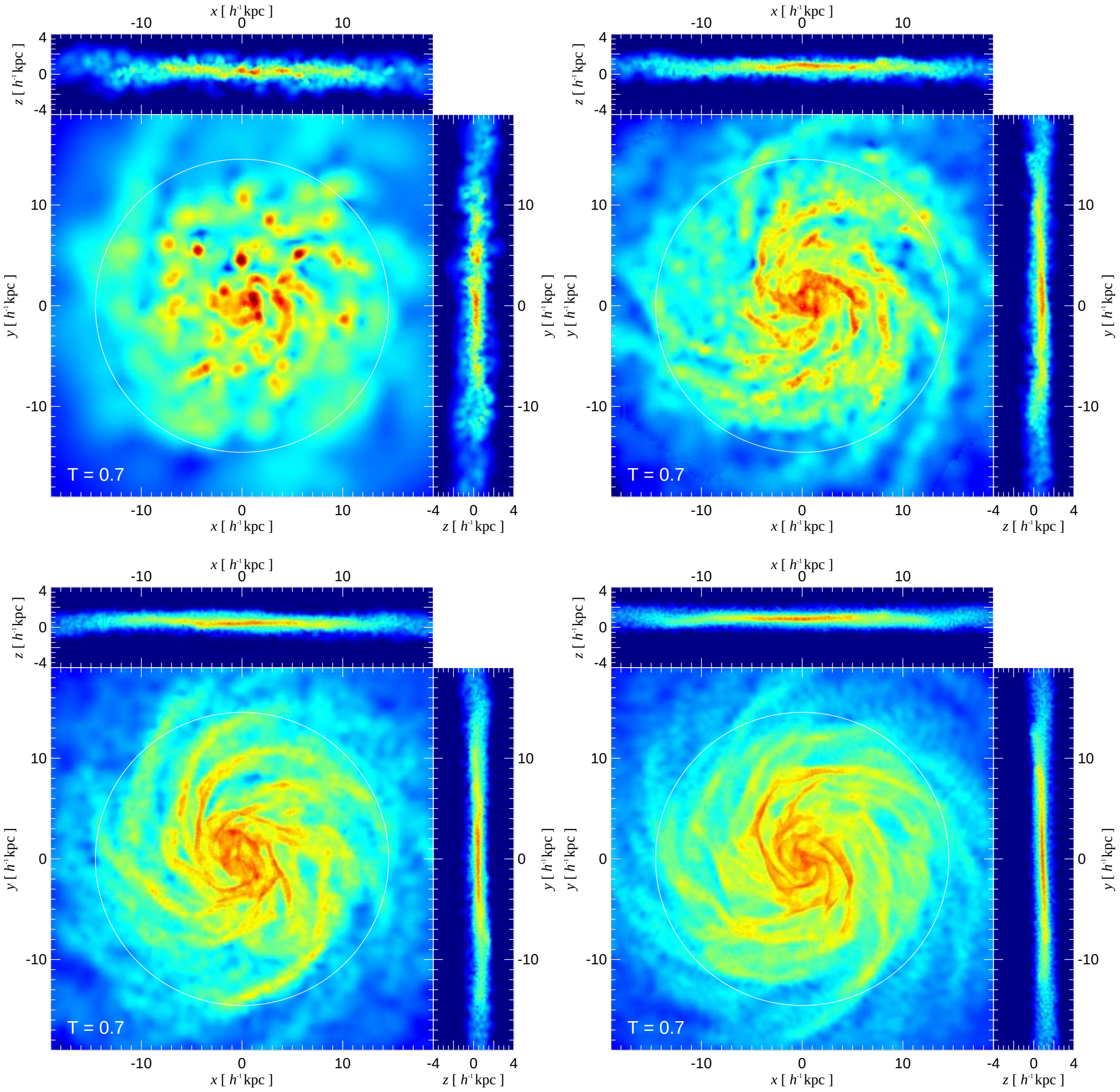


Figure 3.8: Comparison of the gas density distribution in model I3 at different numerical resolution. From the top left panel (I3-a) to the bottom right panel (I3-d), the particle number in each component (gas, disk, halo) increases by a factor of two.

3.4.3 Resolution study

An important aim of our modeling has been to come up with an effective "sub-grid" model for star formation, able to give meaningful quantitative results for the star formation rate, even under conditions of moderate or poor numerical resolution. Since we formulated our feedback scheme without direct involvement of the mass resolution in

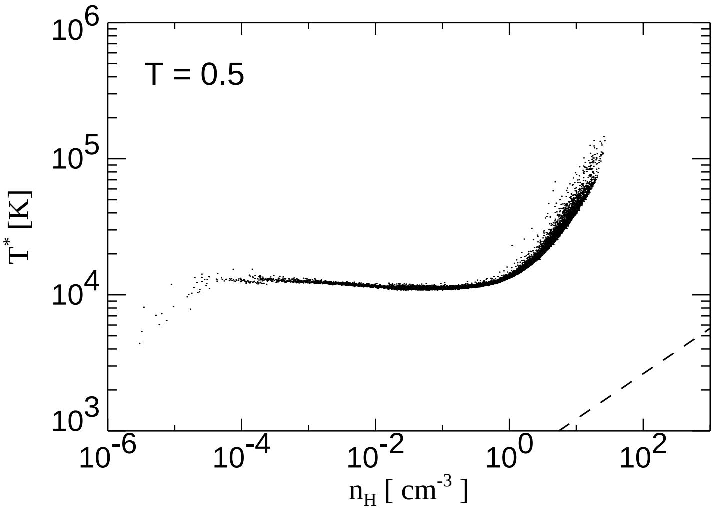


Figure 3.9: Phase diagram of $T^\star \equiv \overline{\mu} P/k\rho$ versus density of simulation I3. Above a density of $\simeq 10^0\,\mathrm{cm}^{-3}$, the energy content in the thermal reservoir q becomes larger than that in the thermal reservoir u. As a result, the feedback due to star formation leads to additional pressure support of the gas, changing the effective equation of state to $P \propto \rho^2$.

the SPH equations, we should be able to obtain converged numerical results, without having to adjust the free parameters α and β.

We test this assertion by repeating simulation I3 at a range of different mass resolutions. We start in realization I3-a with 10000 gas, 10000 disk, and 15000 halo particles. We then successively increase the particle numbers in all three components by a factor of 2, to obtain realizations I3-b, I3-c, and I3-d. The final run I3-d has then 80000 gas, 80000 disk, and 120000 halo particles. Note that I3-b is identical to the simulation I3 described earlier. We have run all four simulations with identical numerical parameters, the *only* difference being the particle number.

In Figure 3.8 we compare the gas surface mass density for the models I3-a to I3-d at an equal point of time. While models I3-c and I3-d show nice spiral patterns, they are less well resolved in I3-b, and broken up in the rather patchy gas distribution of model simulation I3-a. However, the morphology of the gas distribution is quite comparable, at least in models I3-b, I3-c, and I3-d.

In Figure 3.10 we show results for the global star formation rate of disk model I3 in this resolution study. The models I3-b, I3-c, and I3-d agree quite well, indicating that the global SFR has already largely converged in the realization I3-b. The poorly resolved model I3-a leads to an over-prediction of the star formation rate. This seems to be due to the lumpy structure which the gaseous disk develops under the influence of large particle noise. The higher density inside these lumps boosts the global star

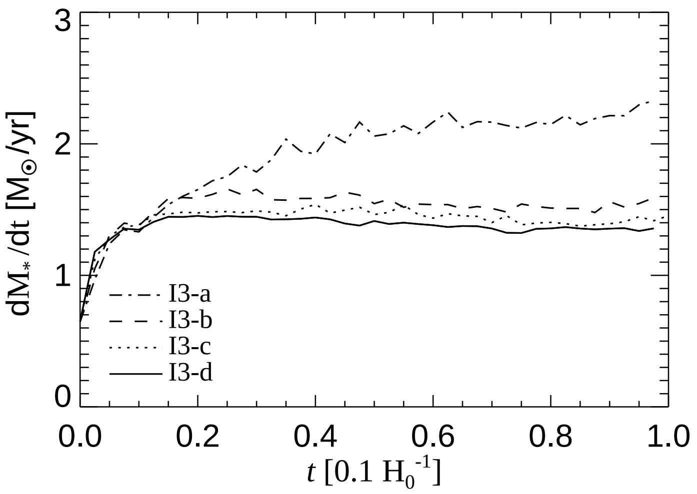


Figure 3.10: Comparison of the star formation rate as a function of time for different numerical resolution. In the models I3-a to I3-d, the particle number successively increases by a factor of 2. It is seen, that the star formation rate has already very well converged for I3-b, while the poor resolution of run I3-a leads to an overestimate of the star formation rate. In this case, the small number of gas particles leads to the development of gaseous lumps in the disk, causing an overestimate of the star formation rate.

formation rate. However, the overall result of this resolution study is very encouraging. It demonstrates that our model has well-posed numerical properties, and that moderate numerical resolution is sufficient to obtain meaningful results for the global properties of the ISM in this model.

3.4.4 Kennicutt's law

We tried to design the feedback scheme such that it reproduces the observed global Schmidt-law of Kennicutt (1998). In particular, we have selected the values of the free parameters α and β a priori such that our models should match quantitatively the observed dependence of star formation rate per unit area on gas surface density as a function of radius in disk galaxies. We now test whether this has worked out.

In Figure 3.11 we show $\Sigma_{\rm SFR}$ versus the azimuthally averaged gas mass density $\Sigma_{\rm gas}$ for the run I3-d. Note that in this simulation, we have not used a cut-off value for the star formation rate. Also shown in this Figure is Kennicutt's law, which we used to compute values for the free parameters α and β of our feedback model. The good agreement between the simulation and this target relation shows that the approximate treatment of the gas as a self-gravitating sheet in hydrostatic equilibrium works surprisingly well.

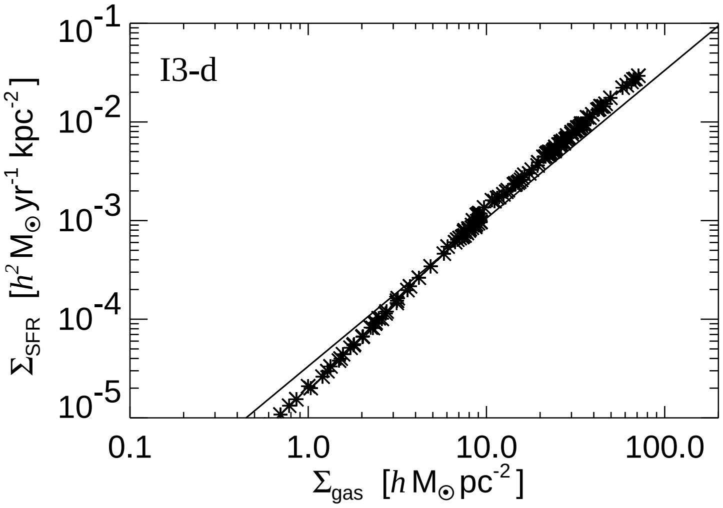


Figure 3.11: Star formation rate per unit area as a function of gas surface density for simulation I3-d. The stars show azimuthally averaged quantities in small circular annuli. The solid line shows the 'target' relation (Kennicutt's phenomenological law) used to determine the free parameters α and β before this simulation was done. Our feedback model thus performs as expected, in particular, it nicely reproduces a $\Sigma_{\rm SFR} \propto \Sigma_{\rm gas}^{1.5}$ law. Note that we haven't included a cut-off in the star formation law in this simulation.

We now consider a number of simulations of isolated galaxies with different structural properties. Model I4 is the model I3 yet again, but this time with the cut-off in the star formation rate included. We also simulate two related versions of this model, differing only in the circular velocity of the halo. In I5, we use $v_{200} = 120\,{\rm km\,s^{-1}}$, and in I6, $v_{200} = 80\,{\rm km\,s^{-1}}$. We further consider two rather extreme galaxies in terms of gas density. In model I7 we chose a low disk mass ($m_{\rm d} = 0.025$), and a low gas content ($f_{\rm gas} = 0.1$) together with a high spin parameter for the halo ($\lambda = 0.1$). This results in a large disk ($R_{\rm d} = 10.75\,h^{-1}{\rm kpc}$) of extremely low gas surface density. Aiming for the opposite extreme, we combined a heavy disk ($m_{\rm d} = 0.1$) with a high gas content ($f_{\rm gas} = 0.4$), and a relatively small spin of the halo ($\lambda = 0.06$). This results in a disk of size $R_{\rm d} = 4.5\,h^{-1}{\rm kpc}$ with very high gas surface density.

In Figure 3.12, we show $\Sigma_{\rm SFR}$ versus the azimuthally averaged gas mass density $\sigma_{\rm gas}$ for the runs I4 to I8. All the models follow the Kennicutt relation very well, including the cut-off at the right place. Note that model I7 practically has no star formation at all, because its gas density remains below the threshold value nearly everywhere.

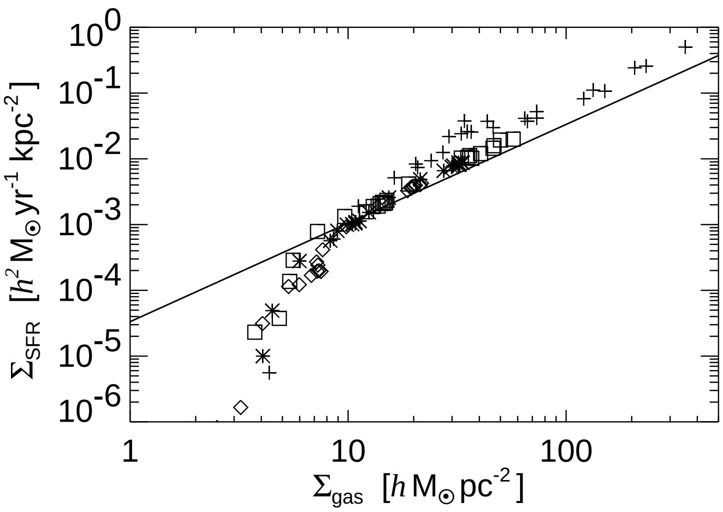


Figure 3.12: Star formation rate per unit area as a function of gas surface density for simulations I4 to I8. In these simulations, an explicit cut-off in the star formation rate was used to reproduce the observed threshold.

3.5 Major mergers

3.5.1 Collision simulations

Even in gas-rich spiral galaxies, the total amount of mass in gas is small compared to the mass of the dark matter and that of collisionless stars. Thus we do not expect a significant modification of the global dynamics of a merging system of galaxies due to dissipative gasdynamics. However, the central starbursts triggered by the interaction might dramatically change the properties of the innermost regions of the galaxies. In the following, we try to address the following questions related to these changes:

1. Does our parameterization of feedback and star formation lead to strong nuclear starbursts as obtained in previous studies?

2. How does the strength and evolution of the starbursts depend on the orbit of the galaxies, and do we find a similar dependence on the presence of a bulge as was found by Mihos & Hernquist (1994c)?

3. What is the effect of dissipation on the structure of the merger remnants? How do the density profiles and the isophotal shapes change?

In order to be able to address these questions in a clean way, we have performed a number of targeted simulations. To this end, we consider four galaxy models, A, B, C, and D. The basic structure of all of them is given by $v_{200} = 120\,\mathrm{s}^{-1}\mathrm{km}$, $c = 5$, $\lambda = 0.05$,

Table 3.1: Summary of runs. The table gives the labeling of our collision simulations, and a short explanation for each of them.

run	disk model	orbit
A1	gas, no bulge	prograde, wide
A2	gas, no bulge	prograde, close
A3	gas, no bulge	inclined
B1	gas, with bulge	prograde, wide
B2	gas, with bulge	prograde, close
B3	gas, with bulge	inclined
C1	collisionless, no bulge	prograde, wide
C2	collisionless, no bulge	prograde, close
C3	collisionless, no bulge	inclined
D1	collisionless, with bulge	prograde, wide
D2	collisionless, with bulge	prograde, close
D3	collisionless, with bulge	inclined

and $m_{\rm d+b} = 0.05$, i.e. they all have the same total mass, and the same total baryonic fraction. In galaxy A, we put all of the baryonic mass into the disk, 20 per cent of it in the form of gas, the rest as collisionless stars. In galaxy B, one quarter of the baryonic mass is put into a bulge, 20 per cent into a gaseous disk, and the rest is used for a collisionless stellar disk. The galaxies C and D correspond exactly to A and B, respectively, except that they contain no gas. Instead, the corresponding mass has also been assigned to collisionless stars. These four galaxies may thus be used to explore consequences of dissipative gas dynamics in merging galaxies with or without a bulge.

For each of these models, we have run three collision simulations on different orbits. These are (1) a very wide prograde encounter with a minimum separation of $R_{\rm kep} = 8.0\,h^{-1}$kpc, (2) a close prograde encounter with $R_{\rm kep} = 0.5\,h^{-1}$kpc, and (3) an inclined encounter with $R_{\rm kep} = 1.0\,h^{-1}$kpc, $\theta_1 = 30^o$, $\phi_1 = 0^o$, $\theta_2 = 60^o$, and $\phi_2 = 90^o$. The angles specify the orientation of the spin vectors of the two disks in ordinary spherical coordinates. In our set-up, the orbital plane coincides with the xy-plane, and the galaxies move such that they reach their Keplerian minimum separation when they cross the x-axes, with their orbital angular momentum pointing along the z-axes. All of the collisions have been started from an initial separation $R_{\rm start} = 240\,h^{-1}$kpc on a parabolic orbit. We list a summary of these runs in Table 3.1. Note that for every gas-dynamical simulation there is one collisionless simulations, with the gas replaced by disk stars. In particular, simulations A1-A3 correspond to C1-C3, and B1-B3 to D1-D3.

3.5.2 Time evolution

In Figure 3.13, we show the gas distribution in run A1 as a typical example for the time evolution of the merger simulations. As has been demonstrated in many previous studies, the encounter between two equal disk galaxies leads to the formation of tidal tails, which are ejected when the galaxies reach orbital pericenter for the first time. Simultaneously, stars from the near side of the encounter are drawn towards the companion, giving rise to bridges between the galaxies as they temporarily separate again. The bridges are destroyed when the galaxies come back together for a second encounter, but the tails survive and grow for a longer time in the relatively quiet regions of the outer potential. Eventually, the disks are scrambled up when the centers of the galaxies finally coalesce to form a spheroidal merger remnant.

The strength of the tidal response depends on the structure of the galaxies, and the orbit of their encounter. A more detailed discussion of this dependence is given elsewhere (Dubinski et al. 1999; Springel & White 1999). We here only study a limited set of simulations to examine the influence of dissipation and star formation.

In the first encounter of the galaxies, the disks develop a bar and are transformed into a pair of open bisymmetric spirals. The response of the gas to the tidal perturbation is markedly different than that of the stellar component. It forms relatively thin arms, and exhibits a phase-lag relative to the stellar arms, such that it loses angular momentum due to gravitational torques. This extraction of angular momentum drives a central inflow of the gas, allowing it to feed a nuclear starburst. Barnes & Hernquist (1996) and Mihos & Hernquist (1996) have studied this mechanism in detail, and our simulations are fully in line with their conclusions. In particular, we also find that the presence of a bulge can prevent an early inflow of the gas by stabilizing the disks against bar formation. This can be most easily seen in the history of the star formation rate, as discussed below.

3.5.3 Evolution of the star formation rate

In Figure 3.14, we show the time evolution of the global star formation rate for those merger simulations that involve gas (A1-A3, B1-B3). Until the first encounter of the disks at $t \sim 1.0$, the approaching galaxies form stars at a constant rate in all models. In the wide prograde encounter of A1, a strong inflow of gas is initiated in this first encounter, leading to a moderate but sustained starburst. Interestingly, simulation B1 shows only a slight elevation of the star formation rate. Hence the early inflow of gas can be almost completely suppressed by the presence of a central stellar bulge.

However, the strength of the early inflow depends also on the orbit of the galaxies, as may be seen by comparing A1 with the narrow prograde encounter of A2. The orbital angular momentum is much smaller in A2/B2 than in A1/B1, and the strong shocks formed in the overlapping disks reduce differences between the early response in models B2 and A2 compared to those showing up between A1 and B1.

When the galaxies come together for a second time, a nuclear starburst develops

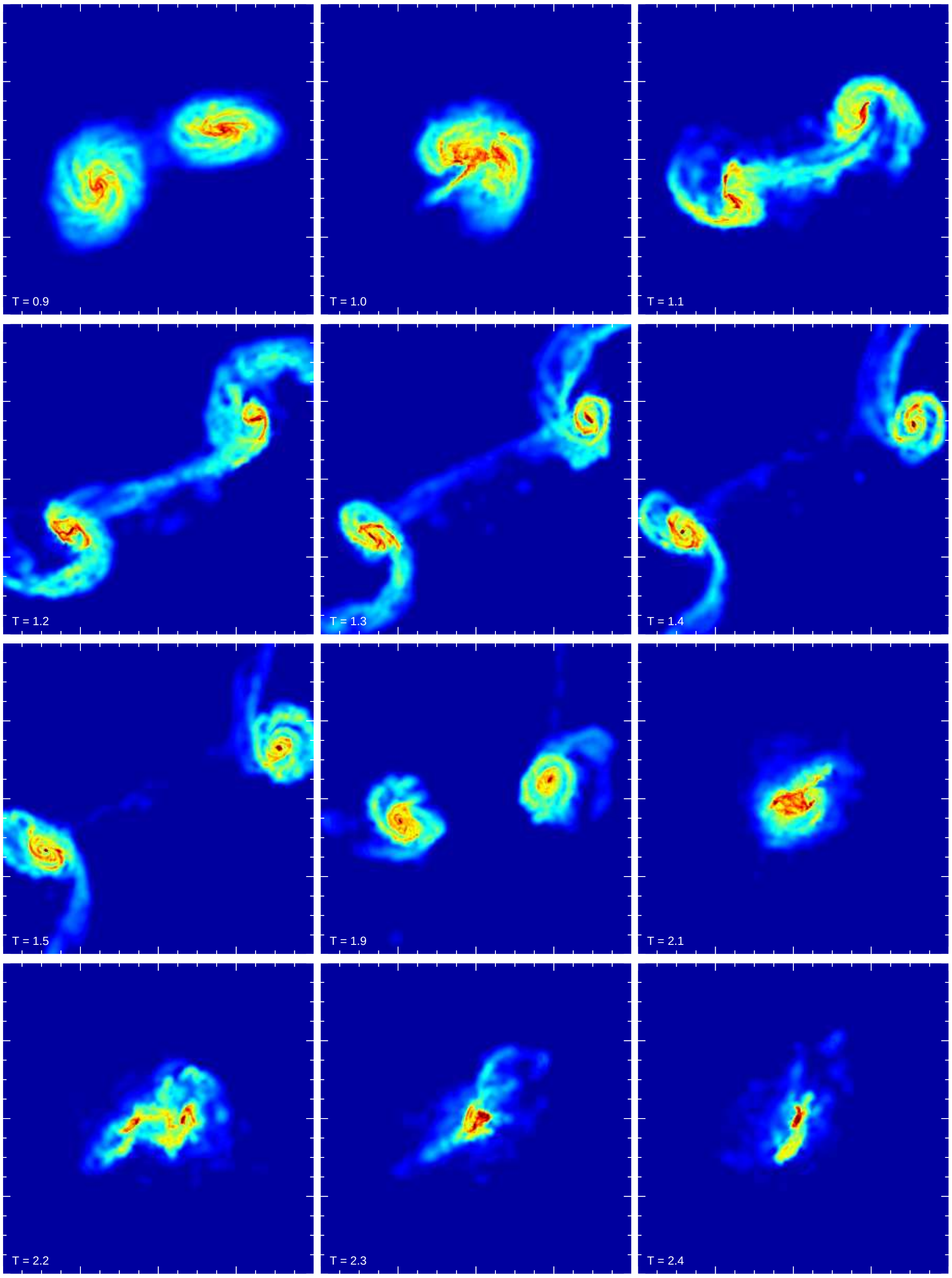


Figure 3.13: Time evolution of the projected gas density in simulation A3, an inclined encounter of two bulge-less disk galaxies. Each panel measures $80\,h^{-1}$kpc on a side, and time T is given in units of $9.8 \times 10^{8} h^{-1}$yr.

in the center of each galaxy. In our simple model, the global star formation rate is increased by not much more than a factor of 10. This is not particularly extreme, and is insufficient to explain the most luminous ULIRGS. However, the absolute values of the star formation densities in the central regions are actually quite large, as we show below.

There is also an interesting double-peak structure in the main starburst at $t \sim 2.0 - 2.5$. This is connected to the dynamics of the final merger. The main starburst is triggered by the second encounter of the galaxies, at a time when they already have lost most of their orbital angular momentum due to dynamical friction. However, the galaxies still spiral about each other for a short time, until their nuclei finally coalesce to form a single remnant. The strongest star formation rate is reached when the nuclei have just merged, or are already so close that the infalling gas 'sees' them as one nucleus.

The difference in the absolute strength of the final starburst in A1 and B1 is largely a consequence of the early influx of gas in A1; here a substantial fraction of the gas is already consumed early on in the encounter, leaving less gas for the starburst at coalescence.

It is also interesting to look at the surface density of the star formation rate in terms of suitably defined averaged values, as it is done in observational studies. In Figure 3.15 we plot the effective density of the star formation rate versus the gas surface density, for three different regimes. These are: (1) whole isolated disks, (2) the centers of isolated disks, (3) the central regions of starbursting disks. In the first case, we have defined the effective star formation density as the average SFR inside the optical radius, while we considered the innermost 3% of the gaseous mass to obtain values for the centers.

By putting these measurements onto a common diagram, see Fig. 3.15, we obtain a global Schmidt-law over a remarkably large dynamic range. The absolute amplitude of the star formation rates in the different regimes is also in very good agreement with the compilation of observational data by Kennicutt (1998, see his Fig. 6).

3.6 Structure of merger remnants

3.6.1 Morphology

Merger remnants are generally triaxial objects, with isophots that are close to perfect ellipses. The triaxiality may be quantified in terms of the axis ratios a and b of the principal axis of the moment of inertia tensor (Barnes 1992; Hernquist 1992, 1993b). To this end, we define the center of the remnant as the position of the particle with the minimum gravitational potential, and we compute the moment of inertia tensor for the most bound half of the particles. In principle, the shape of the remnant can vary if particle groups of increasing binding energy are considered. However, the shape stays roughly constant at least out to the half-mass energy, so we here follow Hernquist (1993b) and characterize the remnants by just considering the inner half of the particles.

In Table 3.2 we list the axis ratios a and b for our merger remnants. Here, $b = \lambda_2/\lambda_1$ and $c = \lambda_3/\lambda_1$, where the eigenvalues λ of the moment of inertia tensor are sorted as

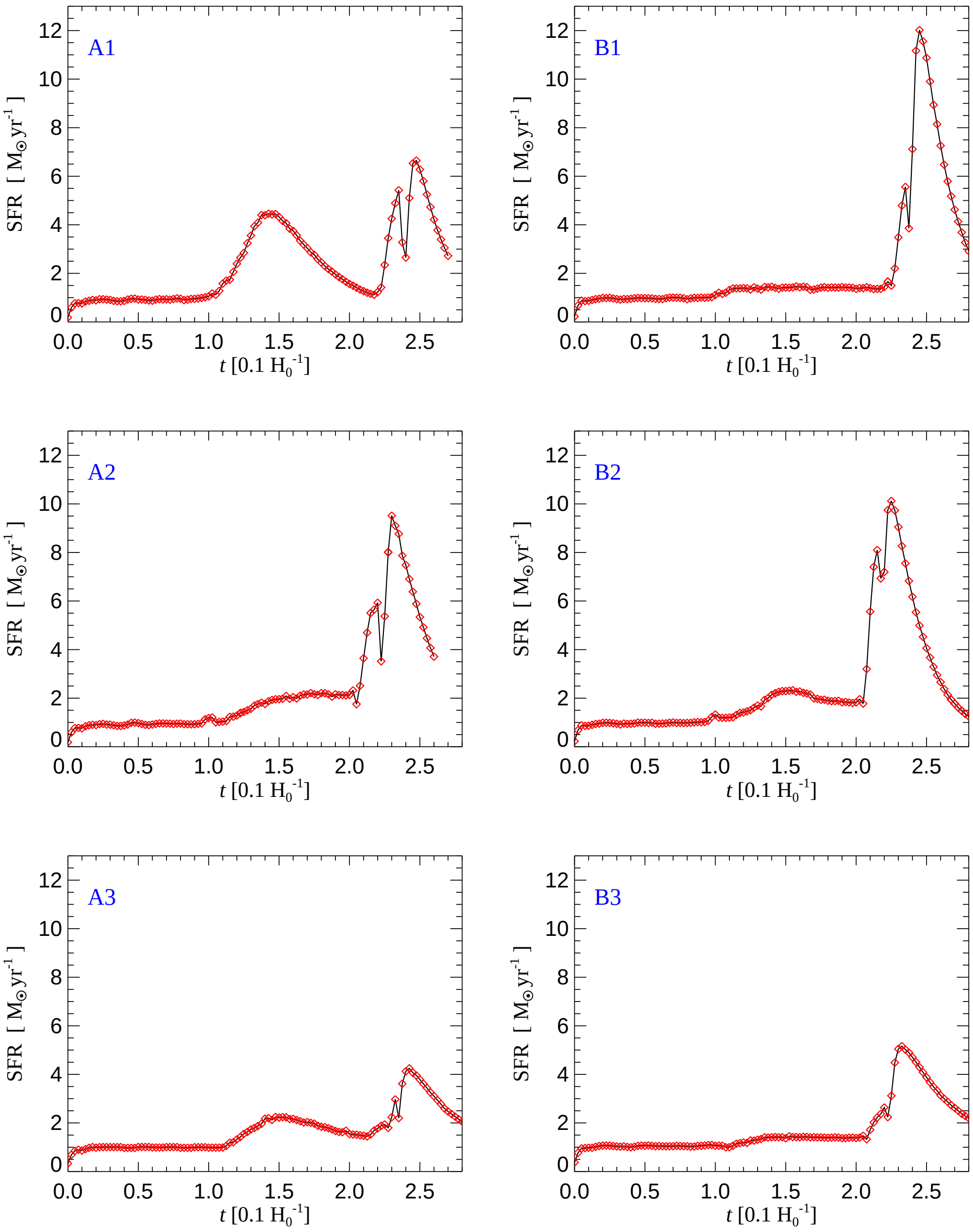


Figure 3.14: Evolution of the global star formation rate in the collision simulations A1-A3, and B1-B3. The runs labeled with a leading 'A' involve disk galaxies that do not have a bulge, while those with 'B' have a central stellar bulge, however, the total mass of gas is equal in both types of galaxies. The simulations labeled with '1' are wide prograde encounters, those with '2' are much narrower prograde collisions, and those with '3' are inclined mergers.

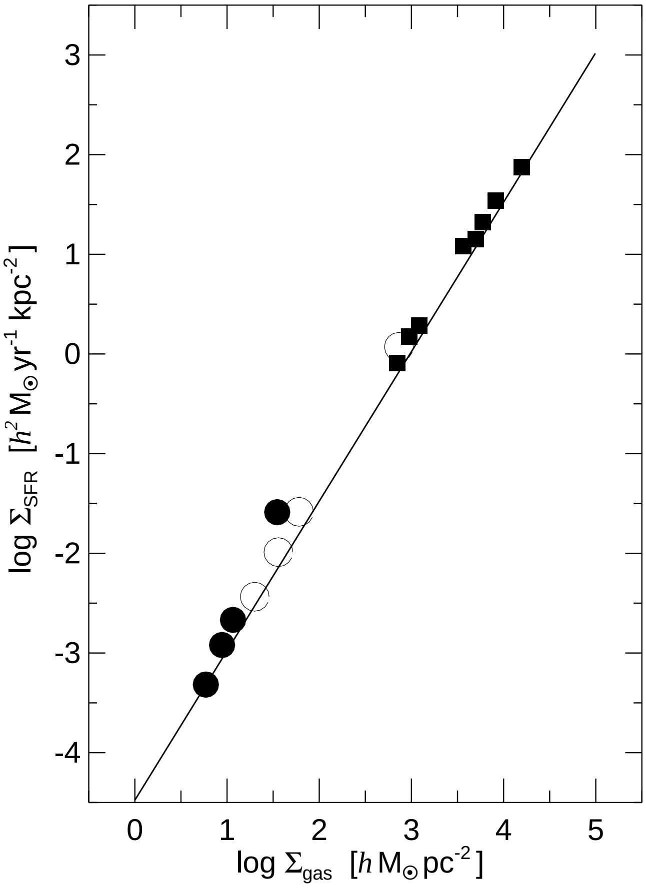


Figure 3.15: Composite Kennicutt law. We here plot the star formation rate per unit area versus the gas surface density for several of our disk models and collision simulations. The filled circles are effective values for the isolated galaxies I4-I8, averaged over the region inside their optical radius. Hollow circles are for the central regions of these galaxies, while the filled squares are for the central regions of the merging galaxies A1, A2, B1, B2 at two times of their evolution. The 'central' values have been obtained for the innermost 3% of the gas mass. The solid line is the 'target' relation we used to compute values for the free parameters of our star formation model. This figure may be directly compared to Fig. 6 of Kennicutt (1998).

$\lambda_1 \geq \lambda_2 \geq \lambda 3$. We also give the triaxiality parameter $T = (1 - b^2)/(1 - c^2)$. For oblate galaxies $T = 0$, while prolate galaxies have $T = 1$ (de Zeeuw & Franx 1991).

Most of the remnants can be classified as oblate spheroids that are close to being axisymmetric. Only the remnants of the collisionless disk-only simulations (C1-C3) show stronger triaxiality. Here, there are also large differences to the corresponding dissipative

Table 3.2: Morphology of merger remnants. b, c, and T give axis ratios and triaxiality of the merger remnants at $t = 2.8$ of our collision simulations.

run	b	c	T
A1	0.94	0.54	0.15
A2	0.96	0.50	0.11
A3	0.91	0.51	0.24
B1	0.90	0.49	0.25
B2	0.91	0.50	0.21
B3	0.97	0.51	0.08
C1	0.68	0.37	0.63
C2	0.74	0.33	0.50
C3	0.88	0.61	0.37
D1	0.87	0.42	0.30
D2	0.86	0.43	0.31
D3	0.95	0.70	0.20

simulations. Apparently, the central density cusps provided either by a bulge or by the newly created stars in the center favour axisymmetry of the remnants.

3.6.2 Density profiles

The coarse-grained phase space density can only decrease in collisionless simulations. It therefore appears that the remnants of collisionless mergers of disk galaxies are not able to reproduce the high central phase space densities of elliptical galaxies, just because the progenitor disks start out with insufficient phase space density (Kormendy 1989). This argument has been held against the merger hypothesis. However, dissipative effects or the inclusion of compact bulges may be invoked to provide a solution to this problem (Hernquist et al. 1993). In previous simulations it was indeed shown that the central influx of gas in mergers, accompanied by star formation, can drastically boost the central phase space density. However, it remains a crucial question, whether the resulting luminous profiles in the center can match the observed profiles of elliptical galaxies.

As expected, the nuclear starbursts in our merger simulations lead to a strong increase of the stellar densities in the merger remnants. In Figure 3.16, we show a comparison of the total luminous mass profile (spherically averaged) of the remnants in simulations B2 and D2. The profile of the dissipative simulation B2 is steeper in the center than those of D2, and it even steepens there compared to the power-law with slope ~ -2.5 found in the main body of the remnant. Such a steep slope is at the upper end of the observed distribution of slopes for early-type galaxies (Gebhardt et al. 1996). Much of the central density in B2 stems from the newly created stars, shown as a dashed

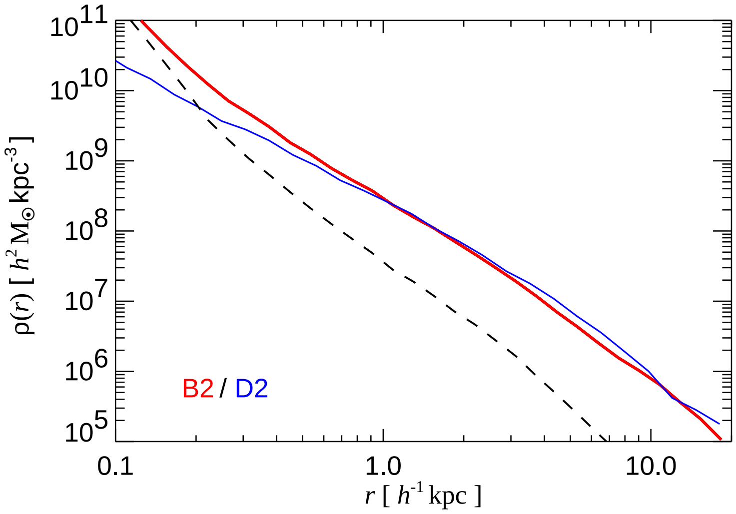


Figure 3.16: Comparison of the spherically averaged, luminous density profiles of the merger remnants of simulations B2/D2. B2 is shown as a thick line, D2 as a thin solid line, and the dashed line gives the newly formed stars in B2. The center of each remnant was defined as the location of the minimum of the gravitational potential.

line in Figure 3.16. However, it is interesting to note that the increase of the central density also affects the other collisionless components in the models. In the two panels of Figure 3.17 we compare the profile of the 'old' collisionless disk and the 'old' bulge, where 'old' refers to material that was already present at the start of the simulations. In B2, both components exhibit a central increase of the density compared to D2 by up to a factor $\sim 4-5$, showing that the dissipational deposition of mass in the center effectively pulls in some of the collisionless material. The merger remnants in our other runs show similar profiles and trends as those of B2/D2, so we omit the corresponding figures here.

We now consider the surface mass density which can be more easily related to observations of surface brightness profiles. Collisionless merger simulations usually lead to $r^{1/4}$-profiles that are consistent with those observed in the majority of early-type galaxies. However, in dissipative simulations, Mihos & Hernquist (1994a) have found a strong concentration of the newly formed stars towards the center, resulting in a luminosity 'spike' in the center, and a surface brightness profile that shows an obvious break between a very steep inner component, and a flatter $r^{1/4}$-profile in the main body of the remnant. Although studies of the cores of ellipticals with HST reveal density cusps in most of them (Gebhardt et al. 1996), the bulk of the elliptical population does not show a steepening in the innermost ~100-500 pc. On the other hand, about 10% of the galaxies in the Nuker sample *do* show such bright cores (Byun et al. 1996). This sug-

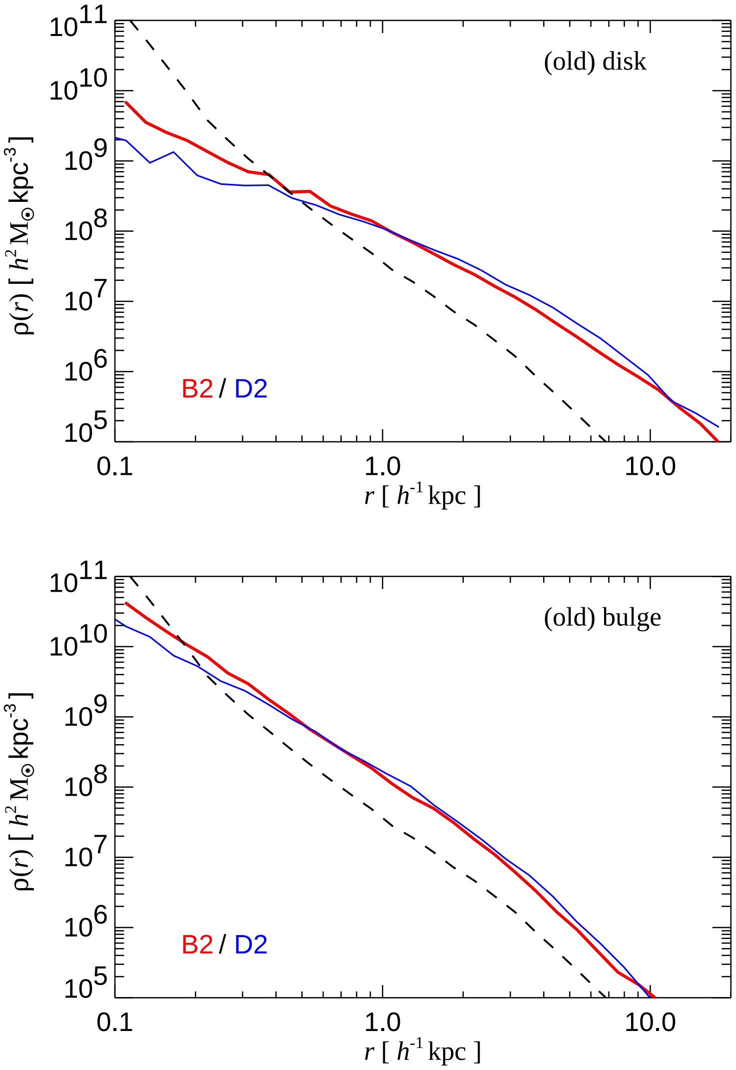


Figure 3.17: The top and bottom panel compare the spherically averaged mass profiles of the old disk and the old bulge components in runs B2/D2 separately. In each case, B2 is shown as a thick line, D2 as a thin solid line, and the dashed line gives the newly formed stars in B2.

gests that gas-rich mergers can evolve into normal elliptical galaxies, while some of them might leave behind a central luminous spike. By studying the luminosity and molecular gas profile of Arp 220, Hibbard & Yun (1999) show that this prototypical ULIRG might indeed be of this kind, while the late-stage remnants of NGC 3921 and NGC 7252 can still be characterized by an $r^{1/4}$-law in their centers.

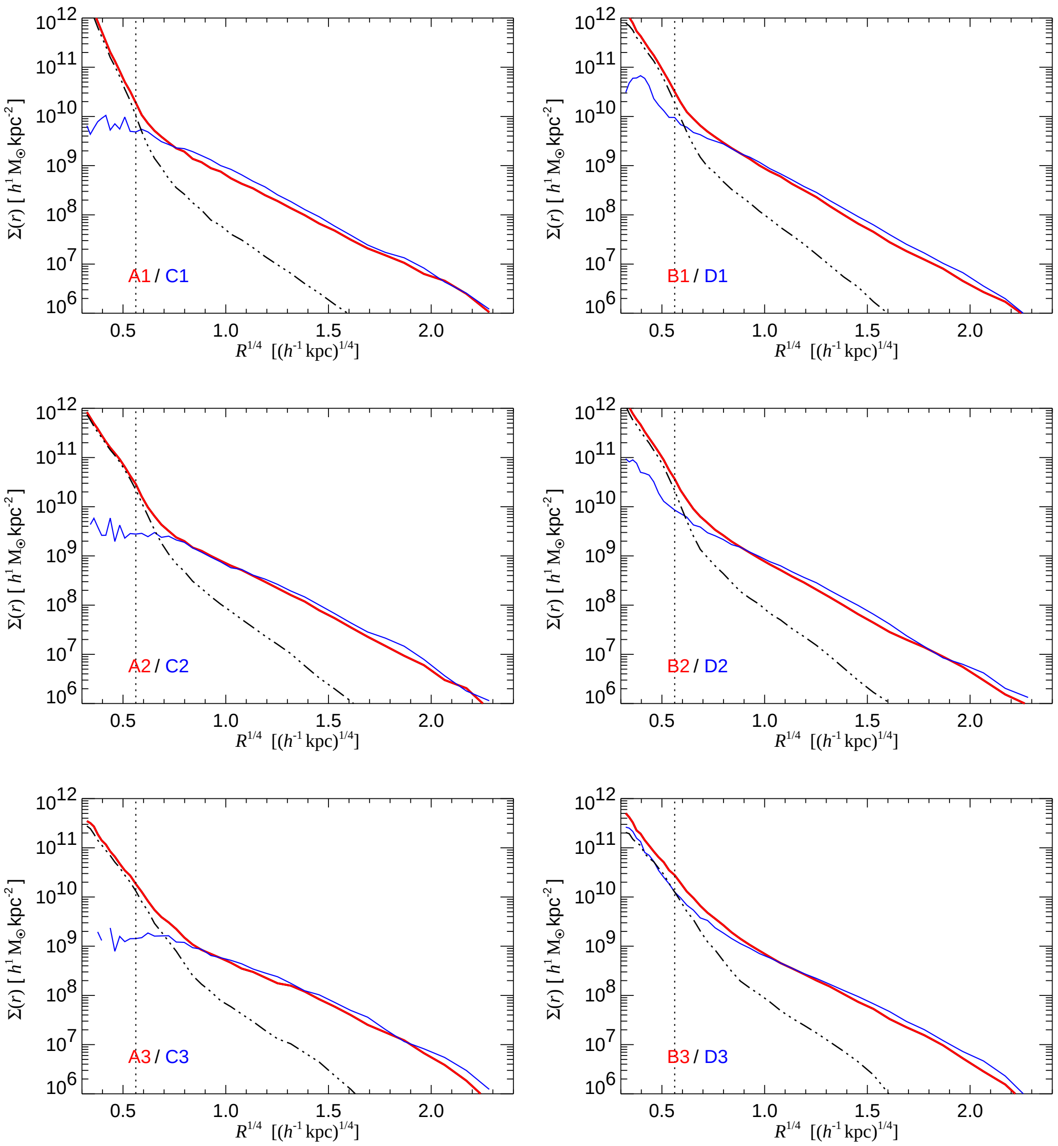


Figure 3.18: Azimuthally averaged surface mass profiles of the remnants of the dissipative simulations A1-A3, B1-B3 (thick), and the dissipationless runs C1-C3, D1-D3 (thin). In each panel, the corresponding luminous profiles are compared with each other, and the dashed line shows the newly formed stars in the dissipative run. The dotted vertical line marks the gravitational softening length of $0.1\,h^{-1}$kpc. Shown here is the xz-projection; other projections are qualitatively very similar.

The surface brightness profiles of our remnants exhibit a central excess of stellar density similar to those seen in the experiments of Mihos & Hernquist (1994a). In Figure 3.18, we compare the azimuthally averaged surface mass densities of our dissipative simulations with those of the corresponding dissipationless run. Unlike the spherically averaged profile, the two-dimensional projection (here along the y-axis) shows the central density excess more clearly. The mass profile of the newly formed stars does not match the remaining collisionless component seamlessly, and this mismatch shows up as a break in the total mass profile. Interestingly, even the collisionless simulations D1-D3 show such a break, although it is weaker than in the runs with dissipation. Apparently, the violent relaxation during coalescence is not effective enough to erase the "memory" of the bulge, and to mix it homogeneously with the rest of the remnant.

It is not yet clear, whether the dense cores due to newly formed stars are a possible problem for the merger hypothesis, because the modeling of star formation and feedback adopted in this study remains highly uncertain. It is however noteworthy that we obtain similar results to Mihos & Hernquist (1994a) using different techniques for our simulations. This suggests that the result is robust, and that any model based on an equally simplistic star-formation/feedback scheme is likely to encounter the same problem. However, there are a number of processes that might prevent the formation of luminous spikes in the nuclear starbursts of gas-rich mergers. If the physics of star formation and feedback is very different in a strong starbursts than captured by our modeling, more diffuse starbursts may be produced, leading to a smoother distribution of the newly created stars. For example, the central starburst might be powerful enough to drive a mass-loaded superwind, carrying a good fraction of the cold gas out of the center. Of course, there is also the possibility that differences in the IMF, or the dust obscuration, between the young starburst component and the old stars change the slopes of the luminous components such that they are consistent with the observed "seamless" profiles of most early-type galaxies.

3.6.3 Isophotal shapes

Isophotes of elliptical galaxies deviate from perfect ellipses (Bender et al. 1987), and it has been proposed that the systematics of these deviations provide important clues for understanding the formation history of ellipticalss. The deviations of isophotes from perfect ellipticity can be quantified by expanding the residuals from an elliptical fit in a Fourier series (Bender & Möllenhoff 1987)

$$\Delta r = \sum_k \left[a_k \cos(k\phi) + b_k \sin(k\phi) \right], \tag{3.27}$$

where Δr are the linear deviations as a function of the azimuthal angle ϕ, measured from the major axes of the ellipse. The first non-trivial coefficient that measures a distortion symmetric to the principal axes of the ellipse is a_4. Negative values of a_4 imply "boxy" ellipses, while positive values lead to a more "disky" shape.

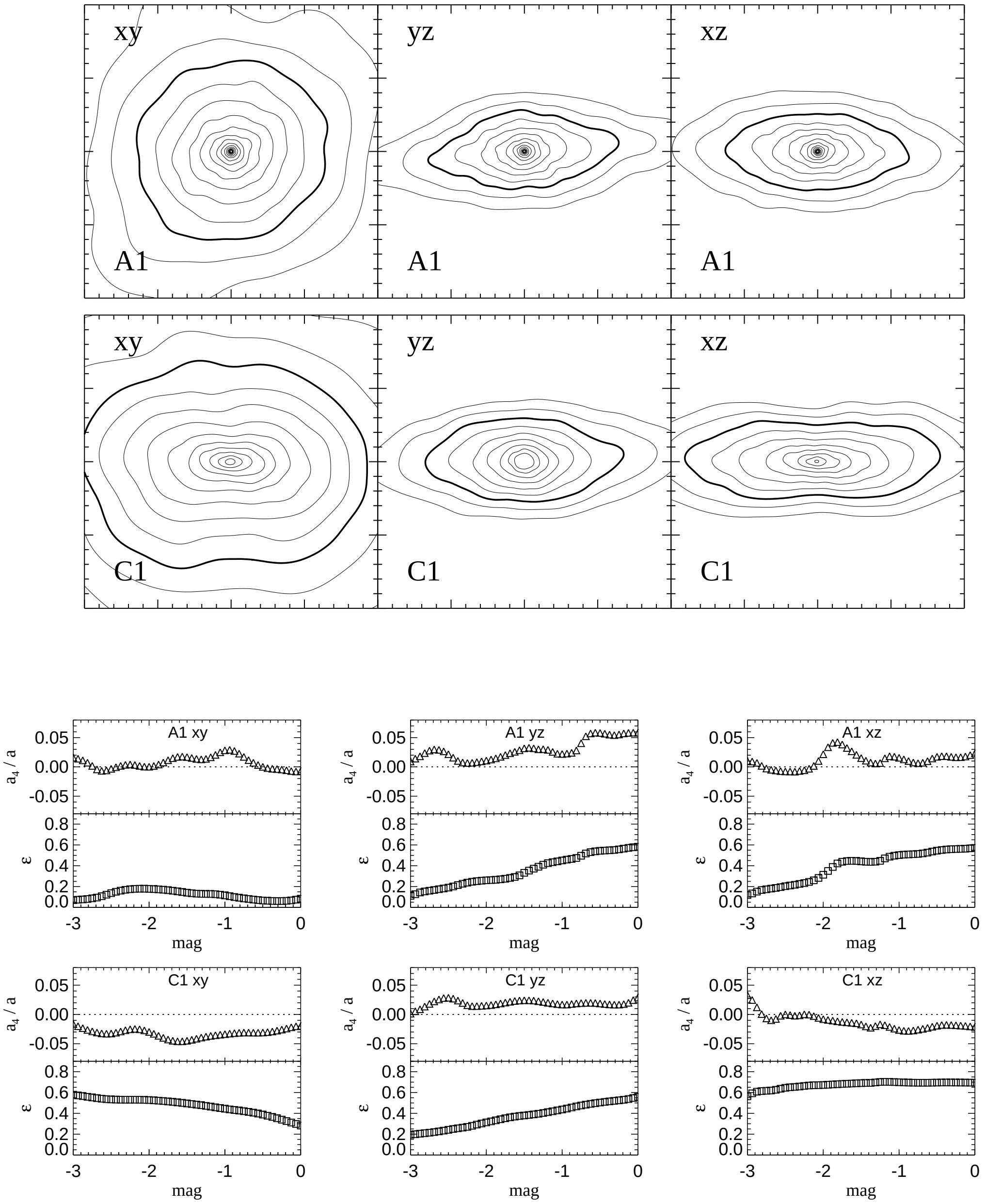


Figure 3.19: Comparison of the isophotal shapes of two merger remnants. The contour plots on the top show the isophots of remnants A1 and C1, projected along their principal axes. Each small panel is $20\,h^{-1}$kpc on a side. The thick isophot contains half the luminosity (stellar mass), and the isophots are spaced 0.5 mag apart in surface brightness. The small panels below show measurements of the diskyness and ellipticity of the isophots for these projections. The isophots are labeled by their magnitude relative to the half-light isophot.

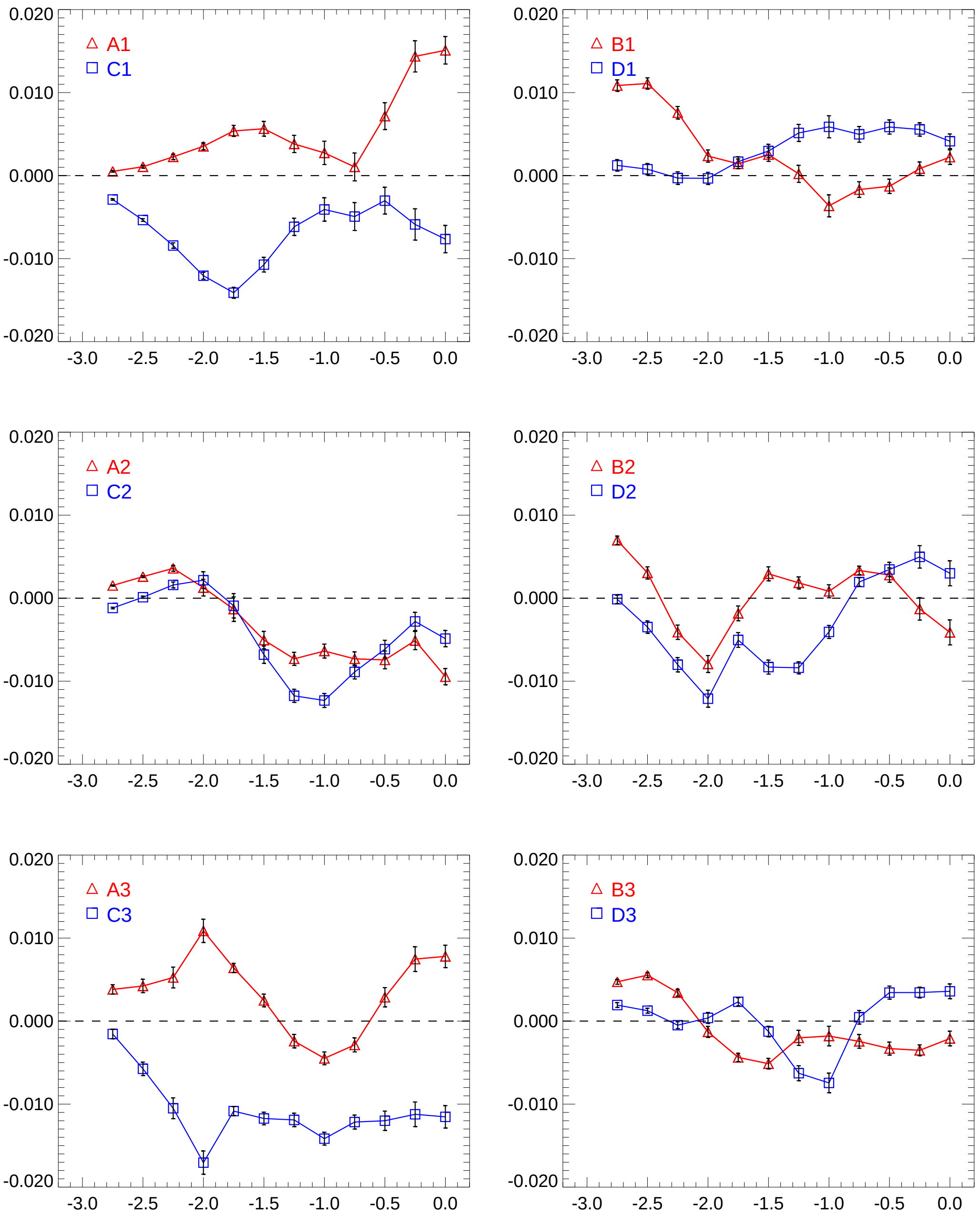


Figure 3.20: Comparison of the isophotal shapes of our merger remnants, averaged over 100 random lines of sight. Each panel gives the average diskyness as a function of the magnitude of the isophot (zero magnitude is assigned to the half-light isophot). The error bars indicate the 1σ uncertainty due to the finite number of projections.

Observationally, rotation seems to be dynamically less important in boxy than in disky ellipticals, and this correlation has been used for a proposed revision of the Hubble classification system by Kormendy & Bender (1996). These authors also provide evidence for a possible dichotomy between (1) normal- and low-luminosity ellipticals that are coreless and disky, and (2) giant ellipticals with cuspy cores and boxy-distorted isophotes. It has also been speculated that dissipation plays a key role in producing disky isophotes, and that the path from boxy to disky ellipticals may thus indicate mergers with increasing importance of dissipation (Bender et al. 1992, 1993). Recent simulations by Bekki & Shioya (1997) seem to support this notion, and we will here examine this question for our merger remnants.

However, collisionless simulations of merging galaxies have repeatedly been shown to produce both boxy and disky isophotes (Governato et al. 1993; Heyl et al. 1994; Lima-Neto & Combes 1995). Furthermore, the isophotal shape strongly depends on the projection, even to the extent that the same merger remnant can exhibit either boxy or disky isophotes when viewed from different directions. This certainly complicates the interpretation of isophotal shapes. Nevertheless, there seems to be a slight surplus of boxy isophotes in large samples of collisionless remnants, and highly flattened remnants also appear to be boxy (Heyl et al. 1994; Lima-Neto & Combes 1995).

In Figure 3.19 we compare the isophotes of simulations A1 and C1, projected along the three principal axes of the remnants. To obtain sufficiently smooth two-dimensional mass density fields, we have employed a Gaussian smoothing kernel of varying smoothing length, i.e. the outer contours are based on more heavily smoothed fields. This is done to beat down particle noise, which strongly increases with distance from the center. The thick isophot in each contour plot contains half of the 'light' (mass), and the spacing of the isophots is 0.5 mag. We have also fitted ellipses to the isophotes inside the half-light radius, and evaluated a_4/a and the ellipticity $\epsilon = 1 - b/a$ as a function of the magnitude of the isophot. These results are shown in the lower panels of Figure 3.19.

The projections along the major and intermediate axes show the strong oblateness of the remnants in these idealized, exactly prograde collisions. Also nicely visible is the central increase of the density in A1 compared to C1. Even by eye one can see that the isophotes of the dissipative simulation A1 seem to be more disky on average than those of C1. This is borne out by the measurements of a_4, which tend to give smaller values for C1. Especially in the xy-projection, the isophotes of C1 are very boxy, while those of A1 appear to be slightly disky and they are also much rounder.

However, Figure 3.19 also demonstrates that the measurement of a_4/a for a single projection is rather noisy, primarily because of the limited number of particles in our remnants, which is very much smaller than the number of luminous stars in real galaxies. But we also see that the same remnant can be both boxy or disky when viewed from different directions (see e.g. the xy/yz projections of C1). The noise can be reduced somewhat if a_4/a is averaged for many lines of sight, which is also warranted due to the projection-dependence of disky-/boxyness. Such an averaging procedure is not possible for real galaxies, but it should give the expected mean ellipticities for a sample of equal objects, observed from random directions. We thus compute elliptical fits for a set of

100 random lines of sight for every remnant. For each remnant, we then average the measurements of a_4/a for the isophots of the same magnitude relative to the half-light isophote.

In Figure 3.20 we compare these measurements for each pair of corresponding dissipative/collisionless simulations. In the bulgeless simulations (A1-A3, C1-C3), the gas-dynamical remnants are clearly much more disky than the corresponding collisionless remnants. This can be understood as a consequence of the strong central increase in density due to the nuclear starburst. As Barnes & Hernquist (1996) have shown, the deep potential well in the center destabilizes box orbits passing close to the center, and turns them e.g. into tube orbits. As a result the mix of orbital families is changed such that the isophots of the remnant turn disky.

This picture is also consistent with the much weaker trend to diskyness seen in the simulations B1-B3, which feature a massive bulge. Here the difference in the central densities between the B/D-pairs is much smaller than in the A/C-pairs, and as a consequence the relative populations of their orbital families remain more similar.

Note that in all cases the dissipative simulations are more disky than the collisionless ones close to the center, i.e. dissipation clearly favours diskyness. However, even with this strong correlation between diskyness and dissipation, it should be kept in mind that it will hold only in the mean for a population of observed merger remnants. A single projection of a remnant might look boxy, even if it is disky in the mean when viewed from many directions.

3.7 Discussion

In this chapter, we have studied a simple model to implement a self-regulating star-formation/feedback-loop in hydrodynamical simulations of disk galaxies. We have allowed the gas to cool radiatively, and to form collisionless stars at a rate determined by the local gas density and the local dynamical time. If only these processes are included, the ISM gives rise to a runaway of the star formation rate. We therefore invoked a feedback process that regulates the star formation rate, and we assumed that the main source of energy for this mechanism stems from kinetic energy of exploding supernovae.

However, since we clearly lack the resolution to follow the physics of the turbulent ISM in three dimensions, we tried to formulate a feedback model that is based on the notion of an effective equation of state. Ideally, such a "sub-grid" model should correctly describe the gross behaviour of the ISM on large scales, while the details on small scales need not to be resolved. A working model that meets these requirements would be extremely valuable; it could be used in simulations of galaxy formation in large cosmological volumes. The catch of course is that currently the "sub-grid" physics is not well enough understood to unambiguously establish the validity of any given "effective" model. While it is quite plausible that supernovae will stir up the ISM in some way, the physical processes that constitute feedback are very difficult to model in detail. We have described this poorly understood kinematics with a fiducial reservoir

q, introduced to account for the kinetic energy in turbulent gas motion on unresolved scales. Assuming that q provides a turbulent pressure for the gas, we arrived at the desired self-regulation mechanism; a local increase in gas density leads to stronger star formation, thereby giving rise to an extra turbulent pressure which works against the local density enhancement.

We have shown that our model is able to describe quiescently star forming disk galaxies. The star formation rates quantitatively match those observed in late-type galaxies, and the model galaxies are stable, and show no signs of strong secular evolution. We formulated the feedback model in a numerically clean way, which does not resort explicitly to the particle formalism of SPH. This approach offers favourable numerical properties. In particular, we have demonstrated that our results are quite insensitive to the mass resolution of our simulations.

In simulations of colliding galaxies we find nuclear inflows of gas and the triggering of nuclear starbursts quite like it was found in previous work. The star formation densities reached in the centers are quantitatively similar to those observed in the central regions of starburst systems. We also confirm that the presence of a bulge can delay the central influx of gas. This is related to the stabilizing effect of the bulge on the disk, where the bulge can prevent the formation of a strong bar in the disk during the first encounter of the galaxies.

We also examined a small number of collisions of galaxies with and without a bulge, and with different orbital configurations. To highlight differences induced by the gas dynamics in the structure of the remnants, we have run a corresponding collisionless simulation for each dissipative merger. Similar to Mihos & Hernquist (1994a), we find that the nuclear starbursts occurring in gas-rich mergers can boost the central star formation rate considerably, even leading to the formation of a dense stellar core, with an accompanying luminosity spike. While there are systems with such central spikes, features of this kind are not observed in the majority of the elliptical population. It thus seems that dissipation may be too effective in our models.

By comparing the isophotal shapes of our dissipative merger remnants with those of corresponding collisionless simulations, we confirm a common speculation that dissipation gives rise to disky isophots. However, it is important to note that such a correlation can only be expected to hold in the mean for a large sample of galaxies. The dependence of isophotal shape on projection is so strong, that the same remnant can exhibit both boxy and disky isophots when viewed from different directions.

Of course, the description of the ISM adopted in this work needs to be treated with caution. It does give reasonable results for isolated disk galaxies, and it was specifically designed to be compatible with the observed global Schmidt-law for the star formation rate. However, it is certainly questionable whether the physical conditions in starbursting nuclei are similar to those in quiescently star-forming disk galaxies, and even if they were, it is not clear whether our attempt to model these processes in a simplistic way gives meaningful results in both regimes. Nevertheless, the agreement between our results with previous work suggests that the main results of these studies, such as the rates of gas inflow, or the development of nuclear starbursts, are robust. While this in general supports the notion that ULIRGS are intimately connected to mergers and interactions, there also remain problems, if one assumes that normal ellipticals are formed by mergers of gas-rich spirals. The most imminent one seems to arise from the formation of luminous spikes at the centers of the merger remnants. They appear difficult to reconcile with the surface brightness profiles of most early-type galaxies.

This might well point to a major shortcoming of the present modeling. Our simplified sub-grid model largely glosses over the shock physics, which is particularly important to drive outflows. For example, OB-associations with an accompanying starburst may generate super bubbles in the ISM and trigger outflows to the surrounding IGM, or even blow away the gas of dwarf galaxies (Mac Low & Ferrara 1999) entirely – it is quite clear that our simple parameterization will not be able to properly model these effects. This highlights that there are at least two important aspects of feedback. One is concerned with the effects of star formation on the IGM and on the hot gaseous phase in the galactic halo, while the other deals with the self-regulation of star formation within the gaseous disk itself. Our model addresses the second of these problems, but it doesn't describe the first one adequately.

Despite these caveats, we think that the general approach offered by simplified sub-grid models is very useful in studies of galaxy formation, and it will likely remain so for quite some time, since a detailed description of the ISM on the scale of whole galaxies is computationally not yet feasible. In the future, high-resolution studies of small patches of the ISM might well give rise to much more accurate sub-grid models, which will then be invaluable tools to model the formation of the luminous component of the Universe on much larger scales.

–This is time, naked time, it comes slowly into existence, it keeps you waiting, and when it comes you are disgusted because you realize that it's been there already for a long time.

Jean-Paul Sartre, Nausea

4

Populating a cluster of galaxies

Abstract

We study the formation of cluster galaxies by combining high-resolution N-body simulations with semi-analytic modeling. To this end, we have simulated the formation of a rich cluster of galaxies in a flat low-density (ΛCDM) cosmology. We simulated the same cluster three times, at mass resolutions of 6.9×10^9, 1.4×10^9 and $2.4\times10^8\,h^{-1}\mathrm{M}_\odot$, allowing direct tests of resolution dependence. The three clusters are resolved with 0.12, 0.61 and 3.5 million particles within the virial regions, respectively. By storing 51 simulation outputs between redshifts $z=20$ and $z=0$, we are able to analyse the merging history of the dark matter in substantial detail. We introduce new algorithms to identify and track substructure within haloes, and we generalize the semi-analytic methodology recently developed by Kauffmann et al. (1999a) to directly include this information for building up the galaxy population of the cluster. We demonstrate that the inclusion of "subhaloes" leads to a substantial improvement of the results of the semi-analytic recipes. In particular, the new scheme endows the cluster luminosity function with a pronounced knee, and it makes it well fit by a Schechter function. It also avoids an excessive brightness of the first ranked cluster galaxies. We show that the main cause for this improvement lies in the ability of the subhalo-scheme to provide more accurate estimates for the merging rate inside the cluster. Our semi-analytic model simultaneously produces a cluster mass-to-light ratio close to observational values and it fits the velocity-based Tully-Fisher relation of spirals in the field. We can directly demonstrate the presence of a morphology-density relation, spatially resolved within the cluster. Towards the center of the cluster, the morphological mix of galaxies becomes increasingly dominated by ellipticals, while in the outskirts of the cluster and in the field, most galaxies are spirals. Because the morphology of our model galaxies is determined almost exclusively by the merging history of the galaxies, this shows that hierarchical clustering naturally produces a strong morphology-density relation. The star formation histories of the cluster galaxies show that already half

of their stars have formed beyond redshift $z \simeq 4$, while the galaxies in the field form the bulk of their stars substantially later; at $z \simeq 2$ half of their stars are in place.

4.1 Introduction

Based on the pioneering work of White & Rees (1978), the last two decades have witnessed substantial progress towards an understanding of galaxy formation within the framework of a universe dominated by cold dark matter (CDM). For an appropriate choice of the cosmological parameters, the CDM theory provides a remarkably successful description of large-scale structure formation, and it is in good agreement with a large variety of observational data. Much of this progress has been achieved by detailed analytical and numerical studies of the collisionless dynamics of the dark matter. As a result, this part of cosmic evolution is now quite well understood. However, the actual formation of the luminous parts of galaxies within CDM universes involves many complex physical processes in addition to gravity, like radiative cooling of gas, or star formation and its regulation mechanisms. The theoretical modeling of important aspects of these processes is still highly uncertain.

Not surprisingly, this lack of precise specifications for the treatment of the relevant physics has also hampered direct numerical studies of galaxy formation. In addition, such studies are confronted with a huge range of scales spanned by the relevant physics. Hydrodynamical simulations therefore still fail to reproduce basic properties of galaxies in detail, although more recent work is beginning to achieve impressive successes (e.g. Steinmetz & Müller 1995; Pearce et al. 1999).

However, much of the current understanding of galaxy formation has been learned from 'semi-analytic' models of galaxy formation, as laid out originally by White & Frenk (1991), Cole (1991), and Lacey & Silk (1991). In these models, the complicated physics involved in galaxy formation is approximated in terms of a simplified yet physical treatment of the most important processes involved. These processes include the formation and merging history of dark matter haloes, the shock heating and virialization of gas within these haloes, the radiative cooling of gas and its settling to a gaseous disk, star formation, and the resulting feedback processes by supernovae and stellar winds. At the expense of some uncertainty introduced by the simplifying assumptions, the semi-analytic techniques can access a much larger dynamic range than present numerical simulations, and they allow a fast exploration of parameter space and of the consequences of possible changes in the model prescriptions.

Over the last few years, a number of groups have studied the properties of semi-analytical models, and they tested their predictions against observational data (Lacey et al. 1993; Cole et al. 1994; Kauffmann et al. 1994; Heyl et al. 1995; Baugh et al. 1996a,b; Kauffmann 1995a,b, 1996a,b; Kauffmann & Charlot 1998; Baugh et al. 1998; Somerville & Primack 1998; Benson et al. 1999). Population synthesis models as introduced by Kauffmann et al. (1993) have allowed detailed photometric comparisons with observations, including studies of the evolution of galaxy populations to high redshift.

With a small number of free parameters, semi-analytic models have been quite successful in reproducing a large range of galaxy properties, such as luminosity functions, the Tully-Fisher relation, number counts, color distributions, and the star formation history. However, depending on the normalization of the models, they have usually not been able to simultaneously fit the zero-point of the Tully-Fisher relation and the luminosity function of field galaxies, although more recent work has been continuously improving on that.

The construction of dark matter merging history trees (Kauffmann & White 1993; Somerville & Kolatt 1999) is an important ingredient in semi-analytic models. In most studies, Monte-Carlo realizations of merging histories for individual objects are generated using the extended Press-Schechter formalism (Press & Schechter 1974; Bond et al. 1991). A disadvantage of this approach is that there is little information about the spatial distribution of galaxies. To still study the clustering properties of galaxies, semi-analytic models have been employed to relate the galaxy and mass distributions of cosmological N-body simulations, effectively placing galaxies into virialized haloes at a single time output. Already White et al. (1987) used such a technique, and many recent studies followed it (Kauffmann et al. 1997; Governato et al. 1998; Somerville & Primack 1998; Benson et al. 1999).

As a natural extension of this approach, one can attempt to use N-body simulations not only to provide spatial positions and masses of haloes, but also to directly measure the full merging history trees of galaxies. White et al. (1987) were the first to develop a crude version of this approach, which avoids the uncertainties and inaccuracies in the Press-Schechter formalism, and it eventually allows semi-analytic models to make more direct contact with the dark matter dynamics. For example, one can imagine using the spin parameters and orbits of dark haloes measured from the simulations to obtain a better description of disk sizes or merging timescales. However, this approach requires simulations with very high mass resolution and the storage of a large number of simulation outputs, resulting in substantial raw data volume.

Recently, Roukema et al. (1997) studied merging history trees directly from N-body simulations using scale-free simulations and a rather limited number of simulation outputs. A much more extensive study has been done by Kauffmann et al. (1999a, hereafter KCDW), who have analysed the merging history trees of a set of high-resolution N-body simulations, with a total of 51 output times between redshift $z = 20$ and $z = 0$. KCDW combined semi-analytic models of galaxy formation with their analysis of the dark matter dynamics to study the clustering strength and the galaxy bias in two cosmological models. In subsequent papers, they used this methodology to predict the evolution of clustering to high redshift (Kauffmann et al. 1999b), and to construct realistically selected mock redshift surveys (Diaferio et al. 1999).

Using a somewhat different technique, van Kampen et al. (1999) also made use of the full merging history of N-body simulations. They modified an existing N-body code such that heavier 'tracer' particles were introduced during run-time. These tracer particles replaced locally overdense groups of particles, which they identified with galaxies. In this way, van Kampen et al. (1999) tried to circumvent the overmerging problem.

Our approach aims in a similar direction, yet we follow the methodology of KCDW and extend it directly to the regime of substructure. To this end, we study three simulations of the formation of a rich cluster of galaxies using very high mass resolution. The simulations follow the same object, a cluster of mass $\simeq 10^{15}\ h^{-1}\mathrm{M}_{\odot}$ in a flat, low-density cosmology, with different mass resolution. We resolve the virial region of the final cluster with 0.12, 0.61 and 3.5 million particles, respectively, and we sample the field in the region around the cluster with about twice as many high-resolution particles in each case. The cosmological tidal field is properly represented by an additional boundary region with 3.1 million particles, extending to a distance of $70\,h^{-1}\mathrm{Mpc}$ from the cluster. This sequence of simulations allows a direct test of the dependence of our results on numerical resolution.

We develop a new algorithm, SUBFIND, to identify substructure within groups forming in these simulations. This algorithm defines "subhaloes" as locally overdense, self-bound particle groups, and it is able to detect hierarchies of substructure using just a single simulation output. As in KCDW, we have stored 51 simulation snapshots from $z = 20$ to the current epoch, and we trace the merging history of groups and their subhaloes from output to output. We modify the semi-analytic recipes employed by KCDW to allow the inclusion of subhaloes, and we analyse the changes resulting from that.

In particular, we study the luminosity function of the cluster in the B-band. We investigate the mass-to-light ratio of the cluster, the Tully-Fisher relation of spirals in the field, and the $B-V$ colors of our model galaxies. We show that the new subhalo-scheme gives rise to a pronounced morphology-clustercentric relation. Our approach also allows an analysis of the star formation history of the field and cluster galaxies, showing that the stars of cluster galaxies are substantially older than those of field galaxies, and that they have formed well before the cluster itself is assembled.

Interestingly, the new subhalo-analysis improves the agreement with observational data considerably, especially with respect to the cluster luminosity function. Most of the bright galaxies in the final, highly-resolved cluster are still connected to well localized subhaloes within the smooth dark matter background of the cluster. There is hence no need to estimate merging timescales for galaxies falling into the cluster. Apparently, inaccuracies of such estimates lead to the problem of excessively bright first ranked cluster galaxies in the standard methodology of KCDW.

This chapter is organized as follows. In Section 2 we describe the N-body simulations, and in Section 3 we briefly review the techniques of KCDW, and our specific implementation of them. In Section 4 we discuss our techniques to identify dark matter substructure within larger haloes, and our methods of tracing it between different simulation outputs. We then describe the implementation of semi-analytic models including this subhalo information, and we present results obtained with these prescriptions in Section 5. Finally, we discuss our findings in Section 6.

Table 4.1: Numerical parameters of cluster simulations. All four simulations compute the evolution of the same cluster, in a ΛCDM universe with cosmological parameters $\Omega_0 = 0.3$, $\Omega_\Lambda = 0.7$, $\Gamma = 0.21$, $\sigma_8 = 0.9$, and $h = 0.7$. The simulations follow a sphere of matter, with comoving diameter $141\,h^{-1}$Mpc. In the Table, we give the particle mass $m_{\rm p}$ used in the central high-resolution zone, the starting redshift $z_{\rm start}$, the gravitational softening ϵ in the high-resolution zone, the number $N_{\rm hr}$ of high-resolution particles, the number $N_{\rm bnd}$ of boundary particles, the total number $N_{\rm tot}$ of particles, and the number $N_{\rm p}$ of processors used in each of the simulations S1-S4. The gravitational softening was kept fixed at the given values in physical coordinates below redshift $z = 9$, and in comoving coordinates above this redshift.

	S1	S2	S3	S4
$m_{\rm p}$ $[h^{-1}{\rm M}_\odot]$	6.87×10^9	1.36×10^9	2.38×10^8	4.68×10^7
$z_{\rm start}$	30	50	80	140
ϵ $[h^{-1}{\rm kpc}]$	6.0	3.0	1.4	0.7
$N_{\rm hr}$	450088	1999978	12999878	66000725
$N_{\rm bnd}$	3029956	3117202	3016932	3013281
$N_{\rm tot}$	3480044	5117180	16016810	69014006
$N_{\rm p}$	16	32	128	512

4.2 Cluster simulations

In this study, we analyse collisionless simulations of clusters of galaxies that are generated by the technique of 'zooming in' on a region of interest (Tormen et al. 1997). In a first step, a cosmological simulation with sufficiently large volume is used to allow the selection of a suitable target cluster. For this purpose, we employed the GIF-ΛCDM[1] model carried out by the Virgo consortium. It has cosmological parameters $\Omega_0 = 0.3$, $\Omega_\Lambda = 0.7$, $h = 0.7$[2], spectral shape $\Gamma = 0.21$, and was cluster-normalized to $\sigma_8 = 0.9$. The simulation followed 256^3 particles of mass $1.4 \times 10^{10} h^{-1}{\rm M}_\odot$ within a comoving box of size $141.3\,h^{-1}$Mpc on a side. Note that this simulation is one of the models recently studied by KCDW. We selected the second most massive cluster that had formed in the GIF simulation for further study. This cluster has a virial mass $8.4 \times 10^{14}\,h^{-1}{\rm M}_\odot$, and it appears to be well relaxed at the present time.

In a second step, we simulated the formation of this cluster once more using a greatly increased mass and force resolution. To this end, the particles in the final GIF-cluster and in its immediate surroundings were traced back to their Lagrangian region in the initial conditions. The corresponding part of the displacement field was then sampled using a glass-like particle distribution with smaller particle masses than in the GIF simulation. Due to the increase in resolution, the fluctuation spectrum could now be extended to smaller scales. We added a random realization of this additional small-

[1]The GIF project is a joint effort by astrophysicists in Germany and Israel.
[2]We employ the convention $H_0 = 100\,h\,{\rm km\,s^{-1}\,Mpc^{-1}}$.

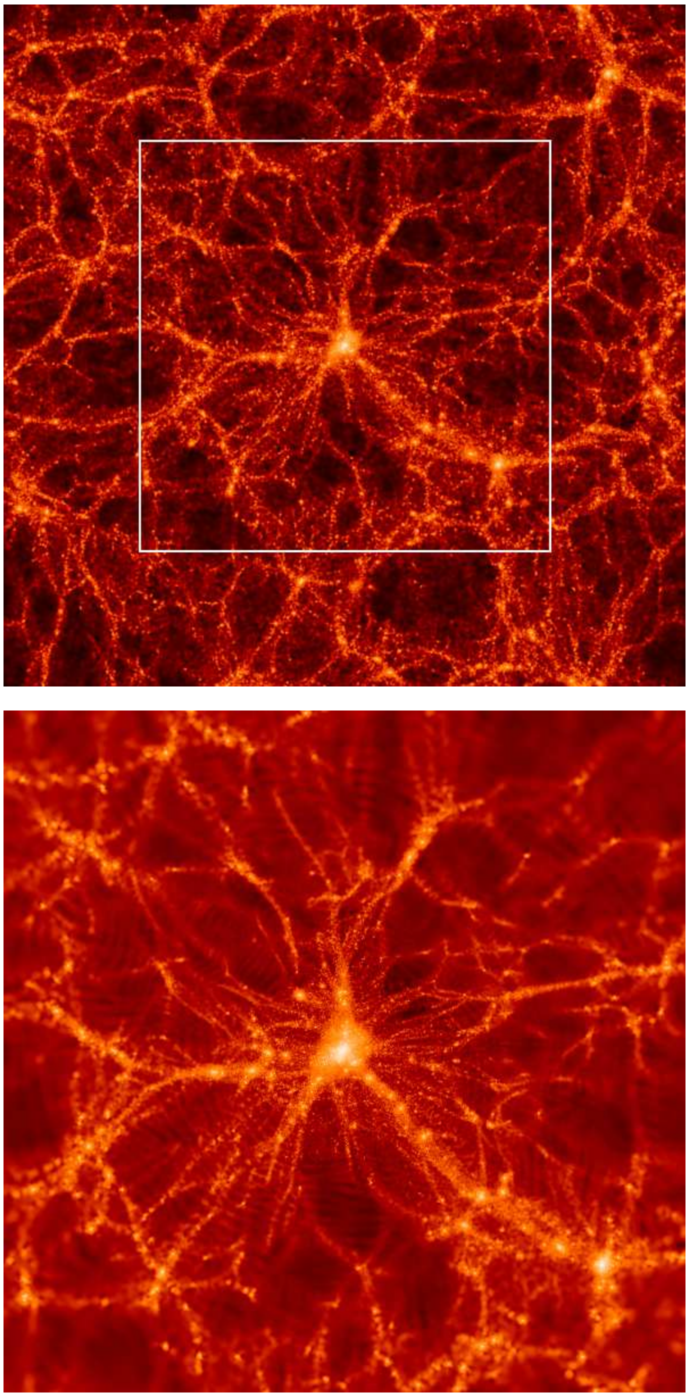


Figure 4.1: The projected mass density fields in slices of thickness $10\,h^{-1}$Mpc around the cluster center in the original GIF simulation (top) and the S2 resimulation (bottom). The left image is $141\,h^{-1}$Mpc on a side, and the white square marks the region ($85\,h^{-1}$Mpc on a side) that is displayed from the corresponding resimulation.

scale power, while we kept all the waves on larger scales that had been used in the GIF simulation.

Outside this central high-resolution region, we gradually degraded the resolution by using particles with masses that grow with distance from the center. In this 'boundary region', we employed a spherical grid whose spacing grew with distance from the high resolution zone. The spherical boundary region extends to a total diameter of $141.3\,h^{-1}$Mpc, which is just the box size of the original GIF simulation. Beyond this region, we assumed vacuum boundary conditions, i.e. a vanishing density fluctuation field. Using comoving coordinates, we then evolved the simulations to redshift $z = 0$ with our parallel tree-code GADGET. This code uses individual timesteps for all particles, and was designed to run on massively parallel supercomputers with distributed memory. Parallelization is achieved explicitly using the communication library of the *Message Passing Interface* (MPI). A full account of the numerical and algorithmic details of GADGET is given in Chapter 5.

To be able to study systematic effects arising from numerical resolution, we simulated the same cluster several times, increasing the resolution step by step. In the first step, refered to as simulation 'S1' from here on, the particle mass was just about two times smaller than in the original GIF simulation. For S1, we used a total of 450000 particles for the high-resolution zone, and 3 million for the boundary region. Note that one would usually try to make the number of boundary particles smaller than that of the high-resolution particles. However, with the sequence of our planned simulations in mind, we here opted for this high number of boundary particles because we wanted to keep the number approximately the same for all runs of our simulation set.

Except for slight changes at its inner rim, the boundary region was then kept fixed for the other simulations were we populated the central zone with many more particles. In simulation 'S2', we used 2 million high-resolution particles, and we reduced the gravitational softening to $\epsilon = 2.0\,h^{-1}$kpc. In simulation 'S3', we then employed a total of 13 million high-resolution particles with mass $2.4 \times 10^8\,h^{-1}\mathrm{M}_\odot$, and a softening of $1.4\,h^{-1}$kpc. In this chapter, we will just report results for the three simulations S1-S3, but we note that we have recently completed yet another step in this sequence. In simulation 'S4' we used a total of 66 million particles for the high-resolution zone, pushing the particle mass down to $4.68 \times 10^7\,h^{-1}\mathrm{M}_\odot$ and the spatial resolution down to $0.7\,h^{-1}$kpc. In each case, roughly one third of the particles in the high-resolution zone end up in the virialized region of the final cluster. This means that S4 resolves a single object with about 20 million particles.

In Table 4.1, we summarize important numerical parameters of our simulations. Note that we have softened gravity using a spline kernel. Our cited values for ϵ are such that the gravitational potential of a point mass at zero lag is $\Phi = -Gm/\epsilon$, and that the softened force becomes Newtonian at a distance $2.8\,\epsilon$. We have kept the softening length fixed in physical coordinates below redshift $z = 9$, and in comoving coordinates at higher redshift. The softening for the boundary particles was set to much larger values, in an inner shell around the cluster to $15\,h^{-1}$kpc, and further outside to $75\,h^{-1}$kpc. For all four simulations, we stored 51 outputs, logarithmically spaced in expansion factor between redshifts $z = 20$ and $z = 0$.

In Figure 4.1, we show two images comparing the original density field of the GIF-simulation with that of our S3 resimulation. The filaments of dark matter around the cluster are nicely reproduced by S3, even relatively far away from the cluster, where the resolution of S3 has already fallen below that of the GIF simulation. Also visible is the much larger resolution of S3 in the region of the actual cluster.

4.3 Modeling galaxy formation using N-body merging trees

In the following, we briefly summarize our specific implementation of the techniques developed by KCDW to combine semi-analytic models for galaxy formation with dark matter merging history trees directly measured from cosmological N-body simulations. We will extend this formalism later to include dark matter substructure, and we will be especially interested in any changes of the results arising from that.

There are essentially two main parts in the modeling: (1) The measurement of dark matter merging trees from a sequence of simulation outputs. (2) The implementation of the actual semi-analytic recipes for galaxy formation on top of these merging trees. Both parts of the modeling are technically complex and warrant a detailed discussion. In this Section, we start by describing our implementation of the techniques of KCDW. First we treat the construction of the merging trees, then the physics of galaxy formation. Having set the ground in this way, we will then describe in Sections 4 and 5 what we change in the two parts to allow the inclusion of subhalo information.

4.3.1 Following the merging trees

For each simulation output, we compile a list of dark matter haloes with the friends-of-friends (FOF) algorithm using a linking length of 0.2 in units of the mean interparticle separation. We only include groups with at least 10 particles in the halo catalogue. The majority of such haloes are already stable, i.e. particles found in 10-particle groups at one output time are also found in a halo in a subsequent simulation output. For each halo, we also determine the most-bound particle within the group, where 'most-bound' here refers to the particle with the minimum gravitational potential.

We now follow the merger tree of the dark matter from output to output. A halo $H_{\rm B}$ at redshift $z_{\rm B}$ is defined to be a *progenitor* of a halo $H_{\rm A}$ at redshift $z_{\rm A} < z_{\rm B}$, if at least half of the particles of $H_{\rm B}$ are contained within $H_{\rm A}$, and the most bound particle of $H_{\rm B}$ is contained in $H_{\rm A}$, too. These definitions already suffice to uniquely define the dark matter merging trees.

4.3.1.1 Defining a galaxy population

So far we are just dealing with catalogues of dark matter haloes. We now supplement this with the notion of a *galaxy population* with physical properties given by the semi-analytic techniques. In our formalism, each dark halo carries exactly one *central* galaxy, and its position is given by the most-bound particle of the halo. Only the central galaxy is supplied with additional gas that cools within the halo.

A halo can also have one or several *satellite* galaxies, where the position of each of them is given by one of the particles of the halo. Satellites are galaxies that had been central galaxies themselves in the past, but their haloes have merged at some previous time with the larger halo they now reside in. Satellite galaxies orbit in their halo and are assumed to merge with the central galaxy on a dynamical friction timescale. Note that they are cut-off from the supply of fresh cool gas, so they may only form stars until their reservoir of cold gas is exhausted.

Finally, we define a class of *field* galaxies, which are introduced to keep track of satellites whose particles are currently not attached to any halo, for example because they have been ejected out of their parent halo. Usually, these "lost" field galaxies are again accreted onto a halo later on.

4.3.1.2 Following the galaxy population in time

At a given output time, we therefore deal with a galaxy population consisting of *central* galaxies, *satellite* galaxies, and *field* galaxies, each attached to the position of a simulation particle. Starting at the first output time at high redshift (when the first haloes have formed), we initialize the galaxy population with a set of central galaxies, one for each halo, with stellar mass, cold gas mass, and luminosity set to zero. The physical properties of these galaxies are then evolved to the next output time, where we obtain a new galaxy population based on a combination of semi-analytic prescriptions and the merging history of the dark matter. Repeating this scheme from output to output forward in time we obtain the galaxy population at the present time, and at all output times at higher redshift.

We now describe in more detail our rules and prescriptions for this evolution. Beginning with the galaxy population at redshift $z_{\rm B}$, we first generate the galaxies of the new population at redshift $z_{\rm A} < z_{\rm B}$ based on the merging history of the dark matter. Using the group catalogues of the corresponding simulation outputs and the galaxy population at $z_{\rm B}$, we construct an 'initial' set of galaxies at $z_{\rm A}$ as follows:

1. Each galaxy at $z_{\rm B}$ is assigned to its new halo at $z_{\rm A}$. If the particle used to tag a galaxy does not reside in any halo, the galaxy becomes a field galaxy.

2. Each halo at $z_{\rm A}$ selects its central galaxy as the central galaxy of its most massive progenitor. This central galaxy is repositioned to the position of the most-bound particle, i.e. a new particle tagging the galaxy is selected. The central galaxies of all other progenitors become satellites of the halo.

3. If a halo has no progenitors, a new central galaxy is created at the position of its most-bound particle. In the event that the halo contains one or more galaxies (particles recovered from the field), the central galaxy is picked as the most massive of them.

Once the new set of galaxies is generated in this way, the properties of the galaxies are evolved according to the physical prescriptions described below, resulting finally in the new galaxy population at redshift $z_{\rm A}$. Note that some of the satellite galaxies generated in the initial set for $z_{\rm A}$ will merge with central galaxies during this evolution, and thus not necessarily be part of the 'final' population at $z_{\rm A}$.

4.3.2 Physical evolution of the galaxy population

We model the following physical processes: (1) Radiative cooling of hot gas onto central galaxies. (2) Transformation of cold gas into luminous stars by star formation. (3) Reheating of cold gas, or its ejection out of the halo, by supernova feedback. (4) Orbital decay of satellites and their merging with central galaxies. (5) Spectrophotometric evolution of the luminous stars. (6) Simplified morphological evolution of galaxies. (7) Metal enrichment of the gas. Below we detail the physical parameterizations adopted for these processes.

4.3.2.1 Gas cooling

Gas cooling is modeled as in White & Frenk (1991). We assume that the hot gas within a dark halo is distributed like an isothermal sphere with density profile $\rho_{\rm g}(r)$. Then the local cooling time $t_{\rm cool}(r)$ can be defined as the ratio of the specific thermal energy content of the gas, and the cooling rate per unit volume, viz.

$$t_{\rm cool}(r) = \frac{3}{2}\frac{kT\rho_{\rm g}(r)}{\overline{\mu}m_{\rm p}n_{\rm e}^2(r)\,\Lambda(T,Z)}. \tag{4.1}$$

Here $\overline{\mu}m_{\rm p}$ is the mean particle mass, $n_{\rm e}(r)$ the electron density, and $\Lambda(T,Z)$ the cooling rate. The latter depends quite strongly on the metallicity Z of the gas, and on the virial temperature $T = 35.9\,(V_{\rm vir}/\,{\rm km\,s^{-1}})^2\,{\rm K}$ of the halo. We employ the cooling functions computed by Sutherland & Dopita (1993) for collisional ionisation equilibrium to represent $\Lambda(T,Z)$.

We define the cooling radius $r_{\rm cool}$ as the radius for which $t_{\rm cool}$ is equal to the current age $t_{\rm age}$ of the universe. If the cooling radius lies well within the virial radius $R_{\rm vir}$ of a given halo, we take the cooling rate to be

$$\frac{{\rm d}M_{\rm cool}}{{\rm d}t} = 4\pi\rho_{\rm g}(r_{\rm cool})r_{\rm cool}^2\frac{{\rm d}r_{\rm cool}}{{\rm d}t}. \tag{4.2}$$

Note that we define the virial radius $R_{\rm vir}$ of a FOF-halo as the radius of a sphere which is centered on the most-bound particle of the group and has an overdensity 200 with respect to the critical density. We take the enclosed mass $M_{\rm vir} = 100H^2R_{\rm vir}^3/G$ as the virial mass, and we assign a virial velocity as $V_{\rm vir}^2 = GM_{\rm vir}/R_{\rm vir}$.

Adopting an isothermal sphere for the distribution of the hot gas of mass $M_{\rm hot}$ within the halo, i.e.

$$\rho_{\rm g}(r) = \frac{M_{\rm hot}}{4\pi R_{\rm vir}r^2}, \tag{4.3}$$

the cooling rate is then given by

$$\frac{\mathrm{d}M_{\mathrm{cool}}}{\mathrm{d}t} = \frac{M_{\mathrm{hot}}}{R_{\mathrm{vir}}} \frac{r_{\mathrm{cool}}}{2t_{\mathrm{age}}}. \tag{4.4}$$

At early times or for low-mass haloes the cooling radius can be much larger than the virial radius. In this case, the hot gas is never expected to be in hydrostatic equilibrium, and the cooling rate will essentially be limited by the accretion rate. We approximate this rate with

$$\frac{\mathrm{d}M_{\mathrm{accr}}}{\mathrm{d}t} = \frac{M_{\mathrm{hot}}\, V_{\mathrm{vir}}}{R_{\mathrm{vir}}}, \tag{4.5}$$

where V_{vir} is the virial velocity of the halo. We adopt the minimum of equations (4.4) and (4.5) as our actual cooling rate.

4.3.2.2 Star formation

We model the star formation rate of a galaxy as

$$\frac{\mathrm{d}M_{\star}}{\mathrm{d}t} = \alpha \frac{M_{\mathrm{cold}}}{t_{\mathrm{dyn}}}, \tag{4.6}$$

where M_{cold} is the mass of its cold gas, and t_{dyn} is the dynamical time of the galaxy. We approximate the latter as

$$t_{\mathrm{dyn}} = \frac{R_{\mathrm{eff}}}{V_{\mathrm{vir}}}, \tag{4.7}$$

with $R_{\mathrm{eff}} = 0.1 R_{\mathrm{vir}}$, i.e. we set the effective stellar radius to a fixed fraction of the virial radius. Note that at a fixed redshift, R_{vir} is proportional to V_{vir}, hence t_{dyn} depends only on redshift. The dimensionless parameter α regulates the efficiency of star formation and is treated as a free parameter. Once a galaxy falls into a larger halo and becomes a satellite, the values of R_{vir} and V_{vir} are not changed any more. The galaxy can then continue to form stars until its reservoir of cold gas is exhausted, but it does not receive new cold gas by cooling processes.

4.3.2.3 Feedback

Assuming a universal initial mass function (IMF), the energy released by supernovae per formed solar mass is $\eta_{\mathrm{SN}} E_{\mathrm{SN}}$, where η_{SN} gives the expected number of supernovae per formed stellar mass, and E_{SN} is the energy released by each supernova. The formation of a group of stars with mass $\Delta M_{\star}$ will thus be accompanied by the release of a feedback energy of $\eta_{\mathrm{SN}} E_{\mathrm{SN}} \Delta M_{\star}$, where we adopt $\eta_{\mathrm{SN}} = 5.0 \times 10^{-3} \mathrm{M}_{\odot}^{-1}$, based on the Scalo (1986) IMF, and $E_{\mathrm{SN}} = 10^{51}$ erg.

One major uncertainty is how this energy affects the evolution of the interstellar medium, and how the star formation rate is regulated by it. We here assume that the feedback energy reheats some of the cold gas back to the virial temperature of the dark halo. The amount of gas reheated by this process is then

$$\Delta M_{\rm reheat} = \frac{4}{3}\,\epsilon\,\frac{\eta_{\rm SN} E_{\rm SN}}{V_{\rm vir}^2}\,\Delta M_\star, \tag{4.8}$$

where the dimensionless parameter ϵ describes the efficiency of this process.

Following KCDW, we consider two alternative schemes for the fate of the reheated gas. In the *retention* scheme, the reheated gas is simply transfered from the cold phase back to the hot gaseous halo, and the reheated gas thus stays within the halo. Alternatively, in the *ejection* scheme we assume that the gas leaves the halo, and it is only re-incorporated into the halo at some time later. If $\Delta M_{\rm ejec}$ is the total gas mass ejected by a galaxy, we model this reincorporation by simply setting $\Delta M_{\rm ejec}$ to zero again whenever the galaxy's halo merges with a larger system, i.e. the amount of hot gas is increased by $\Delta M_{\rm ejec}$ in this event.

4.3.2.4 Mergers of galaxies

In CDM universes, large haloes form by merging processes out of smaller haloes. Similarly, mergers of galaxies are an inevitable process in hierarchical galaxy formation. We assume that the satellite galaxies orbiting within a dark matter halo experience dynamical friction and will eventually merge with the central galaxy of the halo. In principle, mergers between two satellite galaxies are also possible. These events are expected to be rare, but they happen occasionally, as we will show later on. Here, such events are neglected, but they will be taken into account in our subhalo-scheme.

N-body simulations by Navarro et al. (1995) suggest that the merging timescale can be reasonably well approximated by the dynamical friction timescale

$$T_{\rm friction} = \frac{1}{2}\frac{f(\epsilon)}{C}\frac{V_{\rm c} r_{\rm c}^2}{G M_{\rm sat}\ln\Lambda}\,. \tag{4.9}$$

The formula is valid for a small satellite of mass $M_{\rm sat}$ orbiting at a radius $r_{\rm c}$ in an isothermal halo of circular velocity $V_{\rm c}$. The function $f(\epsilon)$ describes the dependence of the decay on the eccentricity of the satellites' orbit, expressed in terms of $\epsilon = J/J_{\rm c}(E)$, where $J_{\rm c}(E)$ is the angular momentum of a circular orbit with the same energy as the satellite. The function $f(\epsilon)$ is well approximated by $f(\epsilon) \simeq \epsilon^{0.78}$, for $\epsilon > 0.02$ (Lacey & Cole 1993). C is a constant with value $C \simeq 0.43$, and $\ln\Lambda$ is the Coulomb logarithm.

We follow Kauffmann et al. (1999a) and approximate $r_{\rm c}$ with the virial radius of the halo when the satellite first falls into it. To describe the orbital distribution, we adopt the average value $\langle f(\epsilon)\rangle \simeq 0.7$, computed by van den Bosch et al. (1999). Note that this differs slightly from KCDW who drew a random orbit uniformly from $\epsilon \in [0.02, 1]$. We identify the mass of the satellite with the virial mass of the galaxy at the time when it was last a central galaxy, and we approximate the Coulomb logarithm with $\ln\Lambda = (1 + M_{\rm vir}/M_{\rm sat})$.

When a satellite merges with a central galaxy, all its stellar mass is transfered to the bulge component of the central galaxy, and the photometric properties of this galaxy are updated accordingly. Similarly, the cold gas of the satellite is transferred to the central galaxy. If the mass ratio between the stellar components of the merging galaxies is larger than some threshold value (we adopt 0.3 for that), we assume that we deal with a *major merger*. In this event, the merger destroys the disk of the central galaxy completely, and all stars form a single spheroid, i.e. they generate a bulge. In addition, we assume that all the cold gas left in the two merging galaxies is rapidly consumed in a starburst. The stars created in this burst are also added to the bulge component. Since the central galaxy is fed by a cooling flow, it can later on grow a new disk component.

4.3.2.5 Spectrophotometric evolution

The photometric properties of our model galaxies can be constructed using stellar population synthesis models (Bruzual & Charlot 1993). Such models assume a distribution of stars according to an initial mass function, and they evolve the spectral energy distributions (SEDs) of these stars according to theoretical evolutionary tracks. In this way, the spectra and colors of a stellar population formed in a short burst of star formation can be followed as a function of time. Once the evolution $F_\nu(t)$ of the SED of a single isochrone population of stars is known, the SED $S_\nu(t)$ of a galaxy can be computed as

$$S_\nu(t) = \int_0^t F_\nu(t - t')\, \dot{M}_\star(t')\, \mathrm{d}t' \tag{4.10}$$

from its star formation history $\dot{M}_\star(t)$. Upon convolution with standard filters, colors and luminosities in the desired bands can be obtained. In principle, this technique also allows the redshifting of spectra, and the incorporation of k-corrections to make direct contact with observational photometric data at high redshift. In this work we use updated evolutionary synthesis models by Bruzual & Charlot (in preparation). In these models, we assume solar metallicity throughout.

4.3.2.6 Morphological evolution

Simien & de Vaucouleurs (1986) have observed a good correlation between the B-band bulge-to-disk ratio, and the Hubble-type T of galaxies. For a magnitude difference $\Delta M \equiv M_{\rm bulge} - M_{\rm total}$ they find a mean relation

$$\langle \Delta M \rangle = 0.324(T+5) - 0.054(T+5)^2 + 0.0047(T+5)^3. \tag{4.11}$$

Following previous semi-analytic studies, we therefore simply assign morphologies based on this equation. Specifically, we will usually classify galaxies with $T < -2.5$ as ellipticals, those with $-2.5 < T < 0.92$ as S0's, and those with $T > 0.92$ as spirals and irregulars. Note that galaxies without any bulge are classified as type $T = 9$.

4.3.2.7 Chemical enrichment

Metallicity can strongly influence both the cooling rates and the photometric evolution of stellar populations. In this work we will only consider its influence on the cooling function. Initially, we assume that all gas is of primordial abundance with zero metallicity. Due to star formation and the accompanying supernova explosions, the interstellar medium is enriched with metals. If some of this processed material is mixed into the hot halo gas, the cooling rates can be increased substantially.

We assume that each generation of stars of mass $\Delta M_\star$ produces a mass $\Delta M_\mathrm{h} = y\,\Delta M_\star$ of heavy elements, where y is the constant yield in the instantaneous recycling approximation. It is not well known how these metals get distributed in galaxies. In our default model, we will assume that the metals always remain in the cold gas phase of each galaxy. In this case, the metallicity of the *halo* gas remains zero, and the mean metallicity of stars will approach the yield y. In this default model, we also take the metallicity of the reheated gas to be zero, despite the enrichment of the cold gas.

However, in our feedback model we assumed that supernovae reheat gas from the cold reservoir of a galaxy. If the latter is enriched with metals, it seems plausible that metals are transported along with the reheated gas. In the retention model, the reheated gas is mixed with the hot halo, thereby enriching it. In the ejection model, the reheated gas leaves the galaxy for a while and thus pollutes the intergalactic medium. Assuming that the accreted gas is of primordial abundance, ejection feedback will then not change the metallicity of the cooling gas.

In Section 4.6.7 we will briefly discuss the effects of metal transport in models with retention feedback. Note that feedback is much more efficient in small haloes, i.e. metallicity effects can be expected in such a transport model to strongly increase cooling rates of low mass galaxies.

4.3.3 More implementation details

In our practical implementation of the physical evolution of the galaxy population, we process the following steps. We first estimate merger timescales for those satellites that have newly entered a given halo, i.e. the galaxies that had not been contained in the largest progenitor of the halo. This 'merger clock' then decreases with time, and the satellite will be merged with the central galaxy when this time is over. Note that the merger clock may be reset before the merger happens if the halo containing the satellite merges with a larger system.

We then compute the total amount of hot gas available for cooling in each halo. Assuming on average a universal abundance of baryons equal to the primordial one, this is simply

$$M_\mathrm{hot} = f_\mathrm{b} M_\mathrm{vir} - \sum_i \left[M_\star^{(i)} + M_\mathrm{cold}^{(i)} + M_\mathrm{ejec}^{(i)} \right], \tag{4.12}$$

where the sum extends over all galaxies within the halo. Here $f_\mathrm{b} = \Omega_\mathrm{b}/\Omega_0$ denotes the baryon fraction of the universe. Using the cooling model of Section 4.3.2.1, we then estimate the cooling rate onto each central galaxy, and we keep this rate constant during the time ΔT between the two simulation outputs.

Once these quantities are known, we solve the simple differential equations describing star formation, cooling and feedback. We typically use a number of $N \simeq 50$ small timesteps of size $\Delta t = \Delta T/N$ for this purpose. At each of these small steps, new cold gas is added to the central galaxies. For each galaxy, we then form some stellar mass $\Delta M_\star$ according to its star formation rate, and we update its current and future photometric properties accordingly. The cold gas mass of each galaxy is reduced by the amount of stars formed, and by the mass of the gas that is reheated or ejected by supernova feedback.

At the end of each of the small steps, the merger clocks of the satellites are reduced by Δt. If a satellite's merging time falls below zero, it is merged with the central galaxy of its parent halo. In practice, this means that the luminosity, the stellar mass, and the gas mass of the satellite are transfered to the central galaxy, and that the satellite is removed from the list of galaxies. In addition, in the event of a major merger all cold gas of the central galaxy is consumed in a short starburst, and all stellar material is transformed into a spheroid.

4.3.4 Choice of model parameters

Following KCDW, we use the I-band Tully-Fisher relation to normalize our models, i.e. to set the free parameters α and ϵ which specify the efficiency of star formation and feedback, respectively. We consider the central galaxies of haloes in the periphery of the cluster, with morphological types corresponding to Sb/Sc galaxies for that purpose. Note that we only use haloes that are not 'contaminated' by heavier boundary particles. The remaining number of galaxies is sufficiently large to construct a well defined Tully-Fisher diagram. We try to fit the velocity based I-band Tully-Fisher relation

$$M_I - 5\log h = -21.00 - 7.68(\log W - 2.5) \tag{4.13}$$

measured by Giovanelli et al. (1997). We will assume that the circular velocity $V_{\rm c}$ of a spiral galaxy is 10% larger than the virial velocity of that galaxies halo. This is motivated by detailed models for the structure of disk galaxies (Mo et al. 1998) embedded in cold dark matter haloes that follow the universal NFW profile (Navarro et al. 1996, 1997). We set the velocity width W as twice the circular velocity.

Keeping other parameters fixed, we find that varying ϵ changes both the slope and the zero-point of the Tully-Fisher relation strongly. In particular, making feedback stronger tilts the Tully-Fisher relation towards steeper slopes, whereas the star formation efficiency only weakly affects the zero-point. However, α has a strong effect on the gas mass fraction left in galaxies at the present time.

In principle, the parameters α and ϵ may be specified using the slope and zero-point of the Tully-Fisher relation *alone*. However, the weak dependence of the Tully-Fisher relation on α makes this impractical. Similar to KCDW, we instead use an additional criterion and require that the cold gas mass in a 'Milky-Way' galaxy of circular velocity $V_\mathrm{c} = 220\,\mathrm{km\,s^{-1}}$ is about $0.6 \times 10^{10}\,h^{-1}\mathrm{M}_\odot$.

Note that the baryon fraction f_b can strongly influence the cooling rates, and thus the absolute normalization of the models. As White et al. (1993) have shown, the baryon content of rich clusters of galaxies strongly argues for a baryon fraction as high as $f_\mathrm{b} = 0.2$. This is inconsistent with big bang nucleosynthesis constraints in a critical density universe, but can be accommodated within low-density cosmologies, like the one considered in our cluster models. We will assume $f_\mathrm{b} = 0.2$ in this study.

4.4 Following halo substructure

4.4.1 Identification of substructure

A basic step in the analysis of cosmological simulations is the identification of virialized particle groups, which give the sites where luminous galaxies form. Perhaps the most popular technique employed for this task is the friends-of-friend (FOF) algorithm. It places any two particles with a separation less than some linking length b into the same group. In this way, particle groups are formed that correspond to regions approximately enclosed by isodensity contours with threshold value $\rho \propto 1/b^3$. For an appropriate choice of b, groups are selected that are close to the virial overdensity predicted by the spherical collapse model. FOF is both simple and efficient, and its group catalogues agree quite well with the predictions of Press-Schechter theory (Götz et al. 1998).

However, FOF has a tendency to occasionally link independent structures across feeble particle bridges, and in its standard form with a linking length of $b \simeq 0.2$ it is not capable of detecting substructure inside larger virialized objects. Using sufficiently high mass resolution, recent studies (Tormen 1997; Tormen et al. 1998; Ghigna et al. 1998; Klypin et al. 1999) were able to demonstrate that substructure in dense environments like groups or clusters may survive for a long time. The cores of the dark haloes of galaxies that fall into a cluster will thus remain intact, and orbit as self-gravitating objects in the smooth dark matter background of the cluster. In previous simulations, haloes falling into clusters were usually quickly evaporated, and the clusters exhibited little signs of substructure. It now appears, that sufficient numerical force and mass resolution is enough to resolve this "overmerging" problem.

The identification of substructure within dark matter haloes is a challenging technical problem, and several algorithms to find "haloes within haloes" have been proposed. In hierarchical friends-of-friends (HFOF) algorithms (Gottlöber et al. 1998; Klypin et al. 1999) the linking length of plain FOF is reduced in a sequence of discrete steps, selecting in this way groups of higher and higher overdensity, eventually capturing true substructure.

Clearly, the need for a well-posed physical definition of "substructure" arises early on in such an analysis. Most authors have required such subhaloes to be locally overdense and self-bound, a requirement that we will impose in the following as well. Note that this implies that any locally overdense region within a dense background needs to be treated with an unbinding procedure. This is because a small halo within a larger system represents only a relatively small fluctuation in the density, and a substantial amount of mass within the overdense region will just stream through and not be gravitationally bound to the substructure itself.

Klypin et al. (1999) also proposed a bound density maximum (BDM) algorithm, where they iteratively determined the bound subset of particles in a sphere around a local density maximum. In the method of Tormen et al. (1998) previous simulation outputs are used to track the infall of particle groups into larger systems. Once such a particle group from the field was accreted by a cluster, they simply determined the subset of these particles that still remained self-bound.

Another approach is followed in DENMAX (Gelb & Bertschinger 1994) and its offspring SKID, where particles are moved along the local gradient in density towards a local density maximum. Particles ending up in the 'same' maximum are then linked together as a group. SKID has been employed by Ghigna et al. (1998) to find substructure in a rich cluster of galaxies, and to study the statistical properties of the detected subgroups.

Integrating the gradient of the density field and moving the particles is not without technical subtleties, for example a suitable stopping condition is needed. The new algorithm HOP of Eisenstein & Hut (1998) tries to avoid these difficulties by restricting the group search to the set of original particle positions, just like FOF does. In HOP, one first obtains an estimate of the local density for each particle, and then attaches it to its densest neighbor. In this way a set of disjoint particle groups are formed. However, a number of additional rules are needed to link and prune some of these groups. For example, HOP may split up a single virialized clump into several pieces of unphysical shape, which have to be joined using auxiliary criteria.

It appears that all of these techniques have different strengths and weaknesses, and none is completely satisfactory yet. We have therefore come up with a new algorithm to detect substructure in dark matter haloes that combines ideas of SKID, HOP and FOF, and adds some new ones. For easier reference, we dub this algorithm SUBFIND (for *subhalo finder*).

4.4.2 The algorithm SUBFIND

Our objective with SUBFIND is to be able to extract substructure defined as locally overdense, self-bound particle groups within a larger parent group. We will take this group as a particle group pre-selected with a standard FOF linking length, although SUBFIND could operate on arbitrary particle groups, or with slight modifications on all of the particles in a simulation at once. The selection of FOF-groups as input data provides a convenient means to organize the groups according to a simple two stage hierarchy consisting of 'background group' and 'substructure'.

Note that it is unlikely to lose any substructure by restricting the search to ordinary FOF-groups. FOF may accidentally link two structures, a case SUBFIND will be able to deal with, but we rarely expect FOF (with a linking length of 0.2) to split a physical structure into two parts.

In SUBFIND, we begin by computing a local estimate of the density at the positions of all particles in the input group. This is done in the usual SPH-fashion, i.e. the local smoothing scale is set to the distance of the $N_{\rm dens}$ nearest neighbor, and the density is estimated by kernel interpolation over these neighbours. The particles may be viewed as tracers of the three-dimensional dark matter density field. We consider any locally overdense region within this field to be a *substructure candidate*. More specifically, we define such a region as being enclosed by an isodensity contour that traverses a saddle point. How could one find these regions? Imagine lowering a global density threshold slowly within the density field. For most of the time, isolated overdense regions will just grow in size as the threshold is lowered, except for the moments when two separate regions coalesce to form a common domain. Note that at these instances the contours of two separate regions join at a saddle point. As a result the topology of the isodensity contour changes as well.

Our algorithm tries to identify all locally overdense regions by imitating such a lowering of the global density threshold. To this end, we sort the finite number of particles according to their density, and we 'rebuild' the particle distribution by adding them in the order of decreasing density. Whenever a new particle i with density ρ_i is considered, we find the $N_{\rm ngb}$ nearest neighbors within the full particle set. Within this set $\mathcal{A}_i$ of $N_{\rm ngb}$ particles, we also determine the subset of particles with density larger than ρ_i, and among them we select a set $\mathcal{B}_i$ holding the two closest particles. Note that this set may contain only one particle, or it may be empty. We now consider three cases:

1. The set $\mathcal{B}_i$ is empty, i.e. among the $N_{\rm ngb}$ neighbors is no particle that has a higher density than particle i. In this case, particle i is considered to mark a local density maximum, and it starts growing a new subgroup around it.

2. If $\mathcal{B}_i$ contains a single particle, or two particles that are attached to the same subgroup, the particle i is also attached to this subgroup.

3. $\mathcal{B}_i$ contains two particles that are currently attached to different subgroups. In this case, the particle i is considered to be a *saddle point*, and the two subgroups labeled by the particles in $\mathcal{B}_i$ are registered as *subhalo candidates*. Afterwards, the particle i is added by joining the two subgroups to form a single subgroup. Note that all subhalo candidates will be examined for self-boundedness later on in the algorithm.

Working through this scheme results in a list of subhalo candidates, which can be efficiently stored in a suitably arranged link-list structure. Note that a given particle can be member of several different subhalo candidates, and that the algorithm is in principle fully capable of detecting arbitrary levels of "subhaloes within subhaloes".

Up to this point, the construction of subhalo candidates has been based on the spatial distribution of particles alone. A more physical definition of substructure is obtained by adding the requirement of self-boundedness. We therefore subject each subhalo candidate to an unbinding procedure to obtain the "true" substructure. To this end, we successively eliminate particles with positive total energy, until only bound particles remain. We perform the unbinding in physical coordinates, where we define the subhalo's center as the position of the most bound particle, and the velocity center as the mean velocity of the particles in the group. We then obtain physical velocities with respect to this group center by adding the Hubble flow to the peculiar velocities. Finally, if more than a minimum number of $N_{\rm ngb}$ particles survive the unbinding, we refer to these particles as a *subhalo*.

There remains the important issue of how one should deal with complications arising from the assignment of particles to several different subhalo candidates. This does not only occur if one deals with genuine "substructure within substructure", but it is actually quite typical for the algorithm. For example, imagine a large halo containing several small subhaloes. Whenever one of the small haloes 'separates' from the main halo, two subhalo candidates are generated according to case (iii) of the algorithm. Each time the larger of these groups describes the bulk of the main halo, which will thus appear several times as a subhalo candidate, although one would like to consider it only once as an independent physical structure.

We approach this issue by considering only the smaller subhalo candidate at each branch of the tree generated by the saddle points. This is based on the notion that we want to examine substructure within some larger object, and this 'background' object is expected to have larger mass than the actual substructure. In addition, we process the subhalo candidates in the inverse sequence as they have been generated, i.e. we work through the saddle points from low to high density. In this way, a smaller subhalo within a larger subhalo will always be processed later than its parent subhalo. As we consider the subhalo candidates in this order, each particle carries a label indicating the subhalo it was last detected to reside in. If in the process it is found to be contained also within a smaller subhalo, this label will be overwritten by the new subgroup identifier.

In this way, the complexity of our analysis is reduced by assigning each particle at most to one subhalo. We are still able to detect a hierarchy of small subhaloes within larger subhaloes, albeit at the expense of reducing the latter by the particles contained at deeper levels of the hierarchy. However, this usually does not affect the corresponding parent subhalo strongly, since the mass of any substructure within a larger group is usually small compared to that of the parent group. Nevertheless, it can happen that the extraction of subhaloes unbinds some additional particles from the parent subhalo. For this reason, we check all the disjoint subhaloes at the end of the process yet again for self-boundedness. Here, we also assign all particles not yet bound to any subhalo to the "background halo" of the group, which we define as the largest subhalo within the original FOF input group, and we check whether they are at least bound to this structure. If not, these particles represent 'fuzz', identified by FOF to belong to the group, but not (yet) gravitationally bound to it.

In summary, SUBFIND decomposes a given particle group into a set of disjoint self-bound subhaloes, each identified as a locally overdense region within the density field of the original structure. The algorithm is spatially fully adaptive, and it has only two free parameters, $N_{\rm dens}$ and $N_{\rm ngb}$. The latter of these parameters sets the desired mass resolution of structure identification, and we usually employ $N_{\rm ngb} = 10$ for this purpose. The results are quite insensitive to the other parameter, the number $N_{\rm dens}$ of SPH smoothing neighbours, which we typically set to a value slightly larger than $N_{\rm ngb}$.

Finally, we want to note that any efficient practical implementation of the algorithm requires the use of hierarchical tree-data structures, and fast techniques to find nearest neighbours and gravitational potentials. For this purpose, we employ techniques lent from our tree-SPH code GADGET.

4.4.3 Examples for substructure identification

In Figures 4.2 and 4.3, we show two typical examples of substructure identification obtained with SUBFIND. In the first case (Fig. 4.2), we selected a small group from the periphery of the S2-cluster, illustrating the tendency of FOF to link structures across feeble particle bridges. By eye, it is apparent that there are at least four independent structures, forming a group that may be about to merge. SUBFIND detects 6 subhaloes in this case, and it cleanly separates the groups prematurely linked together by FOF.

In the second example (Fig. 4.3), we show a somewhat larger object, also extracted at $z = 0$ from the S2-simulation. Again, one can clearly spot substructure embedded in the FOF-group. The algorithm SUBFIND finds 56 subhaloes in this case. The largest one is the 'background' halo, shown in the top right panel of Fig. 4.3. It essentially represents the backbone of the group, with all its small substructure removed. This substructure is made up of 55 subhaloes, which are plotted in a common panel on the lower left.

4.4.4 Subhaloes in the S1, S2 and S3 clusters

Substructure within dark matter haloes is in itself a highly interesting subject that merits detailed investigations. What are the structural properties of subhaloes within haloes? What is their distribution of sizes and masses? What is their ultimate fate? Answers to these questions are highly relevant for a number of diverse topics such as galaxy formation, the stability of cold stellar disks embedded in dark haloes, or to the weak lensing of galaxies in clusters (Geiger & Schneider 1998, 1999).

Tormen et al. (1998) and Ghigna et al. (1998) have addressed some of these questions, and it will be interesting to supplement their work with results obtained from our new group-finding technique. However, such a study is beyond the scope of the present chapter, where we focus on semi-analytic models for the galaxy population of the cluster. We therefore refer a detailed statistical analysis of substructure to future work, and just report the most basic properties of the subhaloes identified in the clusters at redshift $z = 0$.

Especially interesting is a comparison of the mass spectra of subhaloes detected in the three cluster simulations S1, S2 and S3. In this sequence of simulations, the mass and force resolution increases substantially, which clearly leads to a larger number of

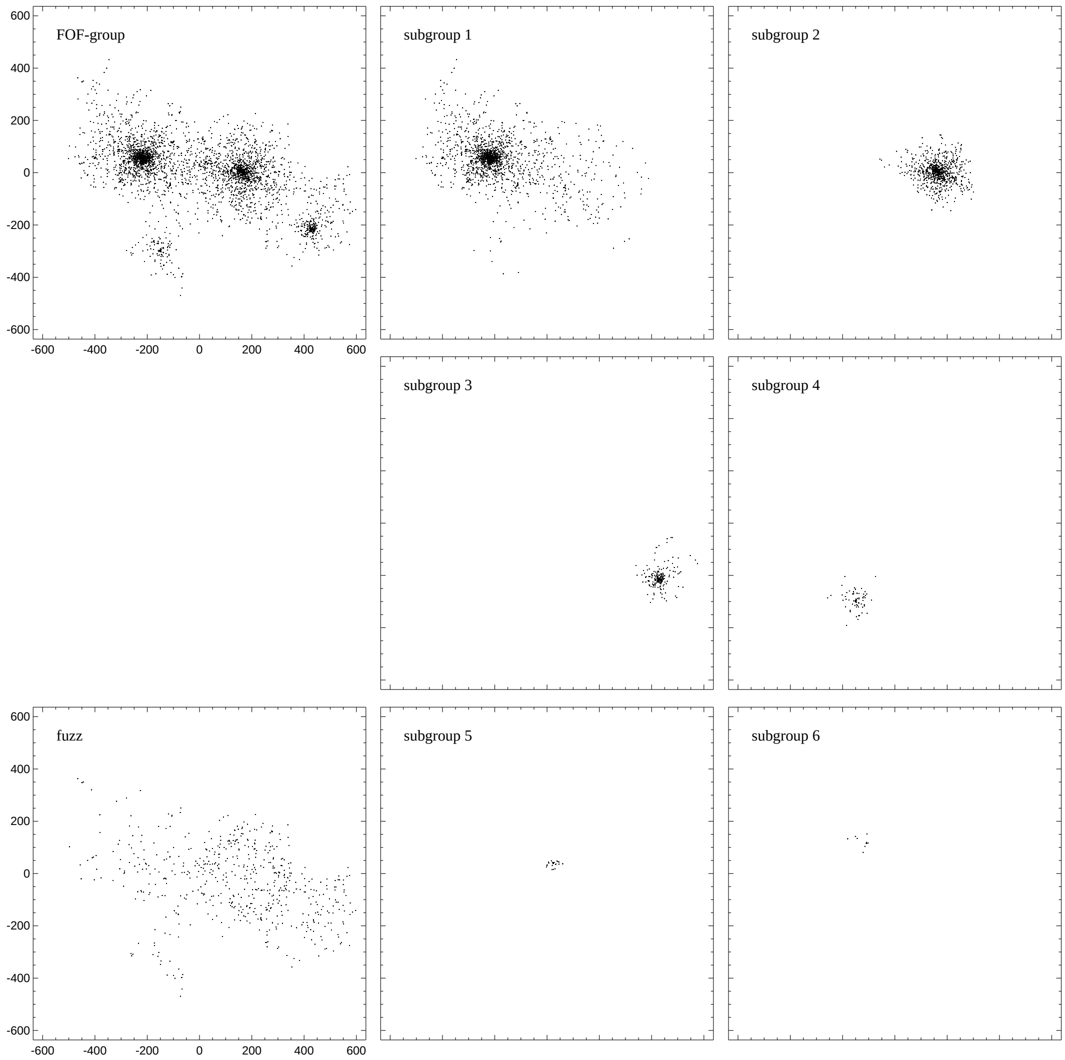


Figure 4.2: Example for a substructure decomposition of a FOF input group. The top left panel shows a small FOF-group (3417 particles) identified at $z = 0$ in the S2 simulation. In this example, FOF has linked structures of comparable mass across thin particle bridges. As shown in the two columns on the right, SUBFIND decomposes this group into 6 constituents. Particles not bound to any of these subhaloes are labeled as "fuzz". Spatial coordinates are given in h^{-1}kpc.

resolved subhaloes. More specifically, we detect 118 subhaloes in the final cluster of S1, 496 in S2, and 1848 in S3. In Figure 4.4, we show a plot of the substructure in the S2-cluster.

The increase of numerical resolution thus unveils an enormous richness of structure in

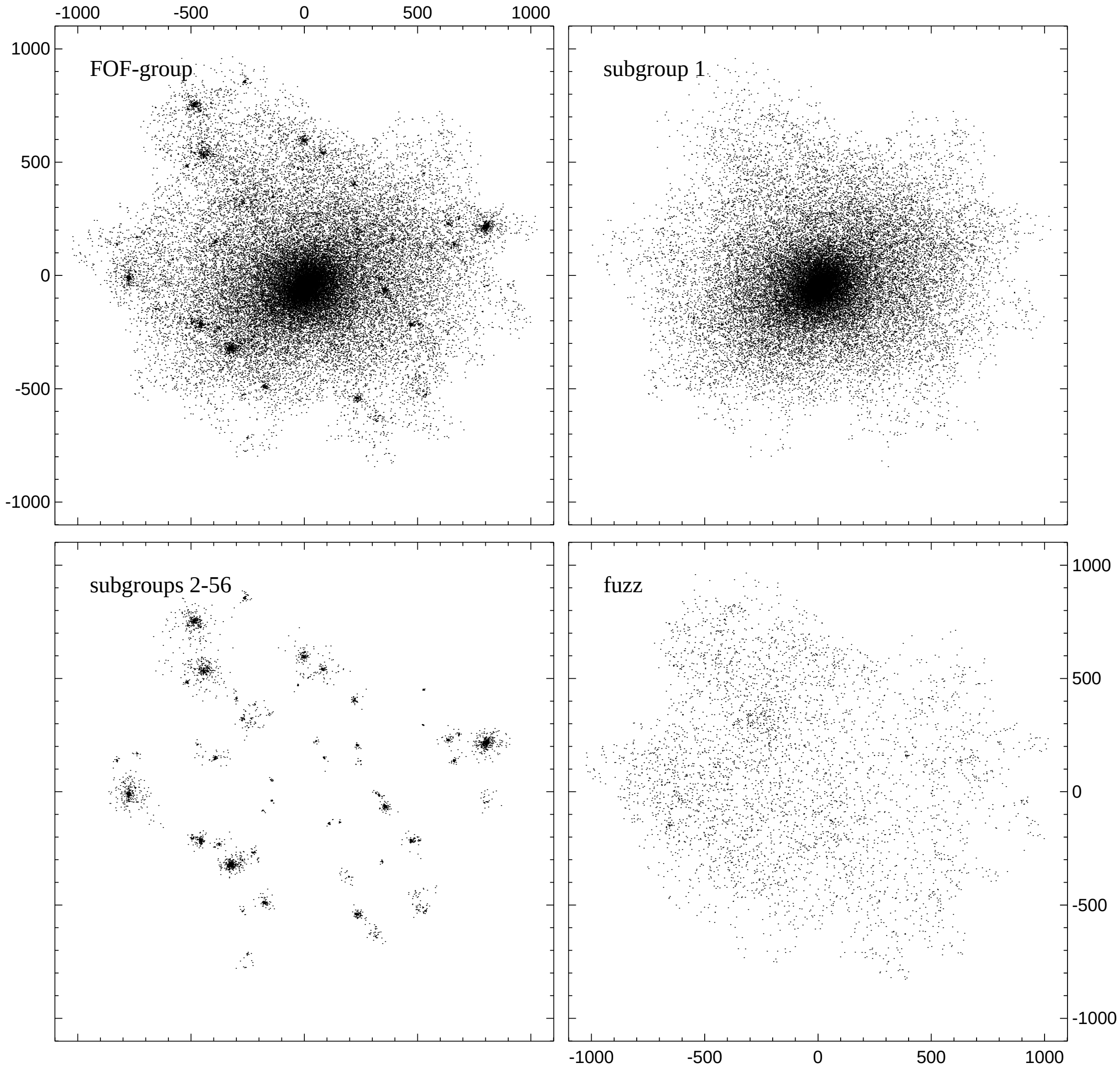


Figure 4.3: Example for a subhalo identification with SUBFIND. The top left panel shows a small FOF-group (44800 particles), identified at $z = 0$ in the vicinity of the S2 cluster. SUBFINDidentifies 56 subhaloes within this group, the largest one forms the background halo and is shown on the top right, while the other 55 subhaloes are plotted on a common panel on the lower left. In this example, the total mass in all the "true" subhaloes 2-56 is about 8% of the group mass. Particles not bound to any of the subhaloes form "fuzz", and are displayed on the lower right. Spatial coordinates are given in h^{-1}kpc.

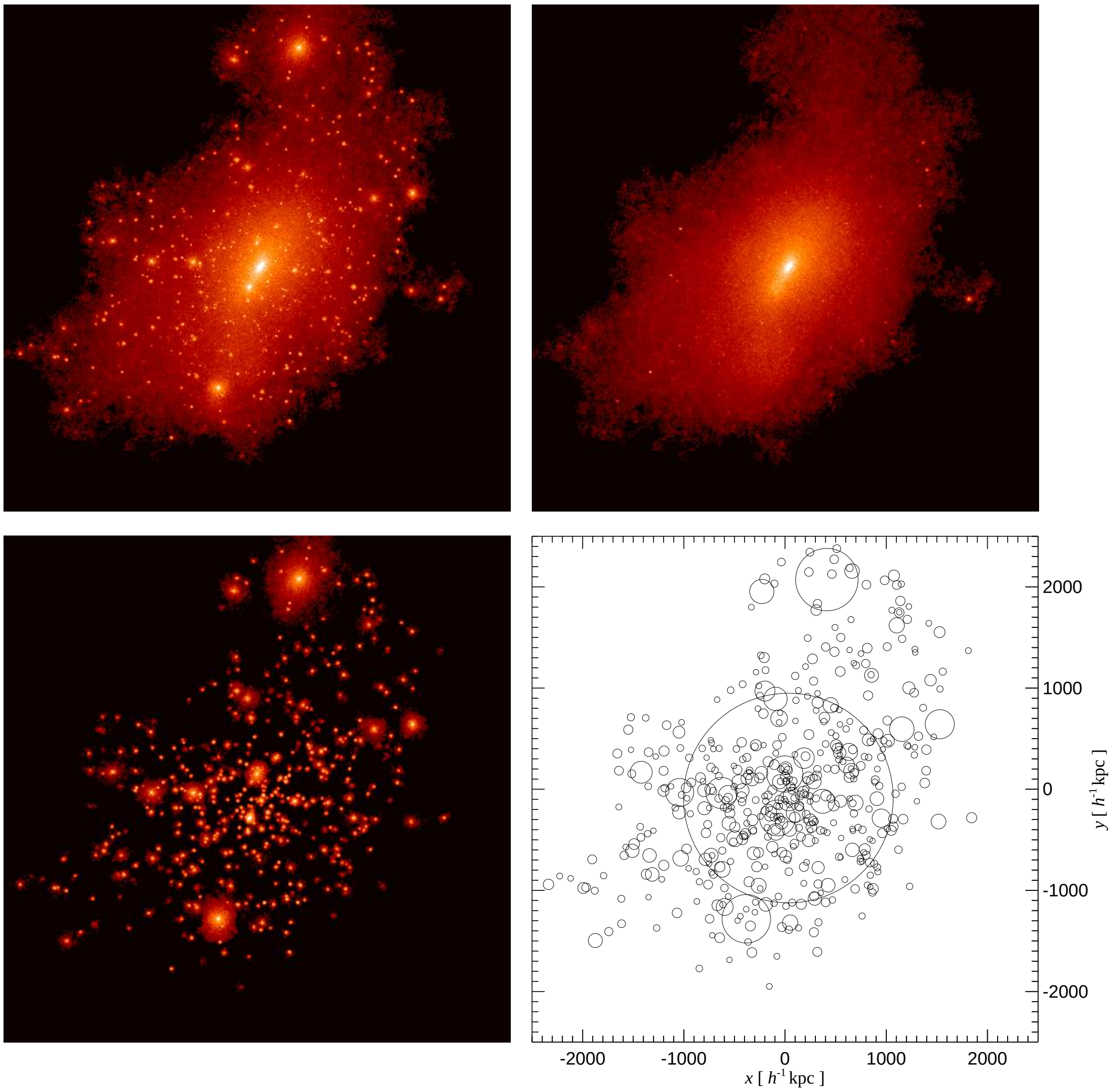
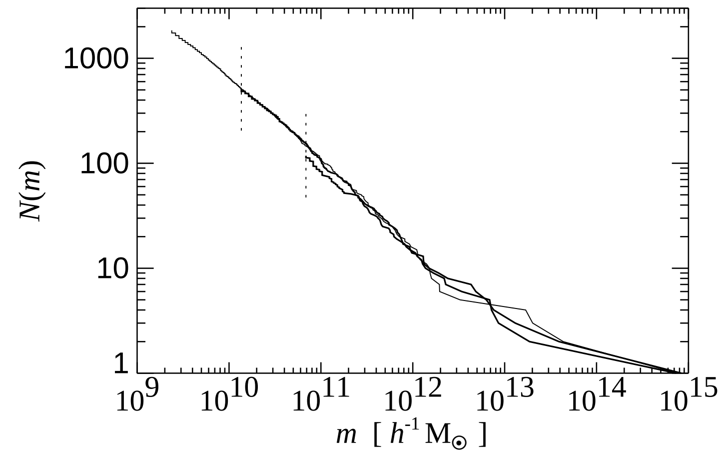


Figure 4.4: Substructure in the S2 cluster at $z = 0$. The top left panel shows a color-coded projection of the FOF-group that contains the cluster. To highlight the substructure, particles have been given a weight proportional to the local dark matter density. In the top right panel we show the largest subhalo identified by SUBFIND, i.e. the background halo. The lower right shows the 495 other subhaloes identified in the object. Finally, on the lower right, we plot in the lower circles at the positions of each identified subhalo, with radius proportional to the third root of the particle number in the subhalo. Note that we even found subhaloes within subhaloes in this example.



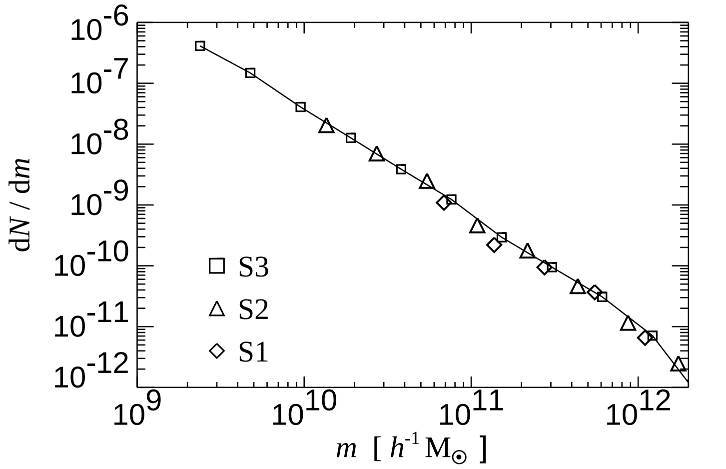


Figure 4.5: Subhalo mass functions in the three clusters S1, S2 and S3 at redshift $z = 0$. In the top panel, we plot the cumulative number $N(m)$ of subhaloes with masses larger than m. The short vertical lines mark the ends of the graphs for the simulations S1 (lowest resolution) and S2 (intermediate resolution). The agreement between the three simulations is remarkably good. This good agreement is also present in the differential mass function $\mathrm{d}N/\mathrm{d}m$, which we show in the lower panel.

the cluster. However, as the comparison of the cumulative and differential mass spectra in Figure 4.5 shows, S1 and S2 are actually well capable of resolving all the objects above their respective resolution limits. Even close to their resolution limit they predict the right abundance of subhaloes of a given mass.

4.4.5 Merging trees using substructure

We now describe our methods to construct merging trees that take the identified subhaloes into account. Note that SUBFIND classifies all particles of a given FOF-group either to lie in a bound subhalo, or to be unbound. Any ordinary FOF-group lacking substructure will also appear in the list of subhaloes identified by SUBFIND, albeit only with its bound subset of particles. Since the requirement of self-boundedness is a reasonable physical condition for the definition of dark haloes, we therefore consider the list of haloes generated by SUBFIND as the 'source list' for our further analysis.

One might then ask why we employ FOF-groups to begin with. We use the FOF-groups as a convenient 'container' to establish a simple two stage hierarchy of haloes. We define the largest subhalo in a given FOF-group as the *main halo* hosting the central galaxy. All other subhaloes found within the group will be considered to be substructure of this main halo.

In the following discussion, the term *subhalo* will refer to any self-bound structure identified by SUBFIND, even if it is just the self-bound part of an ordinary FOF-group that has no real substructure. In addition, we adopt the following definitions:

A subhalo $S_{\rm B}$ at redshift $z_{\rm B}$ is a *progenitor of a subhalo* $S_{\rm A}$ at redshift $z_{\rm A} < z_{\rm B}$, if more than half of the $N_{\rm link}$ most-bound particles of $S_{\rm B}$ end up in $S_{\rm A}$. This definition concentrates on the most-bound core of each structure, and we have found it to be very robust in tracing subhaloes between different output times. One can choose $N_{\rm link}$ as some fraction of the number of particles of subhalo $S_{\rm B}$, say all or half of them. However, such a condition may fail if $S_{\rm B}$ is deprived of its outer halo between two outputs, as it may occasionally happen when a structure of relatively large mass falls into a cluster. We have found that setting $N_{\rm link} = 10$, equal to our lower particle limit for group identification, can satisfactorily treat even these cases. Note that our notion of 'most-bound' refers to the most negative binding energy. SUBFIND automatically stores each subhalo in the order of increasing binding energy to facilitate this kind of linking, i.e. the subhaloes are effectively stored from inside out.

A subhalo $S_{\rm B}$ at redshift $z_{\rm B}$ is a *progenitor of a FOF-group* $G_{\rm A}$ at redshift $z_{\rm A} < z_{\rm B}$, if more than half of the particles of $S_{\rm B}$ are present in $G_{\rm A}$. We also call a FOF-group $G_{\rm B}$ a *progenitor of a subgroup* $S_{\rm A}$, if more than half of the particles of $S_{\rm A}$ are contained in $G_{\rm B}$. Note that in this latter condition we deliberately used the particles of $S_{\rm A}$ to define 'membership' in one of the FOF-group at higher redshift.

We expect that once a self-bound structure has formed, most of its mass will remain in bound structures in the future. However, occasionally it may happen that a group that was just barely above our specified minimum particle number at one output time falls below this limit at the next output time, for example because the group is evaporated by interactions with other material, or because of noise in the identification of groups with size close to our identification threshold. We call such groups *volatile*, and drop them in our analysis. More precisely: All FOF-groups without any bound subhalo are considered to be volatile and disregarded. In addition, if a subhalo is not progenitor to any other subhalo, and not progenitor to any non-volatile FOF-group, it is considered to be volatile too, and dropped from further analysis. Note that basically all subhaloes eliminated in this way have particle numbers very close to the detection threshold.

After the elimination of volatile subhaloes, we link subhaloes between pairs of successive simulation outputs. By construction, every subhalo may have several progenitors, but itself can only be progenitor for at most one other subhalo. In fact, due to the elimination of volatile subhaloes, a subhalo $S_{\rm B}$ will *always* be progenitor to another subhalo, or at least to a FOF-group $G_{\rm A}$. If only the latter is the case, we treat the subhalo $S_{\rm B}$ as a progenitor of the main halo within the FOF-group.

There remains then the important case that a subhalo $S_{\rm A}$ has no progenitor. If it also has no progenitor FOF-group, we call this subhalo a *new structure*, which is considered to have newly formed between the two output times. In the galaxy formation scheme, new galactic 'seeds' will be inserted into these subhaloes.

If however the subhalo $S_{\rm A}$ has no progenitor subhalo, but a progenitor FOF-group $G_{\rm B}$, it is likely that either the subhalo represents a chance fluctuation, or that it was overlooked in the identification process at the time $z_{\rm B}$. In the latter case the corresponding structure will have been merged with the main halo, so we drop these subhaloes. Their absolute number and the total mass contained in them is always very small.

In summary, the above identification scheme allows a detailed tracing of the dark matter merging history tree, from the past to the present. In particular, the scheme is able to deal with groups, that pop into existence, that grow in size by accreting additional background particles, that merge with other haloes of comparable size and lose their identity, or that fall into a larger halo without being destroyed completely. The latter case is particularly interesting, since we expect that a subhalo can survive for some time within the larger structure. However, its mass can be reduced by tidal stripping, and this effect can eventually completely dissolve the structure.

4.5 Substructure and semi-analytic models

4.5.1 Inclusion of subhaloes

One of the questions we want to address in this study is how the inclusion of subhaloes changes the results of semi-analytic models. In order to highlight such changes we will try to modify the 'standard' scheme, which is based on the work of KCDW, in a minimum fashion when subhaloes are included.

Note that a large variety of semi-analytic "recipes" are conceivable, and that the results can depend quite strongly on the detailed assumptions made, as has been recently highlighted by KCDW. Semi-analytic models thus cannot eliminate uncertainties arising from poorly constrained physics. However, they are ideal tools to explore the relative importance of various model ingredients, and hence to constrain their relevance for observed trends in observational properties.

We now describe the changes in our semi-analytic methodology when subhaloes are included. From here on, we will refer to the formalism of KCDW, which only works with FOF groups, as the 'standard'-scheme, and to the analysis that includes substructure as the 'subhalo'-scheme. Please note that the word 'standard' is not meant to imply that the corresponding procedure has the status of a well-established, widely used method in the field – it is just used to refer to the methodology recently developed by KCDW.

Our most important change concerns the definition of the galaxy population at each output time. We define the largest subhalo in a FOF-group to host the central galaxy of the group, and its position is given by the most-bound particle in that subhalo. All gas that cools within a FOF-group is funneled exclusively to the central galaxy. This definition of 'central galaxy' thus corresponds to the one adopted in the standard analysis.

However, with respect to the population of galaxies orbiting in the halo, we can now distinguish between *halo-galaxies* and satellite galaxies. Here we have coined yet another term; 'halo-galaxies' are attached to the most bound particle of the remaining subhaloes in the FOF-group. These halo-galaxies had been proper central galaxies in the past, until their halo has fallen into a larger structure, but the core of their dark halo is still intact, and thus allows an accurate determination of the position of the halo-galaxy within the group. These halo-galaxies may still be viewed as 'central galaxies' of their respective subhaloes, but they are not fed by a cooling flow any more, since their subhalo is not the largest one in the FOF group.

Finally, when two (or more) subhaloes merge, the halo-galaxy of the smaller subhalo becomes a satellite of the remnant subhalo. These satellites are treated like in the standard analysis. Their position is tagged by the most-bound particle identified at the last time they were still a halo-galaxy, and they are assumed to merge on a dynamical friction timescale with the halo-galaxy of the new subhalo they now reside in. We need to introduce such satellites in the subhalo-scheme in order to account for actual mergers between subhaloes, and also for the finite numerical resolution of our simulations, which limits our ability to track the orbits of subhaloes once their mass has fallen below our resolution limit. It also allows to make direct contact with the standard-scheme in the limit of poor resolution. Note that the class of halo-galaxies is absent in the standard analysis, where all of these objects would be treated as satellites.

In the subhalo scheme, we define the virial mass $M_{\rm vir}$ of a subhalo simply as the total mass of its particles. We then assign a virial radius by assuming that the halo has an overdenity 200 with respect to the critical density, i.e. $R_{\rm vir}^3 = GM_{\rm vir}/(100H^2)$, and we define a virial velocity as $V_{\rm vir}^2 = GM_{\rm vir}/R_{\rm vir}$.

4.5.2 Mergers inside the cluster

When the analysis of subhaloes is included, a halo falling into a larger group can be followed as an independent physical entity along its orbit in the larger halo. The effects of orbital decay will thus be properly modeled, up to the point when the subhalo is eventually lost, either due to effects of finite resolution, or due to actual physical disruption.

In this context, the interesting possibility of mergers between halo-galaxies *inside* the cluster arises. Such mergers are believed to occur only rarely due to the large velocity

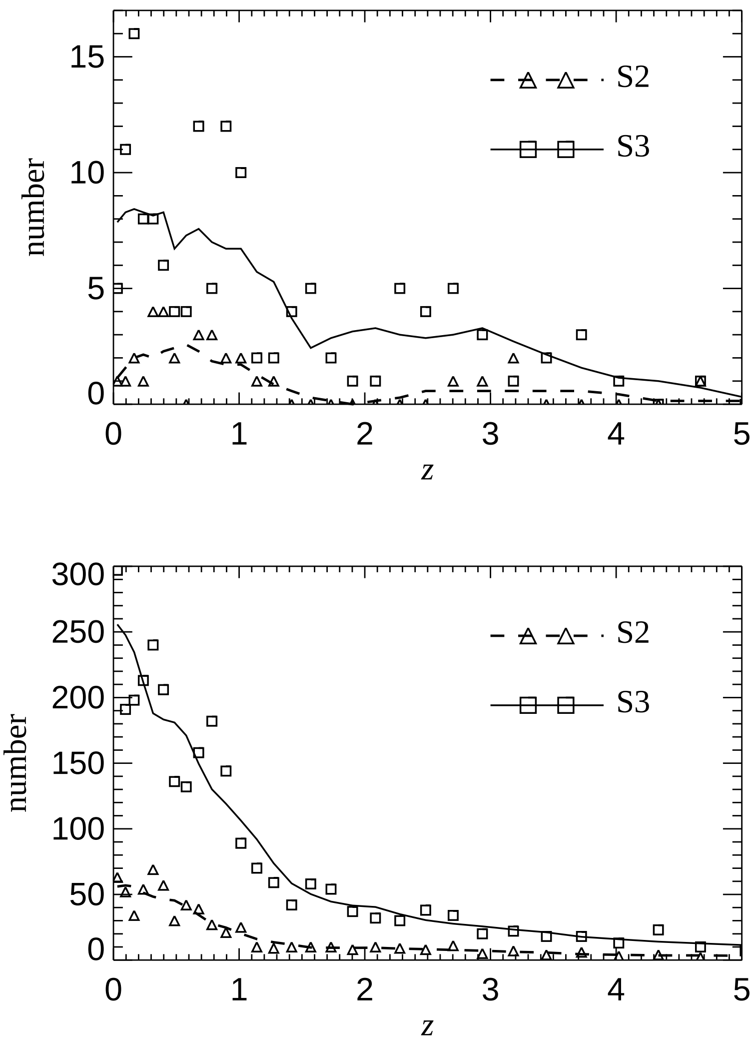


Figure 4.6: Number of merger events between subhaloes *inside* the cluster (top panel). Note that we here define the 'cluster' at some redshift z as the largest progenitor halo of the final cluster. We define an inner merger event by requiring that a subhalo of the cluster, which is not the main subhalo, has at least two progenitor subhaloes residing in the largest cluster progenitor at the previous output time. We plot the number of such events measured in the cluster simulations S2 and S3 for pairs of successive output times. The solid and dashed lines are boxcar averages of the measured points. For comparison, we show in the bottom panel the number of 'normal' mergers with the cluster center. Here we count the number of subhaloes inside the largest cluster progenitor that merge with the main subhalo of the cluster.

dispersion of cluster galaxies. Note however, that our subhalo algorithms are capable of dealing with these events, if they occur.

We are now in a position to find out whether we actually detect such events. Using the merging history trees of the subhalo scheme, we can simply look for situations where a halo-galaxy of the cluster has two (or more) progenitor subhaloes that *both* reside in the largest progenitor FOF-halo of the cluster. In the top panel of Figure 4.6 we show the number of such events measured in the cluster simulations S2 and S3 for pairs of successive output times. While the absolute number of mergers between subhaloes is relatively low, they clearly do occur with a small rate. The larger number of mergers detected in S3 merely reflects the better mass resolution of this simulation. As expected, the strong increase of the abundance of subhaloes towards smaller mass scales is accompanied by an increase of the absolute number of mergers inside the cluster. Note that the standard-scheme of semi-analytic modeling does not account for these merger events, but the subhalo-scheme does.

It is interesting to compare the detected rate of mergers between halo-galaxies with the number of 'normal' mergers, i.e. mergers between a subhalo and the cluster center. To obtain a fair comparison with the events measured above, we now count how many subhaloes within the largest cluster progenitor merge with the cluster center, i.e. are progenitor to the main subhalo of the cluster. We plot the corresponding number of events in the bottom panel of Figure 4.6. For the last 30 outputs of S2 ($z < 5.2$), the total number of mergers 'inside' is 32, while the 'normal' ones occur 645 times. For S3, these numbers are 143 and 2780, respectively. In both cases, the ratio between these two numbers is close to 20.0, suggesting that roughly one out of 20 subhalo mergers occurs with another halo-galaxy instead of with the central object.

4.6 Results

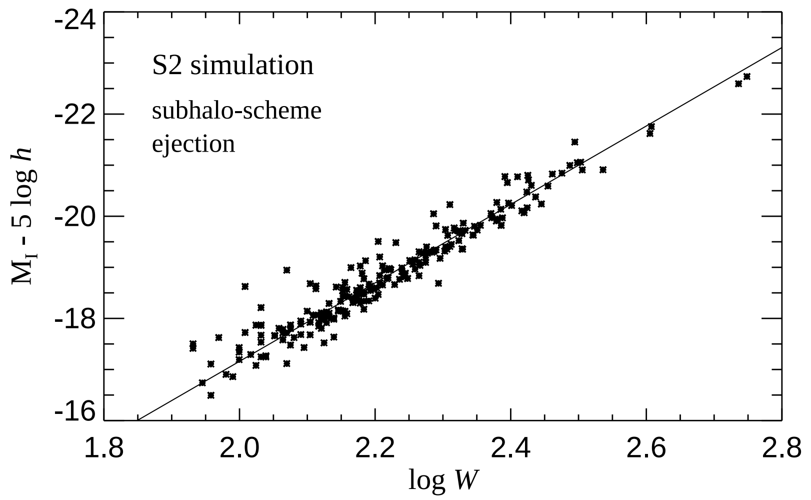
4.6.1 Tully-Fisher relation

We use the velocity-based I-band Tully-Fisher relation

$$M_I - 5\log h = -21.00 - 7.68\,(\log W - 2.5) \tag{4.14}$$

of Giovanelli et al. (1997), and the requirement of a gas mass of $\sim 0.6 \times 10^{10}\,h^{-1}\mathrm{M}_\odot$ in 'Milky-Way' haloes, to normalize our models. We consider two variants for the implementation of feedback, the 'ejection' model, where gas is blown out of small haloes, and the 'retention' model, where reheated gas is always kept within the halo.

In Figure 4.7, we show the best-fit Tully-Fisher relations obtained for these two models, applied to the S2-cluster using the 'subhalo'-scheme. In these plots, we only considered central galaxies of haloes that are peripheral to the cluster, but that are not contaminated by heavier boundary particles. We also applied a morphological cut $1.2 \le M_{\rm bulge} - M_{\rm total} \le 2.5$, approximately selecting Sb/Sc galaxies. In Table 4.2, we list the model parameters thus obtained. In these initial models, we have assumed that



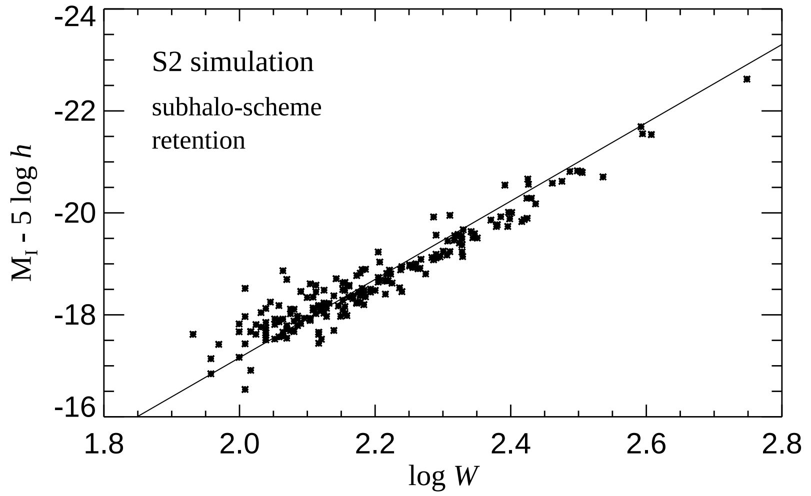


Figure 4.7: The *I*-band Tully-Fisher relation for Sb/Sc galaxies in the S2 simulation, obtained with the subhalo-scheme using ejection and retention feedback, respectively. The galaxies have been selected as central galaxies of uncontaminated haloes in the periphery of the cluster. The solid line represents the recent measurement by Giovanelli et al. (1997).

all metals remain in the cold gas phase. Effects of chemical enrichment of the hot halo will be discussed separately below. In the following, we will use the same normalization also for our other cluster simulations, and for the ‘standard’ semi-analytic scheme.

The Tully-Fisher relations we obtain exhibit remarkably small scatter. This is a

Table 4.2: Numerical parameters adopted for our semi-analytic models. α is the star formation efficiency, ϵ the efficiency of feedback by supernovae, f_b the baryon fraction, and y the mean effective yield of metals per stellar generation.

Model	α	ϵ	f_b	y
ejection	0.10	0.05	0.20	0.02
retention	0.10	0.15	0.20	0.02

result of the tight coupling we assumed between the sizes and circular velocities of the disks of spiral galaxies and the masses of their dark haloes. Note that additional scatter can be expected from the distribution of spin parameters of dark haloes, which gives rise to variations of the disk sizes associated with a halo of a given mass (Mo et al. 1998).

The ejection model fits the slope of the observed TF-relation relatively easily. However, the retention model is less effective in suppressing star formation in low mass haloes. For the same value of ϵ, the retention scheme produces therefore a shallower Tully-Fisher relation than the ejection model. As a result, a much larger value for ϵ is needed to bring the retention model in agreement with the observed steepness of the TF-relation. This strong feedback reduces the overall brightness somewhat, an effect that could be easily compensated for by slightly larger values for α or f_b. Note that a feedback efficiency of $\epsilon = 0.05$ means that feedback will be very efficient in haloes of virial velocity below $V_{SN} = 130\,\mathrm{km\,s^{-1}}(\epsilon/0.05)^{1/2}$. In such haloes, one will have $\Delta M_{\rm reheat} \geq \Delta M_\star$, i.e. the mass of reheated gas exceeds that of newly formed stars.

It is also interesting to compare the Tully-Fisher relations obtained for the different cluster simulations, and for the 'subhalo' and 'standard' variants of semi-analytic modeling (see Figure 4.8). For the sake of brevity, we restrict the comparison between the 'subhalo' and 'standard' schemes to the S2-cluster, and between the simulations S1-S3 to the subhalo-scheme with ejection feedback. In general, there is good agreement between the two variants of semi-analytic modeling, and between the different simulations, despite their great differences in numerical resolution. The 'subhalo' scheme is slightly more conservative at the lower mass end than the standard method. This is because volatile haloes are excluded in the subhalo analysis, and the virial mass of haloes is assigned slightly differently in this scheme. As a result, an Sb/Sc galaxy in the subhalo scheme cannot have a virial mass less than the mass of 10 simulation particles, while close to the resolution the standard analysis sometimes assigns virial masses that can be even below that limit.

Comparing the two lower panels of Figure 4.8 with the left panel in Figure 4.7, one also sees that the agreement between the simulations S1, S2 and S3 is very encouraging. For the same choice of free parameters, the slopes of the Tully-Fisher relations agree very well. However, a detailed examination of the zero-points also shows that there is a slight trend towards fainter zero-points in the sequence of simulations S1 to S3. As we will show below, this effect is a consequence of a systematic increase of the brightness of

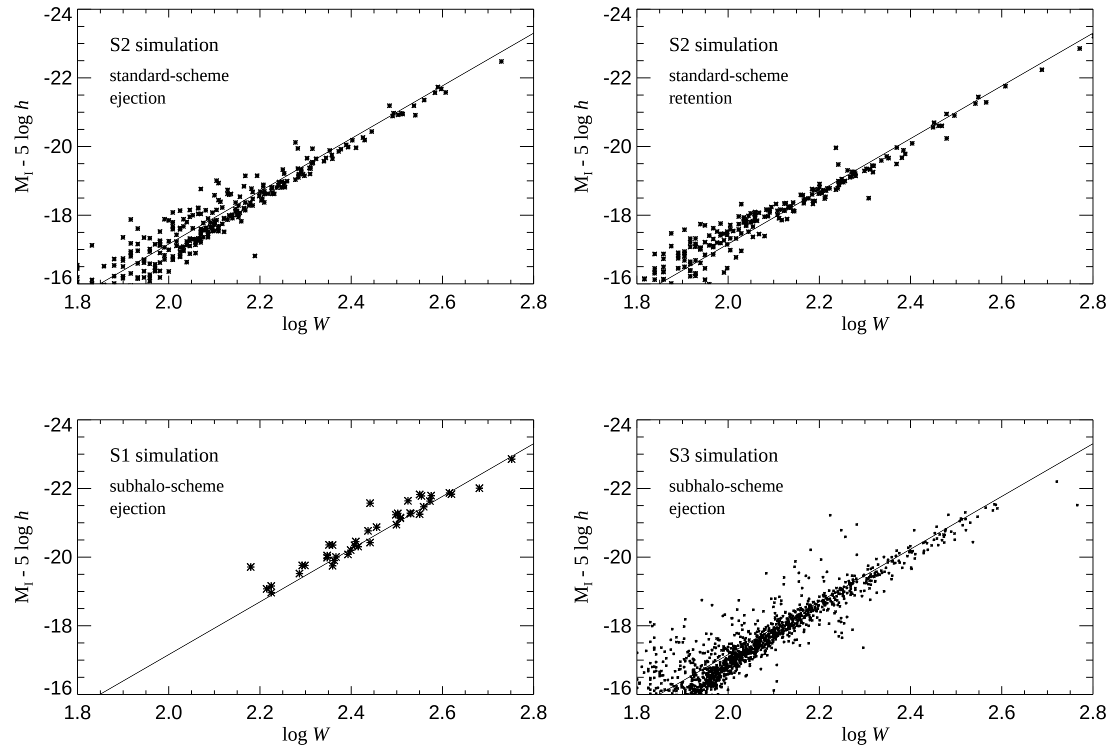


Figure 4.8: The top two panels show the I-band Tully-Fisher relations obtained for the S2 simulation using the 'standard'-methodology of KCDW. These may be compared with the results for the subhalo-scheme displayed in Figure 4.7. The bottom two panels give the TF-relations for the S1 and S3 simulations using the subhalo/ejection-variant of the semi-analytic modeling.

central galaxies due to overmerging processes. In the subhalo scheme, this effect becomes weaker for higher numerical resolution.

4.6.2 Cluster luminosity function

In Figure 4.9, we show the B-band luminosity function of the S2-cluster obtained with the new 'subhalo' methodology, and we compare it to the result of the 'standard' semi-analytic recipe of KCDW. We plot the number of cluster-galaxies in bins of size 0.5 mag, and we fit Schechter functions to the counts. The standard prescription results in a relatively steep slope of -1.35 at the faint-end, and the "knee" of the Schechter function is not well defined. This is just another reincarnation of a well-known problem of semi-analytic studies. They usually predict too many galaxies both at the faint and the bright end, i.e. their luminosity function is not curved enough. These deficiencies can be partly cured by invoking additional physics like dust obscuration, or by alluding

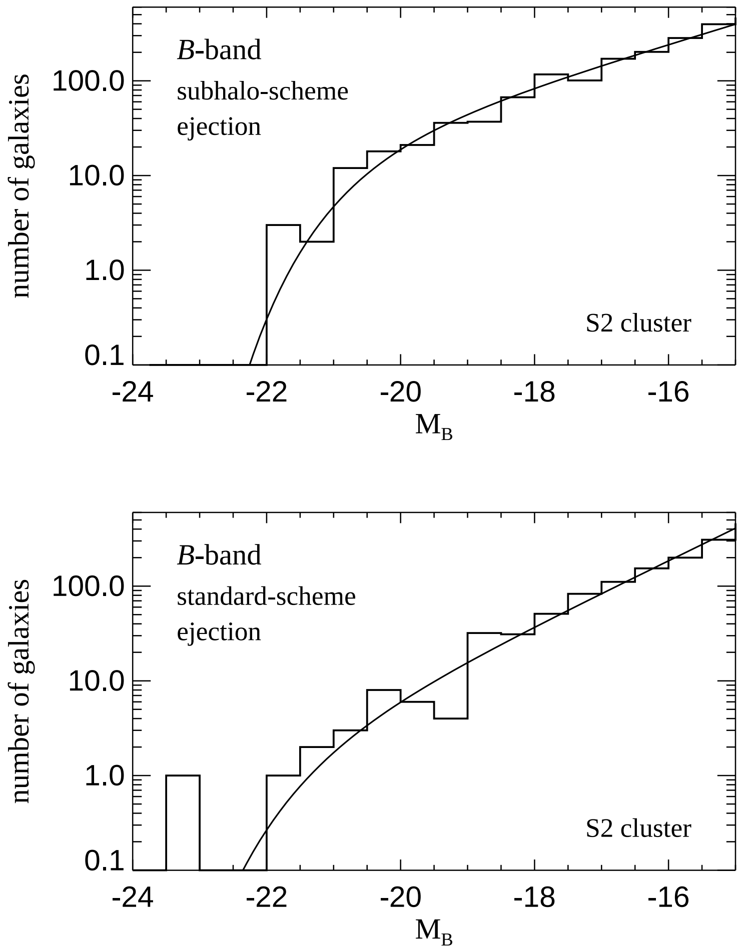


Figure 4.9: The B-band cluster luminosity function obtained for the S2 simulation using the subhalo scheme (top) and the standard scheme (bottom). We here show results for the ejection feedback, and note that the results for the retention model are very similar. The solid lines are Schechter function fits to the histograms of bin-size 0.5 mag. For the subhalo model, the faint end slope is $\alpha = -1.21$, and the turn-off is at $M_\star \simeq -21.6$. The standard scheme results in a steeper slope of $\alpha = -1.35$, and the characteristic magnitude is not well defined. Note that the standard model produces a galaxy of magnitude -23.4, which appears too bright even comparing to the brightest cD galaxies.

to stronger feedback processes. However, most semi-analytic studies have not been successful in simultaneously achieving agreement with the Tully-Fisher relation and the luminosity function.

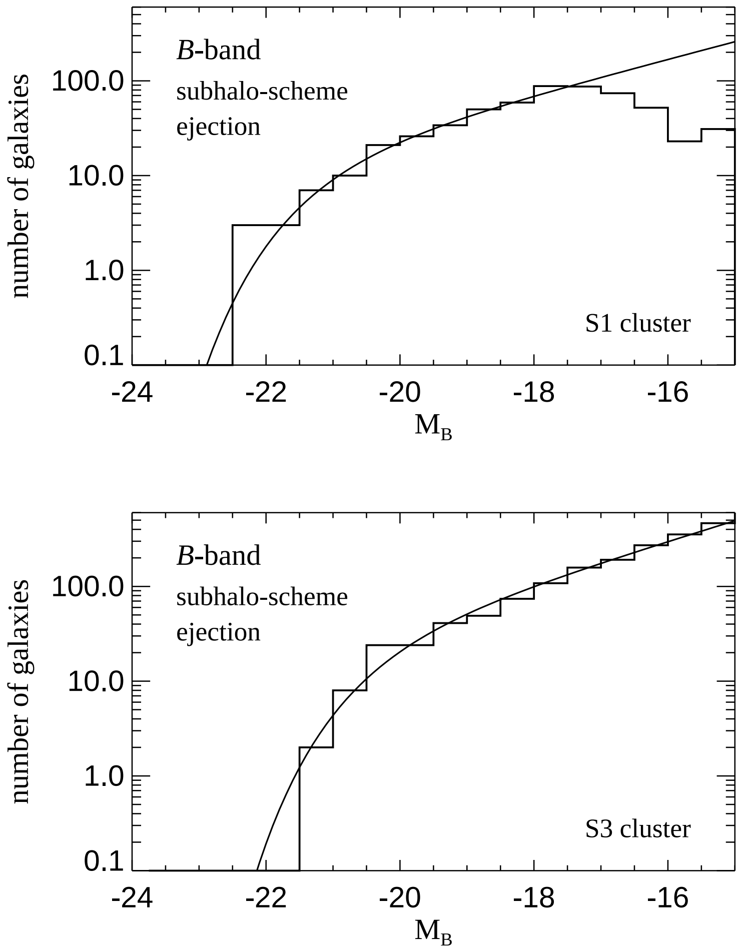


Figure 4.10: The B-band cluster luminosity function obtained for the S1 and S3 simulations using the subhalo scheme and ejection feedback. The solid lines are three-parameter Schechter function fits to the histograms of bin-size 0.5 mag. For the S3 model, the faint-end slope of the fit is $\alpha = -1.22$ with $M_\star = -21.4$, while for S1 we obtain $\alpha = -1.19$ and $M_\star = -22.2$.

Compared to the field, the luminosity functions of clusters tend to be considerably steeper, and a slope of -1.35 can be accommodated easily with existing data. However, the standard-scheme also produces a brightest cluster galaxy with luminosity in excess of most normal cD galaxies. For example, in the S2 cluster, the standard-scheme produces a central galaxy with B-magnitude -23.4.

In contrast to that, the luminosity function obtained with the subhalo formalism has a more reasonable shape. There are more $L_\star$-galaxies, resulting in a flatter faint-end slope of $\alpha = -1.21$, and in a more pronounced "knee", allowing a decent fit with a Schechter function. In addition, the very large brightness of the central galaxy has disappeared. The overall shape of the resulting luminosity function is in reasonably good agreement with known cluster luminosity functions.

What is causing this big difference between the subhalo-scheme and the standard scheme? Note that the excessive brightness of the central galaxy in the latter model cannot be due to an overcooling problem. We have already cut-off the cooling flow for central galaxies in haloes with a virial velocity larger than $350\,\mathrm{km\,s^{-1}}$. Furthermore, the cooling model was essentially the same for both schemes, which is also reflected in their similar overall mass-to-light ratios.

We think that the most important difference between the 'standard' semi-analytic scheme and the 'subhalo'-methodology is that the latter allows a much more accurate estimate of the actual merging rate in any given halo. Recall that one critical assumption in the study of KCDW has been that the position of a satellite galaxy can be traced by a single particle identified as the most-bound particle of the satellite's halo just before it was accreted by a larger system. KCDW also assumed that the time of survival of a satellite can be estimated using a simple dynamical-friction formula. However, if this description turned out to be too crude, the excessive brightness of the central galaxies could simply result from an overestimate of the overall merging rate, or from merging the 'wrong' galaxies with the center, i.e. bright galaxies that actually still orbit in the halo and have not yet merged with the central galaxy.

Before we investigate this issure in more detail below, we plot in Figure 4.10 the cluster luminosity functions for the S1 and S3 clusters, obtained with the subhalo scheme and ejection feedback. Comparing these luminosity functions with the corresponding result obtained for S2 (Fig. 4.9), we clearly see that with decreasing resolution the brightness of the first ranked cluster galaxies increases. Again, we think that this is a reflection of the overmerging problem that troubles simulations with insufficient numerical resolution, and this effect is also responsible for the weak trend in the zero-points of the Tully-Fisher relations (see Figures 4.7 and 4.8).

We now directly test whether such an 'overmerging' problem is responsible for the difference between the subhalo scheme and the standard formalism. To this end we try to answer three questions: (1) How many of the galaxies present in subhaloes of S2 at $z = 0$ are prematurely merged with the center in the standard scheme? (2) How well is the position of subhaloes given by the single particles used to track satellites in the standard scheme? (3) How do the merging timescales estimated in the standard scheme compare with the actual survival times of infalling halos?

To address the first two of these questions, we follow each subhalo of the cluster back in time along its merging history until it is a main halo itself for the first time. The corresponding FOF-halo will host a central galaxy in the standard scheme, and the descendant of this galaxy at $z = 0$ should directly correspond to the galaxy of the originally selected subhalo.

Among the 494 subhaloes identified in the S2-cluster at $z = 0$, we find in this way that 23 of them do not have a directly corresponding satellite galaxy in the 'standard'-

formalism. The satellites that should correspond to these 23 subhaloes have prematurely disappeared by merging processes with the central galaxy. Note that the total number of mergers with the central galaxy in the standard scheme is 94, while this number is 125 in the subhalo formalism. This suggests that the overall rate of mergers is not too high in the standard scheme. However, in the standard scheme the central galaxy accrets 18 galaxies with stellar mass more than $10^{10}\,h^{-1}\mathrm{M}_{\odot}$, among them 6 galaxies with stellar mass larger than $10^{11}\,h^{-1}\mathrm{M}_{\odot}$. On the other hand, in the subhalo formalism there is only 1 merger involving more than $10^{11}\,h^{-1}\mathrm{M}_{\odot}$, and 6 with more than $10^{10}\,h^{-1}\mathrm{M}_{\odot}$. This means that a larger number of very bright galaxies is merged with the central galaxy in the standard scheme. We also note that the 23 subhaloes that appear to have been merged prematurely in the standard scheme tend to be quite bright. If we take the results for the subhalo-model and add the luminosities of the corresponding halo-galaxies to that of the central galaxy, its brightness increases from -21.64 to -23.56 mag, quite close to the -23.4 obtained in the standard-scheme. It thus appears that the excessive brightness of the central cluster galaxy in the standard scheme is mainly caused by an underestimate of the merging timescales for some fraction of the bright galaxies.

With respect to positions of satellites, we find that a fraction of 86.8% of the subhaloes in S2 still contain the most-bound particle that is used in the standard-scheme to track the position of the satellite. This number is 86.6% in S1, and 82.2% in the S3 cluster. Using just a single particle identified at a time before a structure was accreted onto the cluster can thus provide a good estimate of the subhalo's position within the cluster halo at later times.

We now come to the third question raised above. Consider a subhalo that merges at a certain time with another, larger subhalo. As above, we can follow this subhalo back in time along the path defined by its most massive progenitor subhalo at each output time. At some point, the subhalo's progenitor found in this way will be for the first time the main subhalo of a FOF halo. We call the lookback time until this happens the *survival time* of the FOF halo. Note however, that this survival time is really a lower limit to the 'true' survival time, since the corresponding subhalo might have been lost prematurely due to limited resolution effects. To determine the survival times, we process all subhaloes that vanish in mergers, and we store the maximum survival time thus obtained for each FOF halo.

It is now interesting to compare these survival times with the estimates of the merging timescale, as given by equation (4.9). We compute the merger timescales exactly in the way as it is done in the standard scheme for FOF haloes that merge with another halo, but are not a most massive progenitor themselves. In Figure 4.11, we compare the survival time with the estimated merger timescale for these FOF haloes. There is clearly a correlation between the two, albeit with very large scatter. Note in particular that all the points below the diagonal line correspond to merger events that are predicted to happen *too early* in the standard methodology – the subhaloes that trace the cores of the corresponding FOF haloes actually survive for a longer time. In the standard scheme, these premature mergers make the central galaxy too bright.

On the other hand, there is also a substantial number of estimated merger timescales

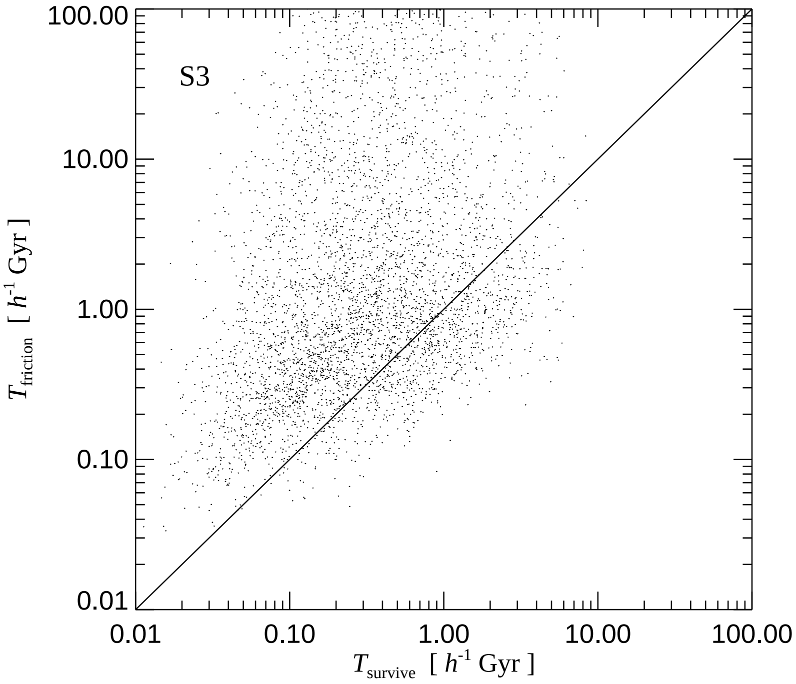


Figure 4.11: Scatter plot comparing the estimated merger times based on the dynamical friction formula (4.9) with the survival times derived by explicitly tracking the merging history of subhaloes. To avoid overcrowding, we have just plotted one fifth of the measured points.

that are very large, in fact much larger than a Hubble time. Note that for the choices we adopted in equation (4.9), the dynamical friction formula is essentially the current dynamical time of virialized haloes times the mass ratio of the halo and its infalling satellite. Hence the large estimates of merger timescales arise when the mass of the satellite is much smaller than that of the halo. However, the dynamical friction estimate does not include the effects of tidal truncation and disruption, which act to reduce the lifetime of satellites. The dynamical friction estimate might therefore systematically overestimate the time of survival of small satellites. In fact, the results of Tormen et al. (1998) indicate that the dependence of the merger timescale on the mass ratio, measured just before the satellite falls in, is much weaker than linear. Presently it is unclear, whether an improved parameterization of the merger timescales in the standard scheme can produce results as good as those obtained by explicitely following the subhaloes.

4.6.3 Morphology density relation

Dressler (1980) has shown that the relative frequency of elliptical galaxies is higher in denser environments. In particular, in the cores of clusters, spirals are quite rare, while they are the dominant type in the field, and in the Universe as a whole. Whitmore et al. (1993) have argued that this morphology-density relation presumably reflects a more fundamental morphology-clustercentric relation; the correlation between morphology and clustercentric separation seems tighter than that between morphology and projected density. However, this assertion is still controversial.

In Figure 4.12, we show the cumulative number of galaxies of different Hubble as a function of distance from the cluster center. The top panel gives all the galaxies with $M_{\rm B} < -16$, while in the middle panel we just show the bright galaxies with $M_{\rm B} < -19$. It is seen that the elliptical galaxies (the $T < -4$ bin) are more concentrated towards the center than the spirals. In fact, from the cumulative distribution it can be infered that the three-dimensional density profile of the ellipticals has a steeper slope than that of the sprirals. It is also interesting to note that the bright galaxies are primarily ellipticals, as seen in the middle panel. In the top two panels all galaxies have been counted, i.e. halo-galaxies that are attached to a subhalo as well as ordinary satellites. In the bottom panel, we just show the distribution of the subhalo-galaxies. Note that there is one subhalo as close as $55\,h^{-1}$kpc to the center, but all other subhalos are more than $\sim 200\,h^{-1}$kpc away.

In Figure 4.13 we show the morphology-clustercentric relation obtained for the S2-cluster, using our subhalo-scheme with ejection feedback. For the classification of galaxies, we applied the same cuts in T-space as above. Note that the fraction of elliptical galaxies strongly rises towards the center of the cluster, while that of the spirals declines accordingly. The quantitative strength of these trends is in good yet not perfect agreement with the results of Whitmore et al. (1993).

We think that this is a very interesting result. Recall that our morphological modeling is solely based on the merging history of galaxies. This simplistic model for morphological evolution already suffices to establish a pronounced morphology density relation. This shows that the morphology density relation is built-in at a very fundamental level in hierarchical theories of galaxy formation.

In passing we note that in the standard recipe, where satellites are just traced by single particles once they have fallen into a larger halo, the morphology-clustercentric relation is also present, yet not so well defined. In Figure 4.14, we show the morphology-clustercentric relation obtained for the standard methodology. Within the cluster, the trend of morphology with radius is somewhat weaker. Note that in Figure 4.13 we only counted galaxies that are still attached to a subhalo. If we also include the satellites, the results become more similar to the standard scheme, i.e. the morphology-clustercentric relation is better defined in the subhalo population of galaxies than in the satellites. This is probably caused by a more accurate treatment of the merging history of those galaxies that can still be traced by a subhalo.

Figure 4.13 also shows that the fraction of S0's in our cluster is too small compared

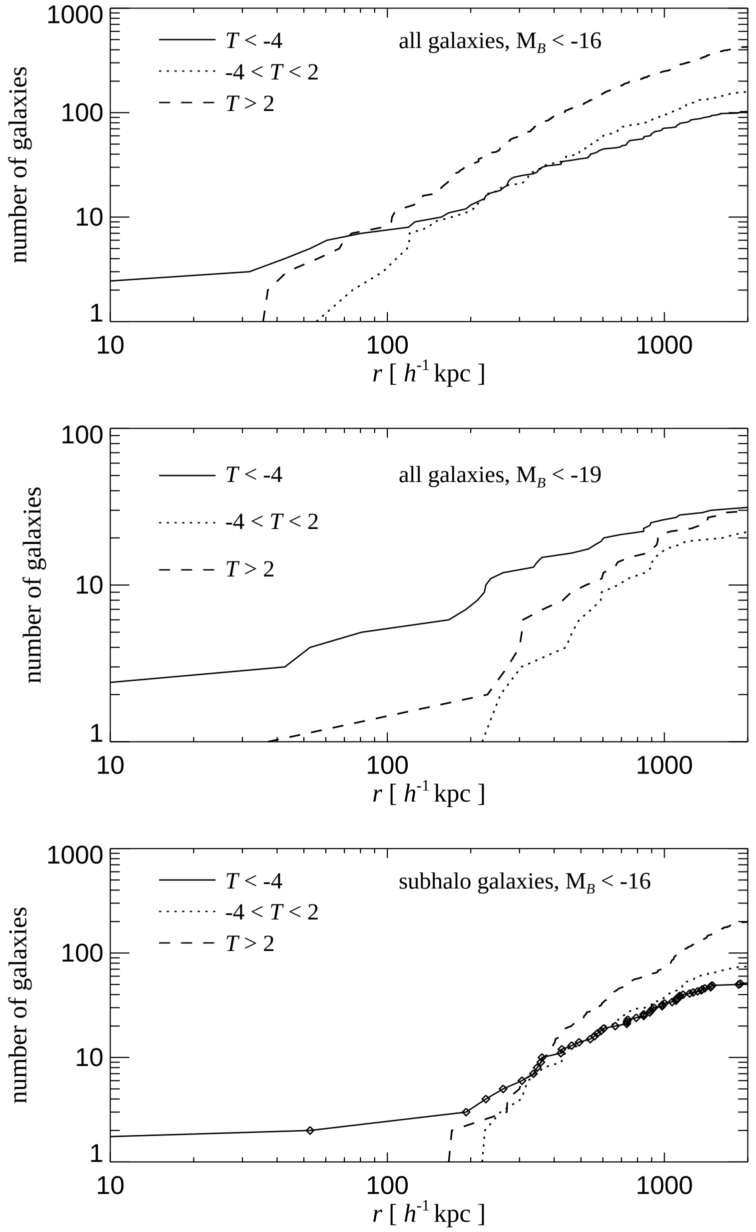


Figure 4.12: Cumulative number of galaxies in the S2 cluster as a function of radial distance to the cluster center. In the top panel, we count all galaxies with $M_\mathrm{B} < -16$, binned into three classes of different Hubble type. In the middle panel, only the bright galaxies with $M_\mathrm{B} < -19$ are shown, while the bottom panel gives only those galaxies that are still represented by a subhalo. Here, the diamonds mark the distance of individual galaxies. The results shown are for the S2 cluster using ejection feedback.

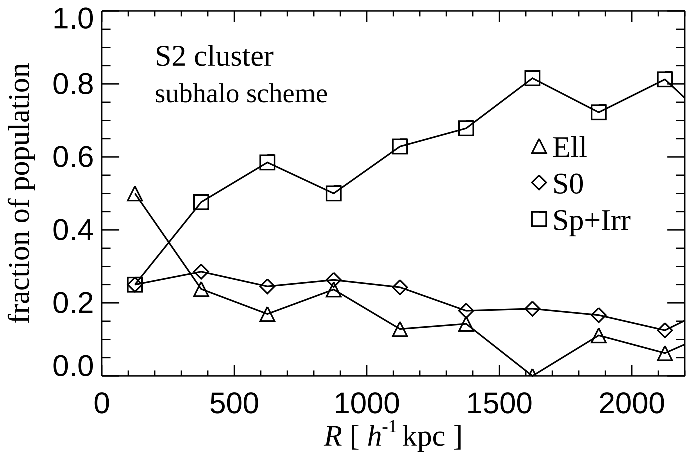


Figure 4.13: Morphological mix of galaxies as a function of clustercentric radius. We here show results for the S2-simulation, using the subhalo modeling with ejection feedback. Only galaxies that are still attached to a subhalo and are brighter than $M_{\mathrm{B}} = -16$ have been used in this plot.

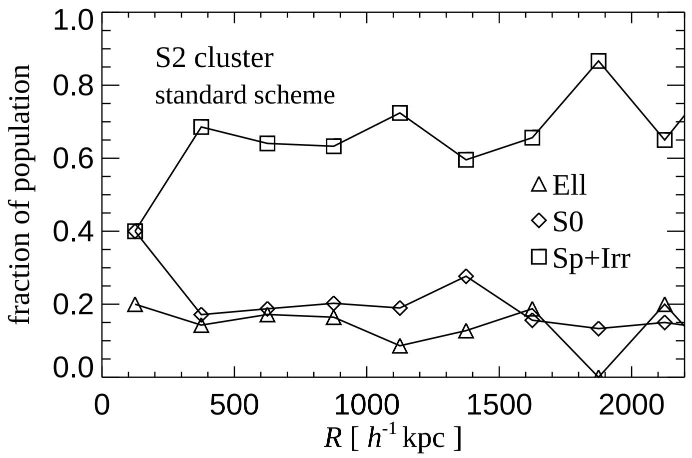


Figure 4.14: Morphological mix of galaxies as a function of clustercentric radius. We here show results for galaxies brighter than $M_{\mathrm{B}} = -16.3$ obtained with the standard scheme of KCDW using ejection feedback.

with that seen in nearby clusters. However, disk galaxies orbiting in a cluster experience high-speed encounters with other galaxies. Together with global cluster tides this can drive a morphological evolution towards spheroids, a process termed galaxy harassment (Moore et al. 1996, 1998). It has been suggested that the spiral galaxies seen abundantly in clusters at moderate redshift are harassed and slowly transformed to S0's at the present time. In its current form, our modeling does not account for this effect, and it might thus be invoked to explain our deficit of S0 galaxies.

4.6.4 Cluster mass-to-light ratio

The observed mass-to-light ratios of clusters of galaxies are known to be much larger than those of individual galaxies, and this fact has been recognized early on as strong evidence for the existence of large amounts of dark matter in clusters. Typical measured values for the cluster mass-to-light range from $\Upsilon_B = 200\,h\,\Upsilon_\odot$ to $\Upsilon_B = 400\,h\,\Upsilon_\odot$ in the B-band. For the Coma cluster, Kent & Gunn (1982) measured $\Upsilon_B = 360\,h\,\Upsilon_\odot$, while X-ray data seem to point to a lower value. For example, Cowie et al. (1987) inferred a mass-to-visual-light ratio for Coma as low as $\Upsilon_V = 180\,h\,\Upsilon_\odot$.

The galaxy population constructed for the S2 cluster using the subhalo model has a total magnitude of $M_B = -25.7$. For its total mass of $8.36 \times 10^{14}\,h^{-1}\mathrm{M}_\odot$, the cluster mass-to-light ratio is thus $\Upsilon_B = 590\,h\,\Upsilon_\odot$. In the V-band, we obtain $M_V = -26.6$ and $\Upsilon_V = 450\,h\,\Upsilon_\odot$. These mass-to-light ratios may be slightly too large compared to the mean of observational results, although there are also measurements that are as large as our values. For example, Kent & Gunn (1983) obtained $\Upsilon_V = 600\,h\,\Upsilon_\odot$ for the Perseus cluster.

However, small changes of our parameters can easily enhance the overall brightness of the cluster, and thus reduce the mass-to-light ratios to a desired value. One simple possibility for that is to change the conversion factor between the virial velocity and the circular velocity. If the circular velocity of a disk in a dark matter halo of given virial velocity is on average 15% larger than we assumed, the normalization to the Giovanelli et al. (1997) Tully-Fisher relation results in an overall increase in brightness of 0.5 mag, and a corresponding reduction of the mass-to-light ratio to 63% of its old value. Another possibility is to be less restrictive in imposing a cut-off for the cooling-flows. In Figure 4.15 we show the luminosity function of the S3 cluster when the corresponding limit is increased from $350\,\mathrm{km\,s^{-1}}$ to $450\,\mathrm{km\,s^{-1}}$. This only affects the central galaxies of very large haloes and makes them brighter. As a result, the mass-to-light ratio goes down to $\Upsilon_V = 387\,h\,\Upsilon_\odot$. For a cut-off at $500\,\mathrm{km\,s^{-1}}$, the central cluster galaxy reaches a brightness of $M_B = -23.09\,\mathrm{mag}$, which might be interpreted as a cD galaxy, and the mass-to-light ratio becomes $\Upsilon_V = 343\,h\,\Upsilon_\odot$.

4.6.5 $B - V$ color distribution

In Figure 4.16, we show the $B - V$ color distribution of galaxies brighter than -19.7 mag in the S2 simulation. A corresponding plot has been shown by KCDW, and we here

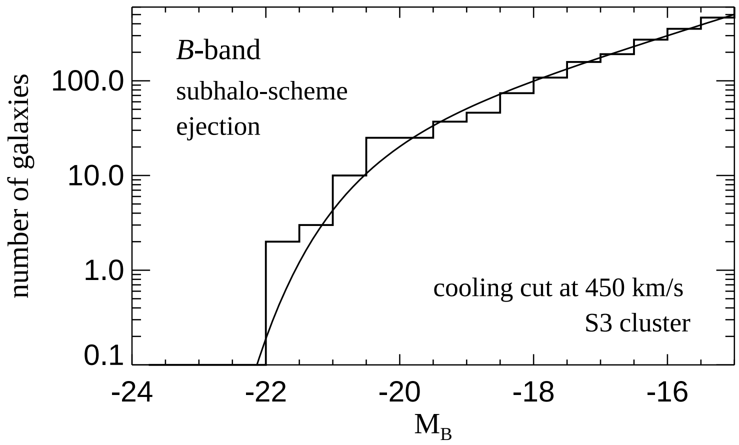


Figure 4.15: B-band cluster luminosity function obtained with the subhalo-methodology for the S3 cluster. We have here cut off the cooling flows in haloes with virial velocities above $450\,\mathrm{km\,s^{-1}}$, i.e. we have been less restrictive than in our normal models (compare Figure 4.10) where the corresponding limit was $350\,\mathrm{km\,s^{-1}}$. As a result of this change, the central galaxies in large haloes become brighter, leading to a reduction of the mass-to-light ratio of the cluster. For easier comparison, the Schechter-function drawn in this plot has the parameters of the fit obtained for S3 in Figure 4.10.

obtain similar result. The distribution of the $B-V$ colors is bimodal, with two peaks at $B-V \simeq 0.7$ and $B-V \simeq 1.0$. By plotting the color distribution for individual morphological types, it becomes apparent that this dichotomy mainly reflects the differences in star formation history between elliptical and spiral galaxies. The recent star formation in spirals makes them blue, while the older stellar populations of ellipticals give them redder colors.

We show the projected spatial distribution of red ($B-V \geq 0.85$) and blue ($B-V < 0.85$) galaxies in Figure 4.17. In this plot, we visualize the dark matter mass distribution inside a cube $4\,h^{-1}$Mpc on a side, centered on the S3-cluster, as a gray-scale image. Overlaid on this picture are red and blue circles at the positions of the semi-analytic galaxies. The area of each circle is proportional to the blue-light luminosity of the corresponding galaxies. It can be seen that most of the galaxies in the cluster are red, and that these red galaxies are heavily concentrated towards the center. Note that only very few galaxies in the region of the cluster are blue, although the blue spirals form the majority of galaxies in the field surrounding the cluster. The concentration of red elliptical galaxies towards the center is consistent with our result for the morphology-clustercentric relation.

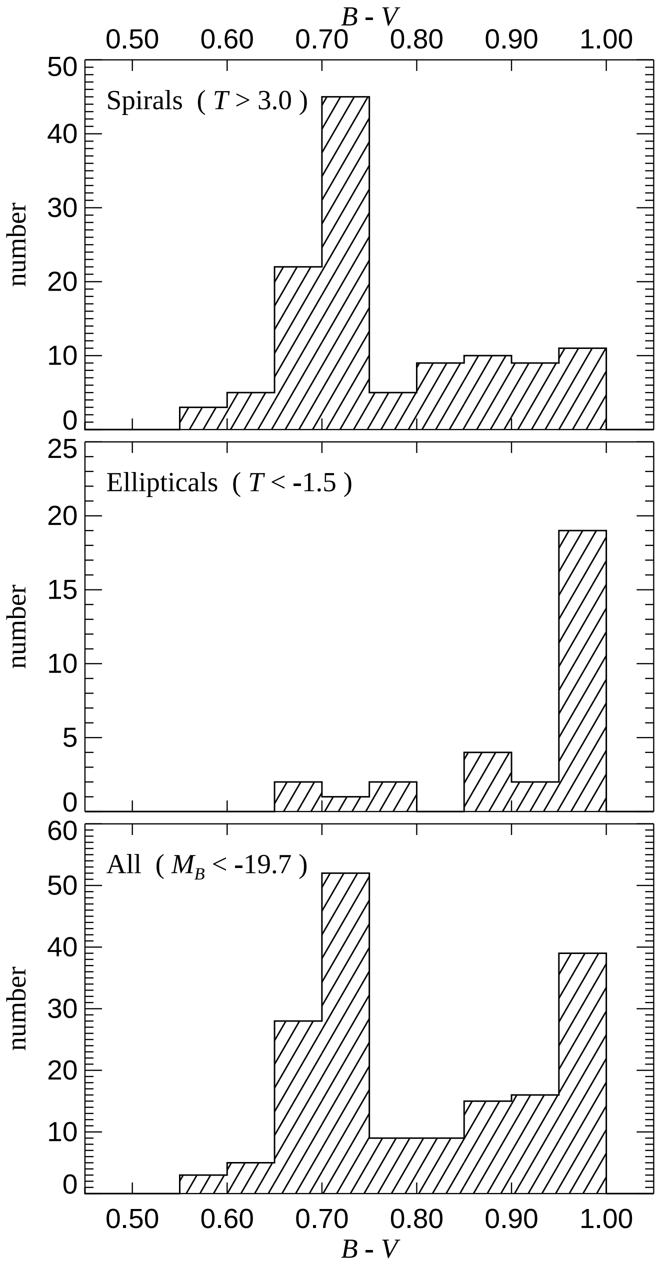


Figure 4.16: The $B-V$ colors of galaxies brigther than -19.7 mag in the S2 simulation. The dichotomy of the distribution mainly reflects the difference in the stellar populations of elliptical and spiral galaxies. Spirals are significantly bluer due to the their recent star formation, while ellipticals have older stellar populations, resulting in redder colors.

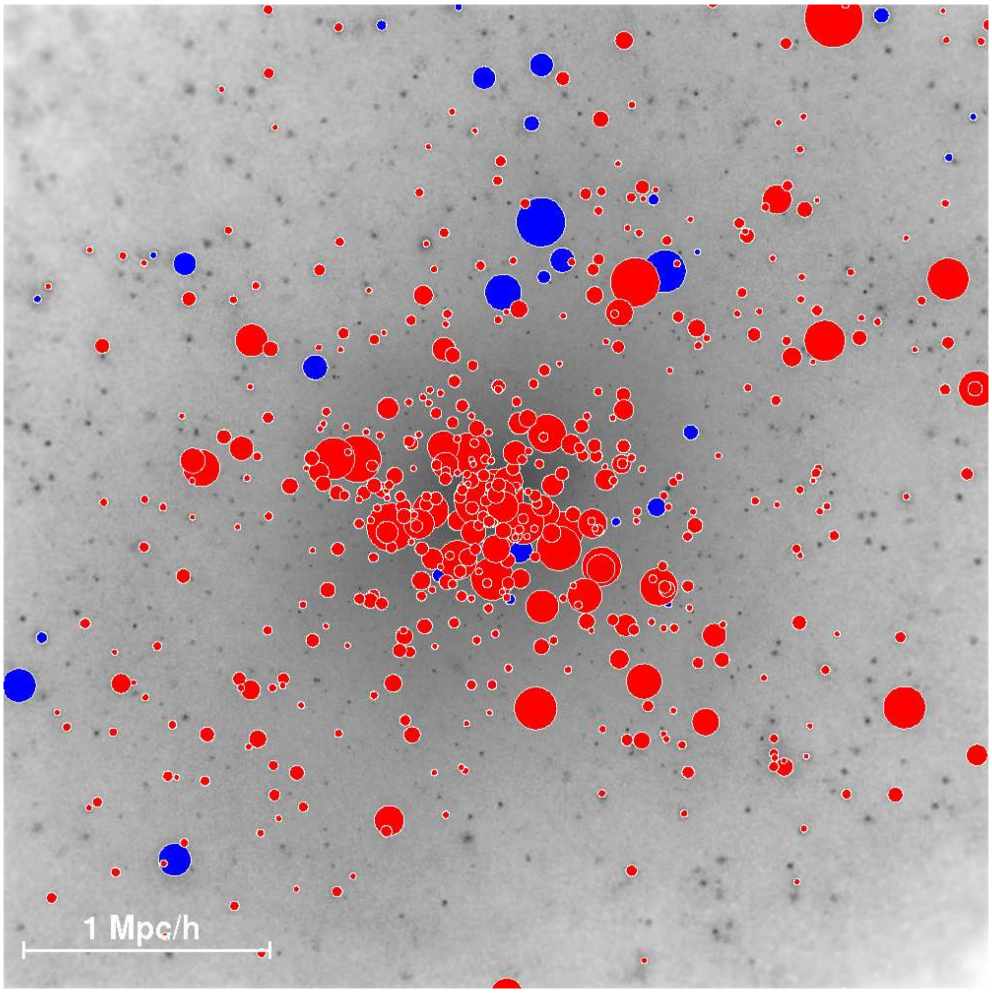


Figure 4.17: The projected distribution of galaxies in the S3 cluster. The circles mark the positions of red ($B-V \geq 0.85$) and blue ($B-V < 0.85$) galaxies, in a box $4\,h^{-1}$Mpc on a side around the cluster center. The area of each circle is proportional to the B-band luminosity, and the gray-scale image in the background visualizes the dark matter density. Only galaxies with $M_{\rm B} < -16$ are shown.

4.6.6 Star formation history

In Figure 4.18, we plot the evolution of the star formation rate for all galaxies (and their progenitors) that belong to the cluster at the present epoch, and we compare this to the history of the star formation rate of 'field' galaxies in the periphery of the cluster. We here show results for the S2 cluster using the subhalo model with ejection feedback. Note that the star formation rate is approximately proportional to the Hα luminosity and to the UV-luminosity, respectively. We find that the total star formation rate of the

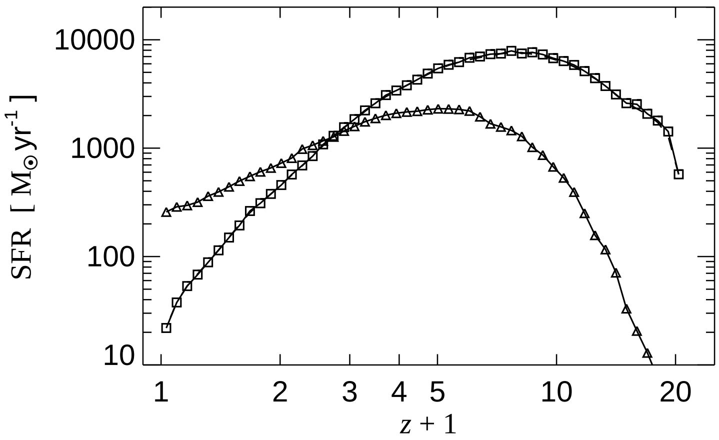


Figure 4.18: Evolution of the star formation rate of all galaxies that end up in the S2-cluster at the present time (small boxes). The small triangles give the corresponding rates for the field galaxies in the periphery of the cluster. Here, the vertical normalization is arbitrary, since it depends on the volume defined as the 'field'. The results shown are for the subhalo-model with ejection feedback. Note that the contribution of the starburst mode in major mergers is negligible for the total star formation rate of the cluster galaxies.

cluster galaxies is almost unaffected when we turn off starbursts in major mergers. In the current model, only a small fraction of all stars is formed in the starburst mode.

The star formation rate of field galaxies has recently been investigated by Madau et al. (1998), suggesting a sharp rise of the star formation rate (SFR) from $z = 0$ to a peak at $z \sim 1.5$, and then a fall towards higher redshift. The peak of the star formation history of our cluster galaxies lies at significantly higher redshift, around $z \simeq 6 - 7$, and the fall towards higher redshift appears somewhat gentler than suggested by the analysis of Madau et al. (1998). On the other hand, the SFR of our field galaxies peaks at lower redshift, around $z \simeq 3 - 4$, and the decline towards the present time is relatively shallow. At present there are still large uncertainties in the observational determination of the cosmic star formation history, primarily due to the dust corrections that have to be applied to optical/UV data. However, both the newest corrections for dust extinction, and the recent determinations of the star formation history in the near-infrared (Blain et al. 1999) move the peak of the observed star formation history to $z \simeq 3 - 4$.

From Figure 4.18 it is already apparent that the stars in the cluster form considerably earlier than those in the field. To highlight this point, we plot in Figure 4.19 the cumulative star formation rate of the cluster galaxies, and compare it to that of the field galaxies. Already at $z \simeq 4$, 12 Gyr ago, half of the stars in the cluster have been formed, while the field galaxies have produced half their stars not until $z \simeq 2$. Note that the

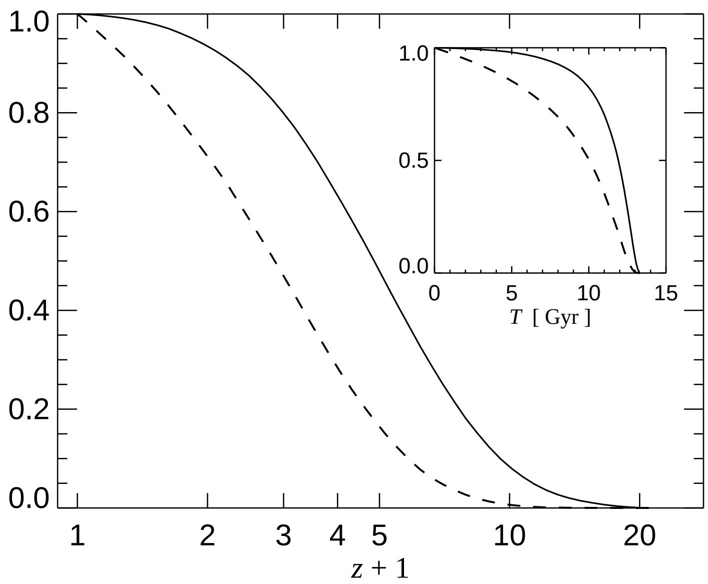


Figure 4.19: The fraction of stars in the cluster that have formed up to redshift z. Half of the stars are in place by $z \simeq 4$. The dashed line gives the same quantity for all galaxies in uncontaminated haloes in the surroundings of the cluster. On average, their stars form considerably later. The small inset gives the same quantities as a function of lookback time for our choice of cosmology. We here show results for the S2 cluster using the subhalo-scheme with ejection feedback.

actual cluster forms at much lower redshift; it has assembled half of its final mass at $z = 0.66$.

4.6.7 Effects of chemical enrichment

In the results presented so far, we have assumed that all the metals released in supernovae explosions remain in the cold phase of the interstellar medium. The metallicity of the halo gas thus remained zero, and the cooling function used was effectively always that for primordial abundance.

However, if feedback does indeed reheat some of the cold gas, there should be a transport of metals along with the reheated, enriched gas. Note that in the retention model, we mix the reheated gas into the hot halo. Feedback should therefore pollute the halo with metals. This is not necessarily the case in the models with ejection feedback, because here the reheated gas is assumed to leave the galaxy.

In Figure 4.20, we show the Tully-Fisher relation obtained for the S2-simulation when retention feedback is used and the metal transport is taken properly into account. Since we modeled feedback to become more efficient in low-mass haloes, the gaseous halo is

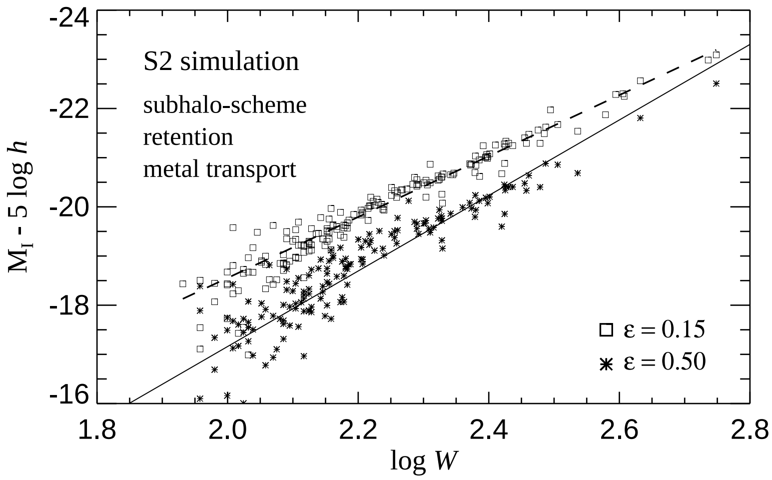


Figure 4.20: Tully-Fisher relation for the S2-simulation, including metal enrichment of the hot gaseous halo. We here assumed that feedback reheats enriched cold gas, and mixes it homogeneously into the halo. For the same choice of parameters as used in our standard retention model, the model with metal transport leads to considerably brighter galaxies, and to a shallower Tully-Fisher relation. The corresponding results are shown as small boxes. To bring the Tully-Fisher relation again into rough agreement with the observational result by Giovanelli et al. (1997) (solid line), a very large feedback efficiency of $\epsilon = 0.5$ is needed. The corresponding galaxies are plotted as stars.

more efficiently enriched with metals in small galaxies. As a result, the brightness of low mass galaxies is more strongly increased by metallicity-boosted cooling rates than that of high mass galaxies. Keeping the parameters α and ϵ fixed, this tilts the Tully-Fisher relation to a shallower slope, apart from increasing the overall brightness of the galaxies. The TF-relation shown in Figure 4.20 is based on the choice of parameters adopted so far for the retention scheme (see Table 4.2). To bring the TF-relation in rough agreement again with the observations of Giovanelli et al. (1997), we have to make the suppression of star formation in low mass haloes stronger. In fact, we need a rather extreme value of $\epsilon = 0.5$ to at least achieve approximate agreement with the observed relation. For this choice, the zero-point of the TF and the overall brightness of the cluster is about right, although the TF relation is still slightly shallow. It can be further steepened by increasing ϵ, albeit at the expense of reducing the overall brightness too much.

It thus seems that metal enrichment of the haloes of low-mass galaxies amplifies the tendency of the retention scheme to produce low-velocity galaxies that are too bright. The ejection model fairs considerably better in this respect. Due to its stronger suppression of star formation in small galaxies, and the absence of metal deposition in the halo, it nicely produces a steep enough Tully-Fisher relation using a moderate value for ϵ. On the other hand, if there was a mechanism that enhances the efficiency of metal enrichment in haloes with *high* circular velocity, such a process could also help to steepen the Tully-Fisher relation. It would act like positive feedback in the larger galaxies. However, presently it is unclear whether such a physical process exists, and if so, how it might operate.

4.7 Discussion

In this study, we have used cosmological N-body simulations combined with semi-analytic techniques to construct the galaxy population of a rich cluster of galaxies. The very high resolution of our simulations allowed us to extend the methodology for galaxy formation of KCDW to the regime of substructure within virialized systems. Our goal in this work has therefore been twofold. We wanted to develop the necessary technical machinery for a galaxy formation scheme that works with subhaloes, and we wanted to compare its results with those obtained with the 'standard' techniques by KCDW.

The detection of subhaloes within haloes is a technically difficult problem, and has only been addressed very recently. Several working algorithms have been described, but none seems to be an ideal solution yet. In this chapter, we presented our new subhalo finding algorithm, SUBFIND. It just relies on the particle positions and velocities of a single output time, and it reliably identifies locally overdense, self-bound particle groups within larger systems as 'subhaloes'. SUBFIND can also detect hierarchies of 'haloes within haloes', a feature that is not well represented in alternative techniques.

The mass spectrum of subhaloes in our cluster appears to be close to a power-law. Our set of simulations allowed a direct study of resolution effects, and it is interesting to note that the lower-resolution simulations predict the right abundance of subhaloes for any given mass above their respective resolution limit.

Using the detected subhaloes, we have shown how they can be traced from simulation output to output. Just like KCDW, we analysed 51 simulation snapshots, logarithmically spaced in expansion factor from $z = 20$ to $z = 0$. This large number of output times together with the very high resolution of our simulations allowed a study of the merging history of the dark matter in unprecedented detail. For example, we detected more than 1840 subhaloes in the final halo of the S3 cluster, and we found around 140 events of mergers or captures between subhaloes orbiting inside the progenitor halo of the cluster.

Using a very small set of modifications, we have adapted the semi-analytic scheme of KCDW to include the analysis of subhaloes. In both schemes, the agreement with the observed Tully-Fisher relation is very good. However, the inclusion of subhaloes results in a substantial improvement of the cluster luminosity function of the models. In the standard scheme, the first ranked cluster galaxies become too bright, while this problem goes away in the subhalo-scheme. We have shown that this is mainly due to overmerging that takes place in the standard model. Here, too many bright galaxies that fall into the cluster are prematurely merged with the central galaxy. The direct tracing of subhaloes until their eventual disappearance allows a more precise estimate of the actual merger rate within haloes. As a result, the luminosity function becomes more curved, and develops a well defined knee and a flatter faint-end slope.

The subhalo-scheme also provides accurate positions for galaxies orbiting in the cluster halo. We have shown that our simple morphological modeling gives rise to a morphology-clustercentric relation that is qualitatively in reasonable agreement with observations. Towards the center of the cluster, the morphological mix of galaxies becomes gradually dominated by ellipticals, while the contribution of spirals strongly declines. Note that the morphology of our model galaxies is primarily determined by their merging history. A morphology density relation arises therefore quite naturally in hierarchical theories of galaxy formation.

Using the subhalo-scheme, our model for the galaxy population of the cluster now produces results that are in good agreement with a variety of data. The cluster-luminosity function has a reasonable shape, the Tully-Fisher relation of field spirals is well fit, the cluster mass-to-light ratio has the right size, a reasonable morphology density relation results, and the $B-V$ color distribution appears to be compatible with observations. Given the approximative treatment of key physical processes, we think that these are remarkable successes. However, it is important to note that changes in some of our model assumptions can have a strong effect on the results. Despite this caveat, semi-analytic models are very useful to study the consequences of any changes in these assumptions, thereby guiding the physical modeling. For example, we have briefly discussed the effects of chemical enrichment of the gaseous halo by retention feedback, and have shown that it is then hard to obtain a Tully-Fisher relation as steep as the observed one.

One powerful strength of semi-analytic models is that they also provide the full history of galaxy formation. If the present time provides the normalization of the models, they can then be used to make predictions at high redshift. Observational data at high redshift can then provide strong constraints on the models. In this work, we have examined the star formation history of the galaxies that end up in the final cluster. The star formation rate of the cluster galaxies peaks at significantly higher redshift than that of the field galaxies. Half of the stars of the cluster are already in place at redshift $z \simeq 4$, while the stars in the field are much younger; half of them have formed below $z \simeq 2$.

The combination of high-resolution N-body simulations and semi-analytic galaxy formation schemes has proven to be able to provide a fairly detailed description of the galaxy population in rich clusters. We think that this approach can be fruitfully exploited to further study the properties of galaxy formation models in hierarchical cosmologies.

–The anomaly showed up most starkly in Base 11 arithmetic, where it could be written out entirely as zeros and ones... Hiding in the alternating patterns of digits, deep inside the transcendental number, was a perfect circle, its form traced out by unities in a field of noughts.

Carl Sagan, Contact

5 GADGET: A code for collisionless and gasdynamical cosmological simulations

Abstract

We describe the newly written code GADGET which is suitable both for cosmological simulations of structure formation and for the simulation of interacting galaxies. GADGET can evolve self-gravitating collisionless fluids with the traditional N-body approach, and a collisional gas by smoothed particle hydrodynamics. Along with the serial version of the code, we present a parallel version that has been designed to run on massively parallel supercomputers with distributed memory. While both versions use a tree algorithm to compute gravitational forces, the serial version of GADGET can optionally employ the special purpose hardware GRAPE instead of the tree. The code uses individual and adaptive timesteps for all particles, and it combines this with a scheme for dynamic tree updates. Due to its Lagrangian nature, GADGET thus allows to bridge a very large dynamic range, both in space and time. Without changes, the code can be used to simulate cosmological volumes as well as problems of interacting galaxies with vacuum boundary conditions. GADGET has been successfully used to run cosmological simulations up to 7.5×10^7 particles, high-resolution simulations of the formation of clusters of galaxies, as well as workstation-sized problems of interacting galaxies. In this study, we detail the numerical algorithms employed, and show various tests of the code.

5.1 Introduction

Numerical simulations of three-dimensional self-gravitating fluids have become an indispensable tool in cosmology. They are now routinely used to study the non-linear gravitational clustering of dark matter, the formation of clusters of galaxies, the interactions of isolated galaxies, and the evolution of the intergalactic gas. Without these numerical techniques the immense progress in these fields would have been impossible, since analytic calculations are often restricted to idealized problems of high symmetry, or to approximative treatments of inherently nonlinear problems.

The advances in numerical simulations have become possible both by the rapid growth of computer performance and also by the implementation of ever more sophisticated numerical algorithms. Note that the development of powerful simulation codes still remains a primary task if one wants to take full advantage of new computer technologies.

While the direct summation method for the gravitational N-body problem remains useful in small stellar dynamical systems, it is hopelessly inefficient for large N due to the $\mathcal{O}(N^2)$ scaling of its computational cost. A large number of groups have therefore developed N-body codes that compute the large-scale gravitational field by means of Fourier techniques. These are the PM, $\mathrm{P^3M}$, and $\mathrm{AP^3M}$ codes (Efstathiou et al. 1985; Couchman 1991; MacFarland et al. 1998). The modern versions of these codes supplement the force computation on scales below the mesh size with a direct summation, and they place mesh refinements on highly clustered regions.

An alternative to these schemes are the so-called tree algorithms, pioneerd by Barnes & Hut (1986) and Jernigan & Porter (1989). Tree algorithms arrange particles in a hierarchy of groups, and compute the gravity at a given point by summing over multipole expansions of these groups. In this way the computational cost of a complete force evaluation can be reduced to a $\mathcal{O}(N \log N)$ scaling.

While mesh-based codes are generally much faster for close-to-homogeneous particle distributions, the tree codes can adapt flexibly to any clustering state without significant losses in speed. This Lagrangian nature is a great advantage, if a large dynamic range in density needs to be covered. Here tree codes can outperform mesh based algorithms. In addition, tree codes are free from any geometrical restrictions, and they can be combined with integration schemes that advance particles with individual timesteps.

Yet another solution to the N-body problem is provided by special-purpose hardware like the GRAPE board (Steinmetz 1996). It consists of custom chips that compute gravitational forces by the direct summation technique. By means of their enormous computational speed they can considerably extend the range where direct summation remains competitive with software solutions.

In recent years, collisionless dynamics has also been coupled to gas dynamics, establishing the link between dark matter and the directly observable quantities. Traditionally, hydrodynamical simulations have usually employed some kind of mesh to represent fluid dynamical quantities. While a particular strength of these codes is their ability to accurately resolve shocks, the mesh also imposes restrictions on the geometry of the problem, and onto the dynamic range of spatial scales that can be simulated. However, new adaptive mesh refinement codes of Norman & Bryan (1998) and Klein et al. (1998) may provide a solution to this problem.

In cosmological applications, it is often sufficient to describe the gas by smoothed particle hydrodynamics (SPH), as invented by Lucy (1977) and Gingold & Monaghan (1977). The particle-based SPH is extremely flexible in its ability to adapt to any given geometry. Moreover, its Lagrangian nature allows a locally changing resolution that 'automatically' follows the local mass density. This convenient feature helps to save computing time by concentration the computational effort on those regions that have the largest gas concentrations. Furthermore, SPH ties naturally into the N-body approach for self-gravity, and can be easily implemented in three dimensions.

These advantages have led a number of authors to develop SPH codes for applications in cosmology. Among them are TREESPH (Hernquist & Katz 1989; Katz et al. 1996), GRAPESPH (Steinmetz 1996), HYDRA (Couchman et al. 1995; Pearce & Couchman 1997), and codes by Evrard (1988); Navarro & White (1993); Hultman & Källander (1997); Davé et al. (1997); Carraro et al. (1998). See also Kang et al. (1994) and Frenk et al. (1999) for a comparison of cosmological hydrodynamic codes.

In this chapter we describe our simulation code GADGET (**GA**laxies with **D**ark matter and **G**as int**E**rac**T**), which can be used both to simulate isolated self-gravitating systems including gas, or for cosmological N-body simulations. We have developed two versions of this code, a workstation version, and a version for massively parallel supercomputers with distributed memory. The workstation code uses either a tree algorithm for the self-gravity, or the special purpose hardware GRAPE , if available. The parallel version works with a tree only. Note that the parallelization required substantial algorithmic changes compared to the serial code.

A particular emphasis of our work has been the development of an efficient time integration scheme that allows individual and adaptive particle timesteps, and on the elimination of sources of overhead both in the serial and parallel code. We have investigated different time step criteria, and have come up with a number of algorithmic improvements that speed up the code, while maintaining its accuracy. As a result, GADGET is a very flexible code without obvious intrinsic restrictions to the dynamic range of the problems that we address with it.

In this work we describe in detail the implemented equations and algorithms of our codes, and we show results for some test problems, and some science applications. In Section 5.2, we give a brief summary of the implemented physics. In Section 5.3, we discuss the computation of the gravitational force both with a tree algorithm, and with GRAPE . We then describe our specific implementation of SPH in Section 5.4, and we discuss our time integration scheme in Section 5.5. The parallelization of the code is described in Section 5.6, and tests of the code(s) are presented in Section 5.7. Finally, we summarize in Section 5.8.

5.2 Implemented physics

5.2.1 Collisionless dynamics

Dark matter and stars are modeled as self-gravitating collisionless fluids, i.e. they fulfill the collisionless Boltzmann equation (CBE)

$$\frac{\mathrm{d}f}{\mathrm{d}t} \equiv \frac{\partial f}{\partial t} + \mathbf{v}\frac{\partial f}{\partial \mathbf{x}} - \frac{\partial \Phi}{\partial \mathbf{r}}\frac{\partial f}{\partial \mathbf{v}} = 0, \tag{5.1}$$

where the self-consistent potential Φ is the solution of Poisson's equation

$$\nabla^2\Phi(\mathbf{r},t) = 4\pi G \int f(\mathbf{r},\mathbf{v},t)\mathrm{d}\mathbf{v}, \tag{5.2}$$

and $f(\mathbf{r},\mathbf{v},t)$ is the mass density in single-particle phase-space. It is exceedingly difficult to solve this coupled system of equations directly with finite difference methods. Instead, we will follow the common N-body approach, where the phase fluid is represented by N particles which are integrated along the characteristic curves of the CBE. In essence, this is a Monte Carlo approach whose accuracy depends crucially on a sufficiently high number of particles.

The N-body problem is thus the task of following Newton's equation of motion for a large number of particles under their own self-gravity. Note that we will introduce a softening into the gravitational potential at small separations. This effectively introduces a lower resolution cut-off, which allows larger integration timesteps, and also prolongs the relaxation time of the system due to two-body encounters.

5.2.2 Gasdynamics

A simple description of the intergalactic medium (IGM), or the interstellar medium (ISM), may be obtained by modeling it as an ideal, inviscid gas. The gas is then governed by the continuity equation

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} + \rho\nabla\cdot\mathbf{v} = 0, \tag{5.3}$$

and the Euler equation

$$\frac{\mathrm{d}v}{\mathrm{d}t} = -\frac{\nabla P}{\rho} - \nabla\Phi. \tag{5.4}$$

Further, the thermal energy u per unit mass evolves according to the first law of thermodynamics, viz.

$$\frac{\mathrm{d}u}{\mathrm{d}t} = -\frac{P}{\rho}\nabla\cdot\mathbf{v} - \frac{\Lambda(u,\rho)}{\rho}. \tag{5.5}$$

Here we used Lagrangian time derivatives, i.e.

$$\frac{\mathrm{d}}{\mathrm{d}t} = \frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla, \tag{5.6}$$

and we allowed for a piece of 'extra' physics in form of the *cooling function* $\Lambda(u,\rho)$, describing external sinks or sources of heat for the gas.

For a simple ideal gas, the equation of state is

$$P = (\gamma-1)\rho u, \tag{5.7}$$

where γ is the adiabatic exponent. We usually take $\gamma = 5/3$, appropriate for a mono-atomic ideal gas. The adiabatic sound speed c of this gas is $c^2 = \gamma P/\rho$.

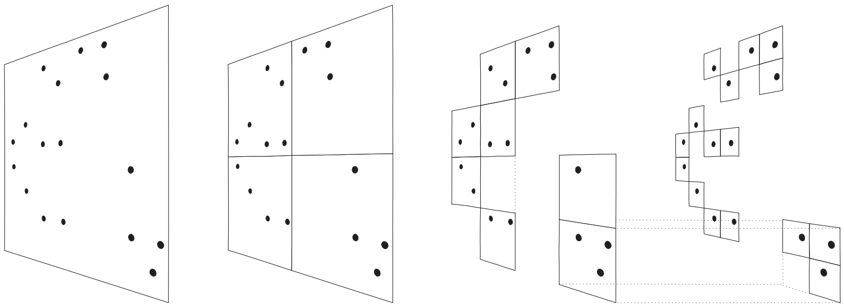


Figure 5.1: Schematic illustration of the Barnes & Hut oct-tree in two dimensions. The particles are first enclosed in a square (root node). This square is then iteratively subdivided in four squares of half the size, until exactly one particle is left in each final square (leaves of the tree). In the resulting tree structure, each square can be progenitor of up to four siblings. Note that empty squares need not to be stored.

5.3 Gravity

5.3.1 Tree algorithm

An alternative to Fourier techniques, or to direct summation, are the so-called tree methods. In these schemes the particles are arranged in a hierarchy of groups. When the force on a particular particle is computed the force exerted by distant groups is approximated by their lowest multipole moments. Usually this approximation is terminated at quadrupole order. Tree codes are a class of algorithms that in this way reduce the computational cost for a complete force evaluation to $\mathcal{O}(N \log N)$ (Jernigan & Porter 1989; Barnes & Hut 1986).

We employ the Barnes & Hut (1986, BH) tree in this work. In this scheme, the computational domain is hierarchically partitioned into a sequence of cubes, where each cube contains eight siblings of half the size. These cubes form the nodes of an oct-tree structure. The tree structure is constructed such that each node (cube) contains either exactly one particle, or it is progenitor to further nodes, in which case the node carries the monopole and quadrupole moments of all the particles in the tree below it. A schematic illustration of the BH tree is shown in Figure 5.1.

A force computation then proceeds by walking the tree, and summing up appropriate force contributions from tree nodes. The multipole expansion of a node of size l is used only if

$$r > \frac{l}{\theta}, \tag{5.8}$$

where r is the distance of the point of reference to the center-of-mass of the cell and θ is a prescribed accuracy parameter. If a node fulfills the criterion (5.8), the tree walk along this branch can be terminated, otherwise it is continued with all its siblings. Equation (5.8) is the standard Barnes & Hut (BH) opening criterion. Following Dubinski et al.

(1996) we have also tried the simple modification

$$r > \frac{l}{\theta} + \delta, \tag{5.9}$$

where the quantity δ gives the distance of the geometric center of the cell to its center-of-mass. This provides protection against pathological cases where the center-of-mass lies close to an edge of a cell. However, except for the first force computation in a simulation, we will usually employ yet another opening criterion to be discussed in Section 5.3.1.2.

A technical difficulty arises when the gravity is softened. In regions of high particle density (e.g. the centers of dark haloes, or cold dense gas knots in dissipative simulations), it can frequently happen that nodes fulfill equation (5.8), and simultaneously we have $r < h$, where h is the gravitational softening length. In this situation, one formally needs the multipole moments of the *softened* gravitational field. Some previous tree codes have ignored this complication altogether, or they have opened nodes always for $r < h$. The first solution can lead to forces significantly in error, while the latter can suffer from severe (and unnecessary) performance degradation for strongly clustered regions (Hernquist & Katz 1989).

However, it is possible to do the proper multipole expansion for the softened potential, and we know discuss it for definiteness. We want to approximate the potential at $\mathbf{r}$ due to a (distant) bunch of particles with masses m_i and coordinates $\mathbf{x}_i$. Note that we use a spline-softened force law, hence the exact potential of the particle group is

$$\Phi(\mathbf{r}) = -G \sum_k m_k \, g\left(|\mathbf{x}_k - \mathbf{r}|\right), \tag{5.10}$$

where the function $g(r)$ describes the softened force law. For Newtonian gravity we have $g(r) = 1/r$, while the spline softened gravity with softening length h gives rise to

$$g(r) = -\frac{1}{h} W_2\left(\frac{r}{h}\right). \tag{5.11}$$

The function $W_2(u)$ is given in the appendix of this chapter. It arises by replacing the force due to a point mass m with the force exerted by the mass distribution $\rho(\mathbf{r}) = mW(\mathbf{r}; h)$, where we take $W(r; h)$ to be the spline kernel used in the SPH formalism.

The spline softening has the advantage that the force becomes exactly Newtonian for $r > h$, while other possible force laws, like the Plummer softening to be discussed below, often converge only relatively slowly to Newton's law.

We now introduce the center-of-mass $\mathbf{s}$, and the total mass M of the particles. Further we define

$$\mathbf{y} \equiv \mathbf{r} - \mathbf{s}. \tag{5.12}$$

The potential may then be expanded in a multipole series assuming $|\mathbf{y}| \gg |\mathbf{x}_k - \mathbf{s}|$. Up to quadrupole order, this results in

$$\Phi(\mathbf{r}) = -G \left\{ M g(y) + \frac{1}{2} \mathbf{y}^T \left[\frac{g''(y)}{y^2} \mathbf{Q} + \frac{g'(y)}{y^3} (\mathbf{P} - \mathbf{Q}) \right] \mathbf{y} \right\}. \tag{5.13}$$

Here we have introduced the tensors

$$\mathbf{Q} = \sum_k m_k(\mathbf{x}_k - \mathbf{s})(\mathbf{x}_k - \mathbf{s})^T = \sum_k m_k \mathbf{x}_k\, \mathbf{x}_k^T - M\mathbf{s}\,\mathbf{s}^T, \tag{5.14}$$

and

$$\mathbf{P} = \mathbf{1}\sum_k m_k(\mathbf{x}_k - \mathbf{s})^2 = \mathbf{1}\sum_k m_k \mathbf{x}_k^2 - \mathbf{1}M\mathbf{s}^2. \tag{5.15}$$

Note that for Newtonian gravity, equation (5.13) reduces to the perhaps more familiar form

$$\Phi(\mathbf{r}) = -G\left[\frac{M}{y} + \frac{1}{2}\mathbf{y}^T\frac{3\mathbf{Q} - \mathbf{P}}{y^5}\mathbf{y}\right]. \tag{5.16}$$

Finally, the quadrupole approximation of the softened gravitational field is given by

$$\mathbf{f}(\mathbf{r}) = -\nabla\Phi = G\left\{Mg_1(y)\mathbf{y} + g_2(y)\mathbf{Q}\mathbf{y} + \frac{1}{2}g_3(y)\left(\mathbf{y}^T\mathbf{Q}\mathbf{y}\right)\mathbf{y} + \frac{1}{2}g_4(y)\mathbf{P}\mathbf{y}\right\}. \tag{5.17}$$

Here we introduced the functions $g_1(y)$, $g_2(y)$, $g_3(y)$, and $g_4(y)$ as convenient abbreviations. Their definition is given in the appendix. In the Newtonian case, this simplifies to

$$\mathbf{f}(\mathbf{r}) = G\left\{-\frac{M}{y^3}\mathbf{y} + \frac{3\mathbf{Q}}{y^5}\mathbf{y} - \frac{15}{2}\frac{\mathbf{y}^T\mathbf{Q}\mathbf{y}}{y^7}\mathbf{y} + \frac{3}{2}\frac{\mathbf{P}}{y^5}\mathbf{y}\right\}. \tag{5.18}$$

Note that although equation (5.17) looks rather cumbersome, its actual numerical computation is only marginally more expansive than that of the Newtonian form (5.18) because all factors involving $g(y)$ and derivatives thereof can be tabulated for later use in repeated force calculations.

5.3.1.1 Tree construction and tree walks

The tree construction can be done by inserting the particles one after the other in the tree. It is possible to compute the multipole moments simultaneously by summing the $\sum_k m_k x_i^{(k)} x_j^{(k)}$ terms for each node, and then looping at the end over all nodes again and subtracting the corresponding center-of-mass moments. It turns out, however, that the particle coordinates have to be refered to a fixed point inside the node (we adopt the geometrical center for that) in this procedure, otherwise intolerable accuracy losses can result if single-precision arithmetic is used. Our code has been written in C and optimized extensively with the help of profilers. As a result, the tree-construction is very fast, and the time spent for it is negligible compared to a complete force walk for all particles.

However, in our time integration scheme we use individual timesteps for all particles, and at each given timestep, only a small fraction of all particles require a force walk. If this fraction drops below ~ 1 per cent, a full reconstruction of the tree can take as long as the force walk. However, most of this tree construction time can be eliminated by *dynamic tree updates*, which we discuss in more detail in Section 5.5.

The most time consuming routine in the code will then always be the tree walk. Hence optimizing it can considerably speed up tree codes. We use pointers to access the data structures and implemented a non-recursive tree walk, similar to Dubinski (1996). The latter speeds up the tree walk by up to 10%.

GADGET also allows different gravitational softenings for groups of particles. In order to guarantee momentum conservation, this requires a symmetrization of the force when particles with different softening lengths interact. We here chose to symmetrize the softenings by

$$h = \max(h_i, h_j). \tag{5.19}$$

However, the usage of different softening lengths leads also to complications for softened tree nodes. Strictly speaking, the multipole expansion is only possible, if all the particles in the node have the same softening. One clean possibility to solve this problem is to construct separate trees for each species of particles with different softening. As long as these species are more or less spatially separated, no severe performance penalty results. However, this is different if the fluids are spatially 'mixed' (e.g. dark halo, stellar disk, gaseous disk, and stellar bulge in simulations of interacting galaxies). Here it can be better to construct a single tree, and open cells always if $r \leq h_i < h_{\max}$, where $h_{\max}$ is the maximum softening of the particles represented by the node, and h_i is the softening of the particle under consideration. GADGET allows a choice between either of the above methods.

5.3.1.2 A new opening criterion

The accuracy of the forces resulting from the tree walk depends sensitively on the criterion used to decide whether the multipole approximation for a given node is acceptable, or whether the node has to be 'opened' for further refinement. The standard BH opening criterion tries to limit the relative error of every particle-node interaction by comparing a rough estimate of the size of the quadrupole term, $\sim Ml^2/r^4$, with the size of the monopole term, $\sim M/r^2$. The result is the purely geometrical criterion of equation (5.8).

As Salmon & Warren (1994) have pointed out, the worst-case behaviour of the BH opening criterion is somewhat worrying. Although very rare in real astrophysical simulations, the BH criterion can occasionally lead to very large force errors, when standard values for the opening angle are used.

Another problem with the BH criterion occurs when one tries to use it at high redshift in cosmological simulations. Here, the density field is very close to homogeneous and the peculiar acceleration is small. For a tree algorithm this is a tough problem, because the tree code always has to sum up partial forces from *all* the mass in a simulation. The small net force at high z then arises in a delicate cancellation process between relatively large partial forces. If a partial force is indeed much larger than the net force, even a small *relative* error in it is enough to result in a large relative error of the net force. For an unclustered particle distribution, the BH criterion therefore requires a much smaller value of the opening angle than for a clustered one.

Similarly, in a cosmological simulation the absolute sizes of forces between a given particle and tree-nodes of a certain opening angle can vary by many orders of magnitude. In this situation, the purely geometrical BH criterion ends up investing a lot of computational effort for the evaluation of all partial forces to the same relative accuracy, irrespective of the actual size of each partial force, and the size of the *absolute* error thus induced. It would be better to invest more computational effort in regions that provide most of the force on the particle and less in regions whose mass content is unimportant for the total force.

As suggested by Salmon & Warren (1994), one may therefore try to devise a cell-opening criterion that limits the absolute error in every cell-particle interaction. In principle, one may use analytic error bounds (Salmon & Warren 1994) to obtain a suitable cell-opening criterion, but the evaluation of the relevant expressions can consume significant amounts of CPU time.

Our approach to a new opening criterion is less stringent. Assume the absolute size of the true total force is already known before the tree walk. In our code, we will use the acceleration of the previous timestep as a handy approximate value for that. We will now require that the estimated error of an acceptable multipole approximation is some small fraction of this total force. Note that we truncate the multipole expansion at quadrupole order. If the hexadecapole order is the leading term in the neglected part of the series, the absolute truncation error will be roughly of the size of this term, which is of order $\simeq M/r^2(l/r)^4$. We can then require that this error should not exceed some fraction α of the total force on the particle. The latter may be estimated from the previous timestep. A tree-node has then to be opened if

$$M\,l^4 > \alpha|\mathbf{a}_{\rm old}|r^6\,. \tag{5.20}$$

We have found this criterion to be more efficient than the ordinary BH criterion, i.e. at a given computational expense it produces forces that are more accurate. Also, this criterion does not suffer from the high-z problem discussed above. The same value of α produces a comparable force accuracy, independent of the clustering state of the material. In Section 5.7.2, we will show some quantitative measurements of the relative performance of the two criteria.

5.3.2 Special purpose hardware

An alternative to software solutions to the N^2-bottleneck of self-gravity is provided by the GRAPE (GRAvity PipE) special-purpose hardware. It is designed to solve the gravitational N-body problem in a direct summation, brute-force approach by means of its superior computational speed. The latter is achieved with custom chips that compute the gravitational force with a hardwired Plummer force law. The Plummer-potential of GRAPE takes the form

$$\Phi(\mathbf{r}) = -G\sum_j \frac{m_j}{(|\mathbf{r}-\mathbf{r}_j|^2+\epsilon^2)^{\frac{1}{2}}}. \tag{5.21}$$

The GRAPE -3A boards installed at the MPA have 40 N-body integrator chips in total with an approximate peak performance of 25 GFlops. Recently, newer generations of GRAPE boards have achieved even higher computational speeds. In fact, with the GRAPE -4 the 1 TFlop barrier had first been surpassed (Makino et al. 1997), and even faster special-purpose machines are in preparation (Spurzem 1997).

The GRAPE -3A boards are connected to an ordinary workstation via a VME interface. The boards consist of memory chips that can hold up to 131072 particle coordinates, and of integrator chips that can compute the forces exerted by these particles for 40 positions in parallel. Note that higher particle numbers can also be processed by splitting them up in sufficiently small groups. In addition to the gravitational force, the GRAPE board returns the potential, and a list of neighbours for the 40 positions within search radii h_i specified by the user. This latter feature makes GRAPE especially attractive for SPH calculations.

The parts of our code that use GRAPE have benefited from the code GRAPESPH by Steinmetz (1996), and are similar to it. In short, the usage of GRAPE proceeds as follows. For the force computation, the particle coordinates are first loaded onto the GRAPE board, then GADGET calls GRAPE repeatedly to compute the force for up to 40 positions in parallel. The communication with GRAPE is done by means of a convenient software interface in C. GRAPE can also provide lists of nearest neighbours. For SPH-particles, GADGET computes the gravitational force and the interaction list in just one call of GRAPE . The host computer then still does the rest of the work, i.e. it advances the particles, and computes the hydrodynamical forces.

In practice, there are some technical complications when one works with GRAPE . In order to achieve the high computational speed, the GRAPE hardware works internally with special fixed-point formats for positions, accelerations and masses. This results in a reduced dynamic range compared to standard IEEE floating point arithmetic. In particular, one needs to specify a minimum length scale $d_{\min}$ and a minimum mass scale $m_{\min}$ when working with GRAPE . The spatial dynamic range is then given by $d_{\min}[-2^{18}; 2^{18}]$ and the mass range is $m_{\min}[1; 64\epsilon/d_{\min}]$ (Steinmetz 1996). In GADGET we take special care that these constraints are always observed, otherwise the simulation could be compromised by numerical errors.

While the communication time with GRAPE scales proportional to the particle number N, the actual force computation of GRAPE is still an $\mathcal{O}(N^2)$-algorithm, because the GRAPE board represents a brute-force approach to the gravitational N-body problem. This implies that for very large particle number a tree code running on the workstation alone will eventually catch up and outperform the combination of workstation and GRAPE . For our current set-up at MPA this break-even point is about at 300000 particles.

However, it seems also possible to combine GRAPE with a tree algorithm (Athanassoula et al. 1997), for example by exporting tree nodes instead of particles in an appropriate way. Such a combination of tree+GRAPE could scale as $\mathcal{O}(N \log N)$ and outperform pure software solutions even for large N.

5.4 Smoothed particle hydrodynamics

SPH is a powerful Lagrangian technique to solve hydrodynamical problems with an ease that is unmatched by grid based fluid solvers (see Monaghan 1992, for an excellent review). In particular, SPH is very well suited for three-dimensional astrophysical problems that do not crucially rely on very accurately resolved shock fronts.

Unlike other numerical approaches for hydrodynamics, the SPH equations do not take a unique form. Instead, many formally different versions of them can be derived. Furthermore, a large variety of recipes for specific implementations of force symmetrization, determinations of smoothing lengths, and artificial viscosity, have been described. Some of these choices are crucial for the accuracy and efficiency of the SPH implementation, others are only of minor importance. See the recent work by Thacker et al. (1998) and Lombardi et al. (1998) for a discussion of the relative performance of some of these possibilities. Below we give a summary of the specific SPH implementation we use. In it we have tried to combine the best formulations that have emerged, while simultaneously emphasizing accuracy and the need that the scheme must be amenable to an individual timestep integrator.

The computation of the hydrodynamic force and the rate of change of internal energy proceeds in two phases. In the first phase, new smoothing lengths h_i are determined for the *active* particles (these are the ones that need a force update at the current timestep, see below). We set h_i equal to the distance of the $N_{\rm s}$-th nearest neighbour of a particle, and we store all the indices of the $N_{\rm s}$ closest particles. Note that holding the number of neighbours exactly constant is only slightly more expansive than letting it fluctuate (the neighbour search is only a negligible fraction of the total computational time anyway). However, a constant number of neighbours leads to much better energy conservation, as shown by Nelson & Papaloizou (1994), and it reduces the 'noise' in SPH related quantities considerably.

For the active particles, we then compute the density of the particle as

$$\rho_i = \sum_{j=1}^{N} m_j W(\mathbf{r}_{ij}; h_i), \tag{5.22}$$

where $\mathbf{r}_{ij} \equiv \mathbf{r}_i - \mathbf{r}_j$, and we compute a new estimate of divergence and vorticity as

$$\rho_i (\nabla \cdot \mathbf{v})_i = \sum_j m_j (\mathbf{v}_j - \mathbf{v}_i) \nabla_i W(\mathbf{r}_{ij}; h_i), \tag{5.23}$$

$$\rho_i (\nabla \times \mathbf{v})_i = \sum_j m_j (\mathbf{v}_i - \mathbf{v}_j) \times \nabla_i W(\mathbf{r}_{ij}; h_i). \tag{5.24}$$

Here we employ the *gather* formulation for adaptive smoothing (Hernquist & Katz 1989).

For the *passive* particles, values for density, internal energy, and smoothing length are predicted at the current time based on the values of the last update of those particles (see Section 5.5). Finally, the pressure of the particles is set to $P_i = (\gamma - 1)\rho_i u_i$.

In the second phase, the actual forces are computed. Here we symmetrize the kernels of *gather* and *scatter* formulations in the way of Hernquist & Katz (1989). We compute the gasdynamical accelerations as

$$\mathbf{a}_i^{\rm gas} = -\left(\frac{\nabla P}{\rho}\right)_i + \mathbf{a}_i^{\rm visc} = -\sum_j m_j \left(\frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2} + \tilde{\Pi}_{ij}\right) \left[\frac{1}{2}\nabla_i W(\mathbf{r}_{ij}; h_i) + \frac{1}{2}\nabla_i W(\mathbf{r}_{ij}; h_j)\right], \tag{5.25}$$

and the change of the internal energy as

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = \frac{1}{2}\sum_j m_j \left(\frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2} + \tilde{\Pi}_{ij}\right) (\mathbf{v}_i - \mathbf{v}_j) \left[\frac{1}{2}\nabla_i W(\mathbf{r}_{ij}; h_i) + \frac{1}{2}\nabla_i W(\mathbf{r}_{ij}; h_j)\right]. \tag{5.26}$$

The artificial viscosity $\tilde{\Pi}_{ij}$ is taken to be

$$\tilde{\Pi}_{ij} = \frac{1}{2}(f_i + f_j)\Pi_{ij}, \tag{5.27}$$

with

$$\Pi_{ij} = \begin{cases} \left[-\alpha c_{ij}\mu_{ij} + 2\alpha\mu_{ij}^2\right]/\rho_{ij} & \text{if } \mathbf{v}_{ij}\cdot\mathbf{r}_{ij} < 0 \\ 0 & \text{otherwise,} \end{cases} \tag{5.28}$$

where

$$f_i = \frac{|(\nabla\cdot\mathbf{v})_i|}{|(\nabla\cdot\mathbf{v})_i| + |(\nabla\times\mathbf{v})_i|}, \tag{5.29}$$

and

$$\mu_{ij} = \frac{h_{ij}(\mathbf{v}_i - \mathbf{v}_j)(\mathbf{r}_i - \mathbf{r}_j)}{|\mathbf{r}_i - \mathbf{r}_j|^2 + \epsilon h_{ij}^2}. \tag{5.30}$$

This form of artificial viscosity is the shear-reduced version (Balsara 1995; Steinmetz 1996) of the 'standard' Monaghan & Gingold (1983) artificial viscosity. Recent studies (Lombardi et al. 1998; Thacker et al. 1998) that test SPH implementations strongly endorse it.

Since we store the full neighbour list for all particles, we can guarantee that the sums contributing to equations (5.25) and (5.26) are always computed exactly, without a possible loss of some of the pairs. For this purpose we simply scan the complete interaction list. For every entry (i.e. the index j of a neighbour of particle i), there is a resulting force component

$$\mathbf{f}_{ij} = -m_i m_j \left(\frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2} + \tilde{\Pi}_{ij}\right) \frac{1}{2}\nabla_i W(\mathbf{r}_{ij}; h_i). \tag{5.31}$$

If we add $\mathbf{f}_{ij}$ to the force on i, and $-\mathbf{f}_{ij}$ to the force on j, the sum of equation (5.25) is reproduced, and the momentum conservation is manifest. This also holds for the internal energy.

5.4.1 Neighbour search

In SPH, one needs to find the nearest neighbours of each SPH particle to construct its interaction list. Specifically, one has the numerical task to find all particles closer than a search radius R. Similar to gravity, the naive solution that checks the distance of *all* particle pairs is an $\mathcal{O}(N^2)$ algorithm which slows down prohibitively for large particle numbers. Fortunately, there are faster search algorithms.

When the particle distribution is approximately homogeneous, perhaps the fastest algorithm works with a search grid that has a cell size equal to R. The particles are then first coarse-binned on this search grid, and link-lists are established that quickly deliver only those particles that lie in a specific cell of the coarse grid. The neighbour search proceeds then by *range searching*; only those mesh cells have to be opened that have a spatial overlap with the search range.

For highly clustered particle distributions and varying search ranges R, the above approach quickly degrades, since the mesh of the coarse grid has not the optimum size for all particles. A more flexible alternative is to employ a geometric search tree. For this purpose we construct a second BH tree just for the SPH particles, but without the computation of multipole moments. A neighbour search is then performed by walking the tree. A cell is 'opened' (i.e. further followed) if it has a spatial overlap with the rectangular search range. If one arrives at a cell with only one particle, this is added to the interaction list if it lies in the search range.

In the algorithm just described, the tree needs to be walked down to its leaves to find all the particles in the interaction list. If the length of this tree walk can be reduced, the speed of the algorithm can be increased. In GADGET we use the following additional trick to achieve that. A tree walk along a branch is terminated, if the cell lies *completely* inside the search range. Then all the particles in the cell can be added to the interaction list. But how can these particles be found in the first place without further walking the tree? It turns out that it is possible to set-up a link-list during the tree construction that allows a retrieval of all the particles that lie inside a given cell, just like it is done in the coarse-binning approach. This trick speeds up the neighbour search by more than a factor of 2, and we will also employ it in the dynamic updates of the force tree to be described later on.

There remains the task to specify an algorithm that finds exactly N_s closest neighbours. We solve this in the following way: We predict a value $\tilde{h}_i$ for the smoothing length based on the length h_i of the previous timestep and the local velocity divergence. We then do a range search with $R = 1.2\tilde{h}_i$, on average resulting in $\sim 2N_\mathrm{s}$ potential neighbours. From these we select the closest N_s with the fast algorithm `select` (Press et al. 1995). If there are fewer than N_s particles in the search range, or if the distance of the N_s-th nearest particle inside the search range is larger than R, the search is repeated for a larger search range. In the first timestep no previous h_i is known, so we follow the neighbour tree backwards from the leave of the particle under consideration, until we obtain a first reasonable guess for the local particle density (from the number N of particles in a node of volume l^3). This provides an initial guess for $\tilde{h}_i$.

5.5 Time integration

As a time integrator, we use a variant of the leapfrog involving an explicit prediction step. The latter is introduced to accommodate individual particle timestep in the N-body scheme, as explained later on.

We here start by describing the integrator for a single particle. First, a particle position at the middle of the timestep Δt is predicted according to

$$\tilde{\mathbf{r}}^{(n+\frac{1}{2})} = \mathbf{r}^{(n)} + \mathbf{v}^{(n)}\frac{\Delta t}{2}, \tag{5.32}$$

and an acceleration based on this position is computed, viz.

$$\mathbf{a}^{(n+\frac{1}{2})} = -\left.\nabla\Phi\right|_{\tilde{\mathbf{r}}^{(n+\frac{1}{2})}}. \tag{5.33}$$

Then the particle is advanced according to

$$\mathbf{v}^{(n+1)} = \mathbf{v}^{(n)} + \mathbf{a}^{(n+\frac{1}{2})}\Delta t, \tag{5.34}$$

$$\mathbf{r}^{(n+1)} = \mathbf{r}^{(n)} + \frac{1}{2}\left[\mathbf{v}^{(n)} + \mathbf{v}^{(n+1)}\right]\Delta t. \tag{5.35}$$

5.5.1 Timestep criterion

In the above scheme, the timestep may vary from step to step. Note that the choice of timestep criterion is very important in determining the overall accuracy and computational efficiency of the integration.

In a static potential Φ, the error in specific energy arising in one step with the above integrator is

$$\begin{aligned} \Delta E \;=\; & \frac{1}{4}\frac{\partial^2\Phi}{\partial x_i \partial x_j} v_i^{(n)} a_j^{(n+\frac{1}{2})}\Delta t^3 + \\ & \frac{1}{24}\frac{\partial^3\Phi}{\partial x_i \partial x_j \partial x_k} v_i^{(n)} v_j^{(n)} v_k^{(n)}\Delta t^3 + \mathcal{O}(\Delta t^4) \end{aligned} \tag{5.36}$$

to leading order in Δt, i.e. the integrator is second order accurate. Here the derivatives of the potential are taken at coordinate $\mathbf{r}^{(n)}$ and summation over repeated coordinate indices is understood.

In principle, one could try to use equation (5.36) directly to obtain a timestep by imposing some upper limit on the tolerable error ΔE. However, this approach is quite subtle in practice. First, the derivatives of the potential are difficult to obtain, and second, there is no explicit guarantee, that the terms of higher order in Δt are really small.

To guarantee that the higher order terms indeed become smaller, one may consider the change of basic quantities describing the particle. For example, the change of the position of the particle is

$$\Delta\mathbf{r} = \mathbf{r}^{(n+1)} - \mathbf{r}^{(n)} = \mathbf{v}^{(n)}\Delta t + \frac{1}{2}\mathbf{a}^{(n+\frac{1}{2})}\Delta t^2, \tag{5.37}$$

and its change in specific kinetic energy is

$$\Delta E_{\rm kin} = \mathbf{v}^{(n)}\mathbf{a}^{(n+\frac{1}{2})}\Delta t + \frac{1}{2}\mathbf{a}^{(n+\frac{1}{2})}\mathbf{a}^{(n+\frac{1}{2})}\Delta t^2. \tag{5.38}$$

For a stable integration, the terms of second order in Δt should be smaller than the first order terms. This suggests a timestep criterion of the form

$$\Delta t = \alpha_{\rm tol}\frac{\mathbf{v}}{|\mathbf{a}|}, \tag{5.39}$$

where α is some dimensionless tolerance parameter. Alternatively, one may constrain the absolute size of the second order displacement by assuming a typical velocity dispersion σ^2 for the particles, corresponding to a scale $E = \sigma^2$ for the typical specific energy. This results in

$$\Delta t = \alpha_{\rm tol}\frac{\sigma}{|\mathbf{a}|}. \tag{5.40}$$

We have found this criterion to give the best results compared to several alternative choices, and we will later on show test calculations illustrating this in more detail.

The quantity $\alpha_{\rm tol}\sigma$ controls the accuracy of the integrations. Our typical choice for simulations of interacting galaxies is of order $\alpha_{\rm tol}\sigma \simeq 10 - 25\,{\rm km/sec}$. Note that the timestep criterion (5.40) is Galilean-invariant and does not make an explicit reference to the gravitational softening length employed.

5.5.2 Integrator for N-body systems

In contrast to a binary hierarchy of timesteps (Hernquist & Katz 1989; Steinmetz 1996), we employ an integrator with completely flexible timesteps, similar to the one employed by Groom (1997) and Hiotelis & Voglis (1991).

Each particle owns a timestep Δt_i, and a current time t_i, where its dynamical state ($\mathbf{r}_i$, $\mathbf{v}_i$, $\mathbf{a}_i$) is stored. The dynamical state of the particle can be predicted at times $t \in [t_i \pm 0.5\Delta t_i]$ with first order accuracy.

The next particle k to be advanced is then the one with the minimum prediction time defined as $\tau_{\rm p} \equiv \min\left(t_i + 0.5\Delta t_i\right)$. The time $\tau_{\rm p}$ becomes the new current time of the system. To advance the particle, we first predict positions for *all* particles at time $\tau_{\rm p}$ according to

$$\tilde{\mathbf{r}}_i = \mathbf{r}_i + \mathbf{v}_i(\tau_{\rm p} - t_i). \tag{5.41}$$

Based on these positions, the acceleration of particle k at the middle of its timestep is calculated as

$$\mathbf{a}_k^{(n+\frac{1}{2})} = -\left.\nabla\Phi(\tilde{\mathbf{r}}_i)\right|_{\tilde{\mathbf{r}}_k}. \tag{5.42}$$

Position and velocity of particle k are then advanced as

$$\mathbf{v}_k^{(n+1)} = \mathbf{v}_k^{(n)} + 2\mathbf{a}_k^{(n+\frac{1}{2})}(\tau_{\rm p} - t_k), \tag{5.43}$$

$$\mathbf{r}_k^{(n+1)} = \mathbf{r}_k^{(n)} + \left[\mathbf{v}_k^{(n)} + \mathbf{v}_k^{(n+1)}\right](\tau_{\rm p} - t_k), \tag{5.44}$$

and its current time can be updated to

$$t_k^{(\mathrm{new})} = t_k + 2(\tau_{\mathrm{p}} - t_k). \tag{5.45}$$

Finally, a new timestep $\Delta t_k^{(\mathrm{new})}$ for the particle is estimated.

At the beginning of the simulation, all particles start out with the same current time. However, since the timesteps of the particles are all different, the current times of the particles distribute themselves nearly symmetrically around the current prediction time, hence the prediction step involves forward and backward prediction to a similar extent.

However, it is impractical to advance only a single particle at any given prediction time, because the prediction itself and the (dynamic) tree updates induce some overhead. For this reason we advance particles in bunches. The particles may be thought of as being ordered according to their maximum prediction times $t_i^{\mathrm{p}} = t_i + \frac{1}{2}\Delta t_i$. The simulation works through this time line, and always advances the particle with the smallest t_i^{p}, and also all subsequent particles in the time line, until the first is found with

$$\tau_{\mathrm{p}} \leq t_i + \frac{1}{4}\Delta t_i \,. \tag{5.46}$$

This condition selects a group of particles at the lower end of the time line, and all the particles of the group are guaranteed to be advanced at least by half or their maximum allowed timestep.

Our practical experience shows that the size M of the group that is advanced at a given step is often only a small fraction of the total particle number N. In this situation it becomes important to eliminate any overhead that scales with $\mathcal{O}(N)$. For example, we obviously need to find the particle with the minimum prediction time at every timestep, and also the particles following it in the time line. A loop over all the particles, or a complete sort at every timestep, would induce overhead of order $\mathcal{O}(N)$ or $\mathcal{O}(N \log N)$, which can become comparable to the force computation itself if $M/N \ll 1$. We solve this problem by keeping the maximum prediction times of the particles in an ordered binary tree (Wirth 1986) at all times. Finding the particle with the minimum prediction time and the ones that follow it are then operations of order $\mathcal{O}(\log N)$. Also, once the particles have been advanced, they can be removed and reinserted into this tree with a cost of order $\mathcal{O}(\log N)$. Together with the dynamic tree updates, which eliminate prediction and tree construction overhead, the cost of the timestep then scales as $\mathcal{O}(M \log N)$.

5.5.3 Dynamic tree updates

If the fraction of particles to be advanced at a given timestep is indeed small, the prediction of *all* particles and the reconstruction of the *full* tree would also lead to significant sources of overhead. However, the geometric structure of the tree, i.e. the way the particles are grouped into a hierarchy, evolves only relatively slowly in time. It would therefore be sufficient to reconstruct this grouping only every couple of timesteps, provided one can still obtain accurate node properties (center of mass, multipole moments) at the current prediction time.

We achieve exactly this by predicting properties of nodes of the tree, instead of *all* the single particles. In order to do this, each node carries a center-of-mass velocity in addition to its position at the time of its construction. New node positions can then be predicted on the fly while the tree is walked, and only nodes that are actually visited need to be predicted. Note that the leaves of the tree point to single particles. If they are used in the force computation, their prediction corresponds exactly to the ordinary prediction as outlined in equation (5.44).

In our scheme we neglect a possible time variation of the quadrupole moment of the nodes. However, we have introduced an additional mechanism that reacts to fast time variations of tree nodes. Whenever a tree node moves more than $0.2\,l$ since the last reconstruction of this part of the tree occurred, the node and all branches of the tree originating from it are completely updated, i.e. the center-of-mass, center-of-mass velocity and quadrupole moments are recomputed from the individual (predicted) particle states.

Finally, the full tree is reconstructed from scratch every once in a while to take into account the slow changes in the grouping hierarchy. We update the tree whenever a total of $0.2N$ force computations have been done since the last full reconstruction. With this criterion the tree construction is only an insignificant fraction of the total computation time, while we have not noticed any significant loss of force accuracy induced by this procedure.

In summary, the algorithms described above result in an integration scheme that can smoothly and efficiently evolve an N-body system containing a very large dynamic range in time scales. At a given timestep, only a small number M of particles are then advanced, and the total time required for that scales as $\mathcal{O}(M \log N)$.

5.5.4 Including SPH

The above time integration scheme may easily be extended to include SPH. Here we also need to integrate the internal energy equation, and the particle accelerations also receive a hydrodynamical component. To compute the latter we also need predicted velocities

$$\tilde{\mathbf{v}}_i = \mathbf{v}_i + \mathbf{a}_{i-1}(\tau_{\rm p} - t_i), \tag{5.47}$$

where we have approximated $\mathbf{a}_i$ with the acceleration of the previous timestep. Similarly, we obtain predictions for the internal energy

$$\tilde{u}_i = u_i + \dot{u}_i(\tau_{\rm p} - t_i), \tag{5.48}$$

and the density of *inactive* particles as

$$\tilde{\rho}_i = \rho_i + \dot{\rho}_i(\tau_{\rm p} - t_i). \tag{5.49}$$

For those particles that are to be advanced at the current system step, these predicted quantities are then used to compute the hydrodynamical part of the acceleration and the rate of change of the internal energy with the usual SPH estimates, as described in Section 5.4.

Note that the timestep criterion (5.40) needs to be supplemented with the Courant condition for the gas particles. We adopt it in the form

$$\Delta t_i = \frac{\alpha_{\rm cour}\, h_i}{h_i|(\nabla\cdot\mathbf{v})_i| + c_i\,(1+0.6\,\alpha_{\rm visc}) + 1.2\,\alpha_{\rm visc}|\mu_i^{\rm max}|}, \tag{5.50}$$

where $\alpha_{\rm visc}$ regulates the strength of the artificial bulk viscosity, and $\alpha_{\rm cour}$ is an accuracy parameter. For the SPH-particles, we use either criterion (5.40) or (5.50), whichever gives the smaller timestep.

5.5.4.1 Additional fixings

As defined above, we evaluate $\mathbf{a}^{\rm gas}$ and $\dot{u}$ at the middle of the timestep, when the actual timestep Δt of the particle that will be advanced *is already set*. Note that in our integration scheme, there is a term in the artificial viscosity that can cause a problem. The second term in equation (5.28) tries to prevent particle inter-penetration. If a particle happens to get very close to another SPH particle in the time $\Delta t/2$, this term can suddenly lead to a very large repulsive acceleration $\mathbf{a}^{\rm visc}$ trying to prevent the particles from getting any closer. However, it is then too late to reduce the timestep. Instead, the velocity of the approaching particle will be changed by $\mathbf{a}^{\rm visc}\Delta t$, possibly *reversing* the approach of the two particles. But the artificial viscosity should at most *halt* the approach of the particles. To guarantee this, we introduce an upper cut-off to the maximum acceleration induced by the artificial viscosity. If $\mathbf{v}_{ij}\cdot\mathbf{r}_{ij} < 0$, we replace equation (5.27) with

$$\tilde{\Pi}_{ij} = \frac{1}{2}(f_i + f_j)\,\min\left[\Pi_{ij}, \frac{\mathbf{v}_{ij}\cdot\mathbf{r}_{ij}}{(m_i+m_j)W_{ij}\Delta t}\right], \tag{5.51}$$

where $W_{ij} = \mathbf{r}_{ij}\nabla_i\left[W(\mathbf{r}_{ij};h_i) + W(\mathbf{r}_{ij};h_j)\right]/2$. With this change, our integration scheme works very well also in regimes with strong shocks. Without this fix, we occasionally got single particles flying out of the shocking regions with high velocities.

5.5.4.2 Implementation of cooling

When radiative cooling is included in simulations of galaxy formation or galaxy interaction, additional numerical problems arise. In regions of strong gas cooling, the cooling times can become so short, that extremely small timesteps would be required to follow the internal energy accurately with the simple explicit scheme used so far.

To remedy this problem, we treat the cooling semi-implicitly in an isochoric approximation. At any given timestep, we first compute the rate $\dot{u}^{\rm ad}$ of change of the internal energy due to the ordinary adiabatic gas physics. In an isochoric approximation, we then solve implicitly for a new internal energy predicted at the end of the timestep, i.e.

$$\hat{u}_i^{(n+1)} = u_i^{(n)} + \dot{u}^{\mathrm{ad}}\Delta t - \frac{\Lambda\left[\rho_i^{(n)}, \hat{u}_i^{(n+1)}\right]\Delta t}{\rho_i^{(n)}}. \tag{5.52}$$

The implicit computation of the cooling rate guarantees stability. Based on this estimate, we compute an effective rate of change of the internal energy, which we take then as

$$\dot{u}_i = \left[\hat{u}_i^{(n+1)} - u_i^{(n)}\right]/\Delta t\,. \tag{5.53}$$

This last step is necessary since our integration scheme requires the possibility to predict the internal energy at arbitrary times. With the above procedure, u_i is always a continuous function of time, and the prediction may be done for all intermediate times.

5.5.5 Integration in comoving coordinates

For simulations in a cosmological context, the expansion of the universe has to be taken into account. Let $\mathbf{x}$ denote comoving coordinates, and a be the dimensionless scale factor ($a = 1.0$ at the present epoch). Then the Newtonian equation of motion becomes

$$\ddot{\mathbf{x}} + 2\frac{\dot{a}}{a}\,\dot{\mathbf{x}} = -G\int \frac{\delta\rho(\mathbf{x}')\,(\mathbf{x}-\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|^3}\mathrm{d}^3x'. \tag{5.54}$$

Here the function $\delta\rho(\mathbf{x}) = \rho(\mathbf{x}) - \overline{\rho}$ denotes the (physical) density fluctuation field.

In an N-body simulation with periodic boundary conditions, the volume integral of equation (5.54) is carried out over *all* space. As a consequence, the homogeneous contribution arising from $\overline{\rho}$ drops out around every point. Then the equation of motion of particle i becomes

$$\ddot{\mathbf{x}}_i + 2\frac{\dot{a}}{a}\,\dot{\mathbf{x}}_i = -\frac{G}{a^3}\sum_{\substack{j\neq i\\ \text{periodic}}}\frac{m_j\,(\mathbf{x}_i-\mathbf{x}_j)}{|\mathbf{x}_i-\mathbf{x}_j|^3}. \tag{5.55}$$

However, one may also employ vacuum boundary conditions. Here, one simulates a spherical region of radius R around the origin, and neglects density fluctuations outside this region. In this case, the background density $\overline{\rho}$ gives rise to an additional term, viz.

$$\ddot{\mathbf{x}}_i + 2\frac{\dot{a}}{a}\,\dot{\mathbf{x}}_i = \frac{1}{a^3}\left[-G\sum_{j\neq i}\frac{m_j\,\mathbf{x}_{ij}}{|\mathbf{x}_{ij}|^3} + \frac{1}{2}\Omega_0 H_0^2\mathbf{x}_i\right]. \tag{5.56}$$

Currently, GADGET does not support periodic boundary conditions, which might in principle be implemented by means of the Ewald summation method (Hernquist et al. 1991). Cosmological simulations therefore need to employ equation (5.56).

In linear theory, it can be shown that the kinetic energy

$$T = \frac{1}{2}\sum_i m_i\mathbf{v}^2 \tag{5.57}$$

in peculiar motion grows proportional to a, at least at early times. This implies that $\sum_i m_i \dot{\mathbf{x}}^2 \propto 1/a$, hence the comoving velocities $\dot{\mathbf{x}} = \mathbf{v}/a$ actually diverge for $a \to 0$. Since cosmological simulations are usually started at redshift $z \simeq 30-100$, one therefore needs to follow a rapid deceleration of $\dot{\mathbf{x}}$ at high redshift. So it is numerically unfavourable to solve the equations of motion in the variable $\dot{\mathbf{x}}$.

To remedy this problem, we instead use a new velocity variable

$$\mathbf{w} \equiv a^{\frac{1}{2}}\,\dot{\mathbf{x}}, \tag{5.58}$$

and we use the expansion factor itself as time variable. Then the equations of motion become

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}a} = \frac{\mathbf{w}}{S(a)}, \tag{5.59}$$

$$\frac{\mathrm{d}\mathbf{w}}{\mathrm{d}a} = -\frac{3}{2}\frac{\mathbf{w}}{a} + \frac{1}{a^2 S(a)}\left[-G\sum_{j\neq i}\frac{m_j\,\mathbf{x}_{ij}}{|\mathbf{x}_{ij}|^3} + \frac{1}{2}\Omega_0 H_0^2 \mathbf{x}_i\right], \tag{5.60}$$

with $S(a)$ given by

$$S(a) = H_0\sqrt{\Omega_0 + a(1-\Omega_0-\Omega_\Lambda) + a^3\Omega_\Lambda} = a^{\frac{3}{2}} H(a). \tag{5.61}$$

Using the Zel'dovich approximation, one sees that $\mathbf{w}$ remains constant in the linear regime. Strictly speaking this holds only for an Einstein-de-Sitter universe at all times, however, it is also true for other cosmologies at early times. Hence equations (5.58) to (5.61) in principle solve linear theory for arbitrarily large steps in a. This allows to traverse the linear regime with maximum computational efficiency. Furthermore, equations (5.58) to (5.61) represent a convenient formulation for general cosmologies, and for our variable timestep integrator. Since $\mathbf{w}$ does not vary in the linear regime, predicted particle positions based on $\tilde{\mathbf{x}}_i = \mathbf{x}_i + \mathbf{w}_i(a_\mathrm{p} - a_i)/S(a_\mathrm{p})$ are quite accurate. Also, the acceleration entering the timestep criterion may now be identified with $\mathrm{d}\mathbf{w}/\mathrm{d}a$, and the timestep becomes

$$\Delta a = \alpha_{\mathrm{tol}}\sigma\left|\frac{\mathrm{d}\mathbf{w}}{\mathrm{d}a}\right|^{-1}. \tag{5.62}$$

The above equations only treated the gravity part of the dynamical equations. However, it is straightforward to express the hydrodynamical equations in the variables ($\mathbf{x}$, $\mathbf{w}$, a) as well. For gas particles, equation (5.60) receives an additional contribution due to hydrodynamical forces, viz.

$$\left(\frac{\mathrm{d}\mathbf{w}}{\mathrm{d}a}\right)_{\mathrm{hydro}} = -\frac{1}{a\,S(a)}\frac{\nabla_{\mathbf{x}} P}{\rho}. \tag{5.63}$$

For the energy equation, one obtains

$$\frac{\mathrm{d}u}{\mathrm{d}a} = -\frac{3}{a}\frac{P}{\rho} - \frac{1}{S(a)}\frac{P}{\rho}\nabla_{\mathbf{x}}\cdot\mathbf{w}. \tag{5.64}$$

Here the first term on the right hand side describes the adiabatic cooling of gas due to the expansion of the universe.

5.6 Parallelization

Massively parallel computer systems with distributed memory have become increasingly popular recently. They can be thought of as a collection of workstations, connected by a fast communication network. This architecture promises large scalability for reasonable cost. Current state-of-the art machines of this type include Cray T3E, IBM SP/2, and Intel Paragon.

However, an efficient use of parallel distributed memory machines often requires substantial changes of existing algorithms, or the development of completely new ones. Conceptionally, parallel programming involves two major difficulties in addition to the task of solving the numerical problem in a serial code. First, there is the difficulty of how to divide the work and data *evenly* among the processors, and second, a communication scheme between the processors needs to be devised.

In recent years, a number of groups have developed parallel N-body codes, all of them with different parallelization strategies, and different strengths and weaknesses. An early attempt for parallelization has been done by Theuns & Rathsack (1993). Later, Warren et al. (1992) parallelized the BH-tree code for the first time on massively parallel machines. Dubinski (1996) presented the first parallel tree code based on MPI. Also, Dikaiakos & Stadel (1995) have developed a parallel simulation code (PKDGRAV), that works with a balanced binary tree.

We have developed a new parallel tree code, that implements for the first time individual particle timesteps. We have used the *Message Passing Interface* (MPI) (Pacheco 1997; Snir et al. 1995), which is an explicit communication scheme, i.e. it is entirely up to the user to control the communication. Messages containing data can be sent between processors, both in synchronous and asynchronous modes. A particular advantage of MPI is its flexibility and portability. Our simulation code uses only standard C and standard MPI, and should therefore run on a variety of platforms. We have confirmed this so far on Cray T3E and IBM SP/2 systems.

5.6.1 Domain decomposition

The typical size of problems attacked on parallel computers is usually much too large to fit into the memory of individual computational nodes, or into ordinary workstations. This fact alone, but of course also the desire to distribute the work among the processors, requires a partitioning of the problem onto the individual processors.

For our N-body/SPH code we have implemented a spatial domain decomposition. The particular algorithm we use for the split is an orthogonal recursive bisection (ORB) (Dubinski 1996). In the first step, a split is found along one spatial direction, e.g. the x-axis, and the collection of processors is grouped into two halves, one for each side of the split. These processors then exchange particles such that they end up hosting only particles lying on their side of the split. In the simplest possible approach, the position of the split is chosen such that there are an equal number of particles on both sides. However, for an efficient simulation code the split should try to balance the work done

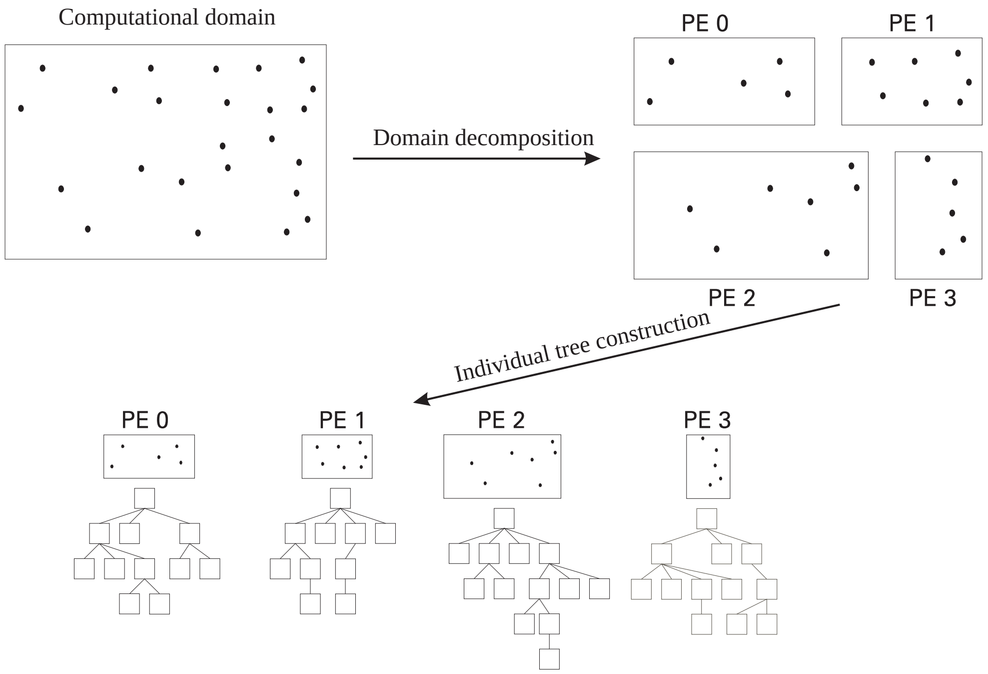


Figure 5.2: Schematic representation of the domain decomposition in two dimensions, and for four processors. Here, the first split occurs along the y-axis, separating the processors into two groups. They then independently carry out a second split along the x-axis. After completion of the domain decomposition, each processor element (PE) can construct its own BH tree just for the particles in its part of the computational domain.

in the force computation on both sides. This aspect will be discussed further below.

In a second step, each group of processors finds a new split along a different spatial axis, e.g. the y-axis. This splitting process is repeated recursively until the final groups consist of just one processor, which then hosts a rectangular piece of the computational volume. Note that this algorithm constrains the number of processors that may be used to a power of two.

A two-dimensional schematic illustration of the ORB is shown in Figure 5.2 for four processors. Note that each processor can construct a local BH tree for its domain, and this tree may be used to compute the force exerted by the processors' particles on arbitrary test particles in space.

5.6.2 Parallel force computation

Our algorithm for parallel force computation differs significantly from that of Dubinski (1996). He uses the notion of *locally essential trees*. These are trees that are sufficiently detailed to allow the full force computation for any particle local to a processor, without

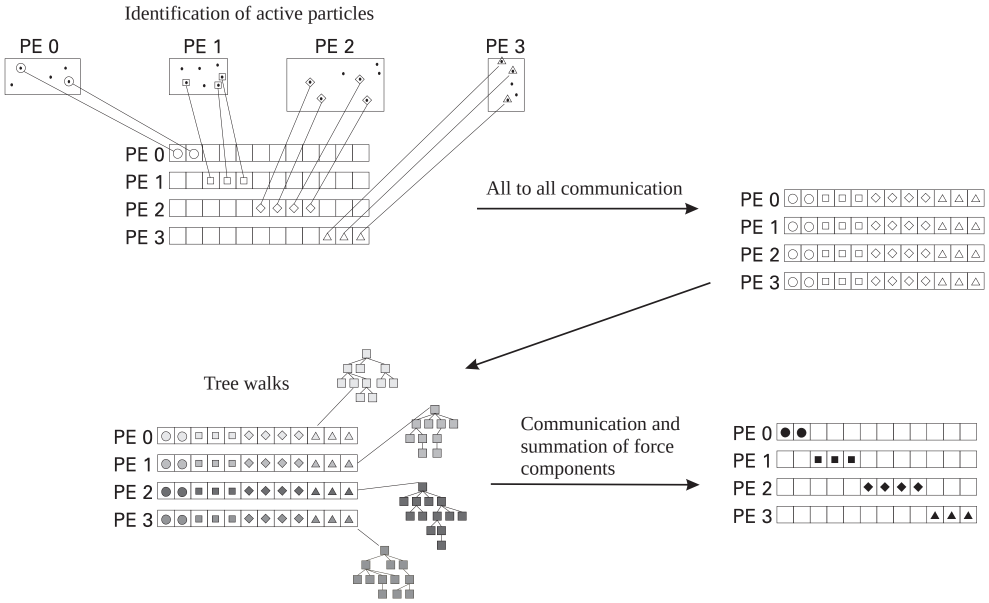


Figure 5.3: Schematic illustration of the parallelization scheme of GADGET for the force computation. In the first step, each PE identifies the active particles, and puts their coordinates in a communication buffer. In a communication phase, a single and identical list of all these coordinates is then established on all processors. Then each PE walks its local tree for this list, thereby obtaining a list of force contributions. These are then communicated in a collective process back to the original PE that hosts the corresponding particle coordinate. Each processor then sums up the incoming force contributions, and finally arrives at the required total forces.

further need for information from other processors. The locally essential trees can be constructed from the local trees by pruning and exporting parts of these trees to other processors, and attaching these parts as new branches to the local trees. To determine which parts of the trees need to be exported, special tree walks are required.

A disadvantage of this technique is that the construction of the locally essential trees is a complicated, time-consuming process that needs to be repeated every time-step. While this should not be an issue for an integration scheme with a global timestep, the tree construction time would dominate the computational time for our individual timestep scheme, where typically of order 1 per cent of all particles require a force update at one of the (small) system timesteps. Therefore we chose a different parallelization scheme that essentially scales linearly with the number of particles that need a force computation.

Our strategy starts from the observation that each of the local processor trees is able to provide the force exerted by its particles for any location in space. The full force might thus be obtained by adding up all the force contributions from the local trees. As long as the number of these trees is less than the number of typical force-node interactions, this computational scheme is practically not more expansive than a tree walk of the corresponding essential tree.

A force computation therefore requires a communication of the desired coordinates to all processors. These then walk their local trees, and send the force components back to the original processor that sent out the corresponding coordinate. The total force then results by summing up the incoming contributions.

In practice, a force computation for a *single* particle would be badly imbalanced in work in such a scheme, since some of the processors could stop their tree walk already at the root node, while others would have to evaluate several hundred particle-node interactions. However, our time integration scheme advances at a given timestep always a group M of particles of size about 0.5-5 per cent of the total number of particles. This group represents a representative mix of the various clustering states of matter in the simulation. Each processor contributes some of its particle positions to this mix, but the total list of coordinates is the same *for all* processors. If the domain decomposition is done well, one can arrange that the cumulative time to walk the local tree for all coordinates in the list is the same for all processors.

Our force computation scheme proceeds therefore as sketched schematically in Figure 5.3. Each processor identifies the particles that are to be advanced in the current timestep, and puts their coordinates in a communication buffer. Next, an all-to-all communication phase is used to establish the same list of coordinates on all processors. This communication is done in a collective fashion: For $N_{\rm p}$ processors, the communication involves $N_{\rm p}-1$ cycles. In each cycle, the processors are arranged in $N_{\rm p}/2$ pairs. Each pair exchanges their original list of coordinates. In this way, the communication is done fully in parallel, and the time it requires scales as $\mathcal{O}(MN_{\rm p})$. On the T3E, the communication bandwidth is large enough, that only a small fraction of the overall simulation time is spent in this phase.

In the next step, all processors walk their local trees and replace the coordinates with the corresponding force contribution. Note that this is the most time-consuming step of the simulation (as it should), hence work-load balancing is most crucial here. After that, the force contributions are communicated in a similar way as above between the processor pairs. The processor that hosted a particular coordinate adds up the incoming force contributions and finally ends up with the full force for that location. This can then be used to advance its particles, and determine new timesteps for them. In these phases of the N-body algorithm, as well as in the tree construction, no further information from other processors is required.

Currently, the parallelization of the SPH part of the code is still in an experimental stage. In principle, hydrodynamics is easier to parallelize than gravity, because it is a *local* interaction. Note that in contrast to this, the gravitational N-body problem has the nasty property that at all times *each particle interacts with every other particle*, making gravity intrinsically difficult to parallelize on distributed memory machines.

Most SPH particles do not rely on information from other domains. However, there are those whose smoothing kernel extends across domain boundaries. Here it becomes a tricky problem of designing an efficient communication and bookkeeping algorithm to exchange neighbours across domain boundaries. We are still working on our solution.

5.6.3 Work-load balancing

Due to the high communication bandwidth of parallel supercomputers like the T3E or the SP/2, the time required for force computation is dominated by the tree walks, and this is also the dominating part of the simulation as a whole. It is therefore crucial that this part of the computation parallelizes well. In the context of our parallelization scheme, this means that the domain decomposition should be done such that the time spent in the tree walks is the same for all processors.

It is helpful to note, that the list of coordinates for the tree walks is *independent* of the domain decomposition. However, we can think of each patch of space, represented by its particles, to cause some *cost* in the tree-walk process. A good measure for this cost is the number of particle-node interactions originating from this region of space. To balance the work, the domain decomposition should therefore try to make this cost equal on both sides of each domain split.

In practice, each tree-node carries a counter for the number of node-particle interactions it participated in since the last domain decomposition. Before a new domain decomposition starts, we then need to assign this cost to individual particles. For this purpose, we have implemented a method to walk the tree backwards from a leave (i.e. a single particle) to the root node. In this walk, the particle collects its total cost by adding up its share of the cost from all its parent nodes. In this respect, our work-load balancing scheme differs from that of Dubinski (1996).

Note that an optimum work-load balance often results in substantial memory imbalance. Tree-codes consume plenty of memory, so that the problem size is usually memory rather than CPU-time limited. For example, a single node with 128 Mbyte on the Garching T3E is already filled to roughly 60 per cent with 1.5×10^5 particles, including all memory for the tree structures. In this example, the remaining free memory would usually not be enough to guarantee optimum work-load balancing in strongly clustered simulations. Unfortunately, such a situation is not untypical in practice, since one usually strives for *large N* in N-body work, so one is always tempted to fill up most of the available memory with particles, without leaving much room to balance the work-load. Of course, GADGET can only try to balance the work within the limits set by the available memory.

5.7 Results and tests

5.7.1 Tests of timestep criteria

As first tests of the relative performance of different timestep criteria, we integrate a particle orbit in the potential of a point mass, and in the potential of an isothermal sphere. We consider three simple timestep criteria, which are based on the acceleration

Table 5.1: Test of the timestep criteria (5.65) to (5.67) for the integration of Kepler ellipses in the field of a point mass. The free parameters of the criteria are set such that they all integrate a circular orbit using the same number of timesteps. We then consider orbits of increasing eccentricity ϵ, keeping the total energy of the orbiting particle constant. Averaged over five orbits, the table lists the mean relative drift of the total energy, and the number in parenthesis gives the number of force computations per orbit. An ∞-sign means that the integration catastrophically failed.

ϵ	A	B	C
0	$< 10^{-10}$ (50)	$< 10^{-10}$ (50)	$< 10^{-10}$ (50)
0.5	-3.0×10^{-3} (57)	-1.9×10^{-3} (50)	-5.5×10^{-3} (70)
0.8	-0.018 (82)	-0.038 (48)	-0.041 (158)
0.9	-0.034 (113)	∞	-0.076 (311)
0.99	-0.072 (352)	∞	∞

and/or velocity of the particle in the previous timestep. These criteria are

$$\text{(A)} \qquad \Delta t = \frac{\eta_{\mathrm{A}}}{|\mathbf{a}|}, \tag{5.65}$$

$$\text{(B)} \qquad \Delta t = \frac{\eta_{\mathrm{B}}}{\sqrt{|\mathbf{a}|}}, \tag{5.66}$$

$$\text{(C)} \qquad \Delta t = \frac{\eta_{\mathrm{C}}}{|\mathbf{v}||\mathbf{a}|}. \tag{5.67}$$

Criterion (A) is the of form suggested in this work, while both (B) and (C) have been used in previous studies (among others: Heller & Shlosman 1994; Quinn et al. 1997).

We now consider a test particle in the field of a point mass. We adjust the coefficients η_{A}, η_{B}, η_{C} of the timestep criteria such that they all integrate a circular orbit of given radius with $N = 50$ force computations. For the circular orbit, the timestep remains fixed, and the integration is symplectic for all three criteria, without any secular drift in energy. It is now interesting to consider how the various criteria perform when the eccentricity ϵ of the orbit is increased. In Table 5.1, we show the drift in total energy per orbit (averaged over 5 orbits) and the number of force computations needed per orbit for the three timestep criteria. Interestingly, only criterion (A) is able to stably integrate orbits of very high eccentricity. Furthermore, it delivers higher accuracy than (C) at a smaller computational cost. Similarly, (A) clearly outperforms (B), at least for $\epsilon \geq 0.5$.

Table 5.2: Test of timestep criteria for the integration of orbits in the potential of a singular isothermal sphere. We set the free parameters of the criteria (5.65) to (5.67) such that they all need the same number of timesteps for a circular orbit. We then consider orbits with smaller velocity $v = \kappa V_c$ at apocenter. The table lists the mean relative drift of total energy over a time $2\pi V_c/a_0$. The number in parenthesis gives the number of force computations occurring over that time.

κ	A	B	C
0	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$
	(50)	(50)	(50)
0.5	-8.3×10^{-4}	2.3×10^{-4}	-3.4×10^{-3}
	(78)	(61)	(85)
0.2	-7.7×10^{-3}	0.035	-0.037
	(101)	(63)	(132)
0.1	-0.021	-0.17	-0.13
	(117)	(99)	(196)
0.05	-0.047	∞	-0.28
	(135)		(341)

Of larger relevance for cosmological N-body simulations is the potential of an isothermal sphere, which may serve as a simple approximation to a dark matter halo. Using the potential

$$\Phi(r) = -V_c^2 \left[\log\left(\frac{r}{a_0}\right) - 1 \right], \tag{5.68}$$

we perform a similar test as above. We set the free parameters of the timestep criteria such that they require 50 force computations for a circular orbit of radius a_0. We now consider orbits with a velocity v at apocenter a_0 which is just a fraction $\kappa = v/V_c$ of the circular velocity. In Table 5.2, we list the mean relative drift of the total energy, given for a time interval $2\pi a_0/V_c$, and we list the number of force computations needed over that time. Again, the criterion (A) by far outperforms the other two criteria (B) and (C). For a relatively small number of force computations it provides both the best stability and the best accuracy for a given computational expense.

5.7.2 Force accuracy and opening criterion

In Section 5.3.1.2, we have pointed out that standard values of the BH-opening criterion can result in very high force errors for the initial conditions of cosmological N-body simulations. Here, the density field is very close to being homogeneous, so that the small peculiar accelerations arise from a cancellation of relatively large partial forces. We now investigate this problem using a sphere of radius $7\,h^{-1}$Mpc, carved out of the

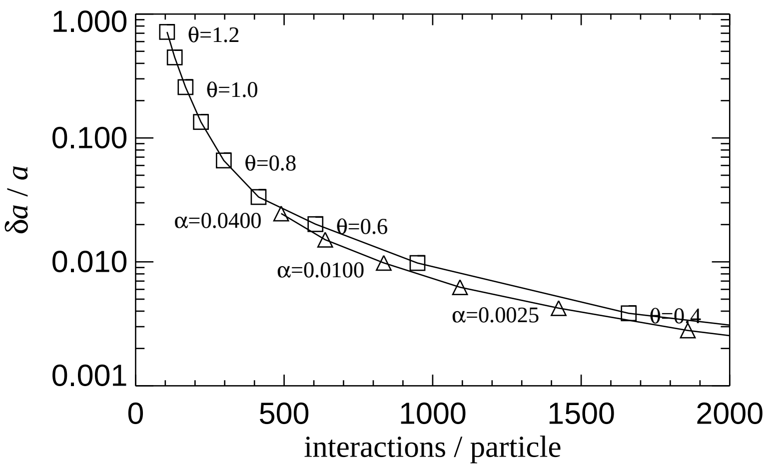


Figure 5.4: Force errors for the initial conditions of a cosmological simulation at $z = 80$. We show results for the BH-criterion (boxes), and our new opening criterion (triangles) as a function of their tolerance parameters. The plot gives the 90% percentile of the error distribution, i.e. 90% of the particles have force errors below the cited values. The horizontal axes measures the computational expense.

initial conditions (redshift $z = 80$) of a high-resolution cosmological simulation. This sphere contains slightly more than 500000 particles of mass $m_{\rm p} = 2.38 \times 10^8 h^{-1}{\rm M}_\odot$. To determine the distribution of force errors, we computed the true force using direct summation for a random sample of 1000 particles, and we compared it to the force obtained by the tree algorithm.

In Fig. 5.4, we show the force errors measured for the BH opening criterion and for our new criterion (5.20) as a function of θ and α, respectively. The plot shows the 90% percentile of the error distribution, i.e. 90% of the particles have force errors below the cited values. Note that the horizontal axes measures the computational expense of the force computation in terms of the average number of particle-cell interactions per force evaluation. What is interesting are the very large errors resulting from the BH criterion when 'normal' values like $\theta \simeq 0.8 - 1.0$ are used. Our new criterion does better in this respect. Since it tries to make the absolute error in any particle-cell interaction smaller than some fraction of the *total* force, the error on the latter always stays within reasonable bounds.

However, an even more important strength of the new criterion lies in its better performance for clustered particle distributions. In Figure 5.5, we show results of a force-accuracy test for an isolated galaxy with 30000 dark matter particles, and an embedded stellar disk with 20000 particles. Comparing these results with Figure 5.4, the most striking difference is the very different performance of the BH criterion in

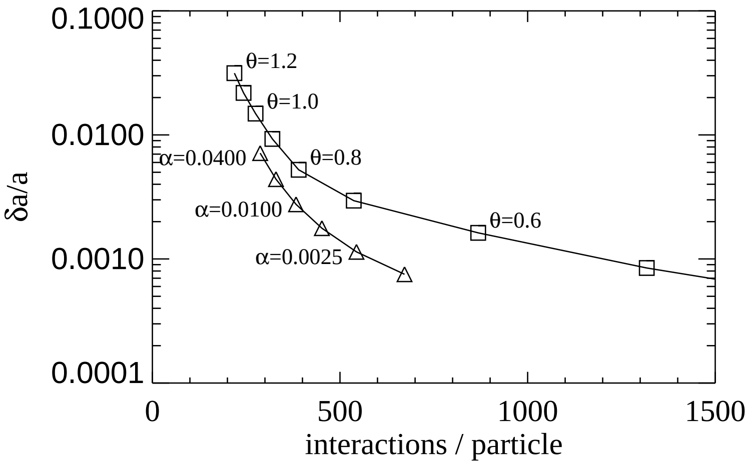


Figure 5.5: Force errors for an isolated halo/disk galaxy with the BH-criterion (boxes), and the new opening criterion (triangles). The plot shows the 90% percentile of the error distribution, i.e. 90% of the particles have force errors below the cited values. The horizontal axes measures the computational expense.

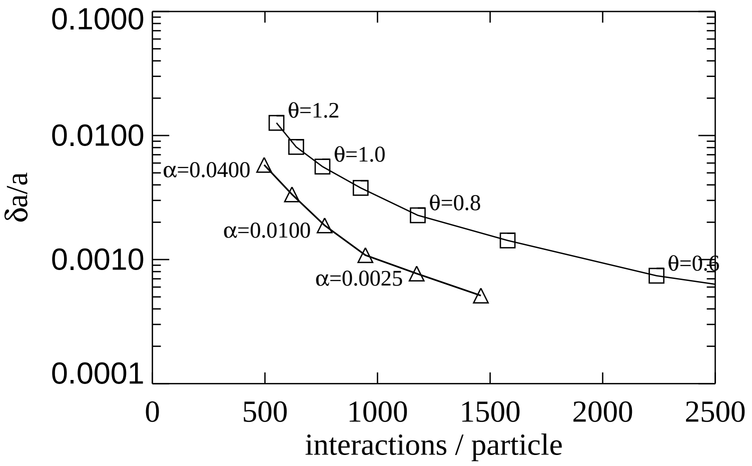


Figure 5.6: Force accuracy in a cosmological simulation of a cluster of galaxies, measured at $z = 0$. Using 16 processors, the parallel version of the code was used for this test. We plot the 90% percentile of the error distribution, i.e. 90% of the particles have force errors below the cited values. Results are given for the BH-criterion (boxes), and the new opening criterion (triangles).

the two different regimes. While already $\theta = 1.0$ gives sufficient force accuracy for the clustered mass distribution of the galaxy, this value leads to intolerable force errors for a close-to-homogeneous mass distribution. The new opening criterion does not suffer from such a severe accuracy degradation in the homogeneous case; its error remains well controlled.

Another advantage of the new criterion is that it is actually more efficient than BH, i.e. it achieves higher force accuracy for a given computational expense. For a strongly clustered particle distribution in a cosmological simulation, the implied saving can easily reach a factor 2-3, and the simulation is sped up by the same factor. This can be understood as follows: The purely geometrical BH criterion doesn't care for the dynamical significance of the mass distribution. For example, it will invest a large number of cell-particle interactions to compute the force exerted by a large void to a high relative accuracy, while actually this force might be of just small absolute size, and it would be much better to concentrate on those regions that provide most of the force on the current particle. The new opening criterion follows this latter strategy, improving the force accuracy at a given number of particle-cell interactions. This can be seen in Figure 5.6 where we plot the results of a measurement of the force accuracy at $z = 0$ in a cosmological simulation with 3.2 million particles.

5.7.3 Colliding disk galaxies

As a test of the performance and accuracy of the integration of collisionless systems, we here consider a pair of merging disk galaxies. Each galaxy has a massive dark halo consisting of 30000 particles, and an embedded stellar disk, modeled with 20000 particles. The dark halo is modeled according to the NFW-profile, adiabatically modified by the central exponential disk, which contributes 5% of the total mass. The halo has a circular velocity $v_{200} = 160\,\mathrm{km\,s^{-1}}$, a concentration $c = 5$, and spin parameter $\lambda = 0.05$. The radial exponential scale length of the disk is $R_\mathrm{d} = 4.5\,h^{-1}\mathrm{kpc}$, while the vertical structure is that of an isothermal sheet with thickness $z_0 = 0.2R_\mathrm{d}$. The gravitational softening of the halo particles is set to $0.4\,h^{-1}\mathrm{kpc}$, and that of the disk to $0.1\,h^{-1}\mathrm{kpc}$.

Initially, the two galaxies are set-up on a parabolic orbit, with separation such that their dark haloes just touch. Both of the galaxies have a prograde orientation, but are inclined with respect to the orbital plane. In fact, the test considered here corresponds exactly to the simulation 'C1' computed in Chapter 3, where more information about the construction of the initial conditions can be found (see also Chapter 2).

We first consider a run of this model with a set of parameters equal to our typically employed values for a simulation of this kind. For the time integration, we used the parameter $\alpha_\mathrm{tol}\sigma = 25\,\mathrm{km\,s^{-1}}$, and for the force computation with the tree algorithm, the new opening criterion with $\alpha = 0.04$. The simulation was then run from $t = 0$ to $t = 2.8$ in internal units of time (corresponding to $2.85\,h^{-1}\mathrm{Gyr}$). During this time the galaxies have their first close encounter at around $t \simeq 1.0$, where tidal tails are ejected out of the stellar disks. Due to the braking by dynamical friction, the galaxies eventually fall together for a second time, after which they quickly merge and violently relax to form

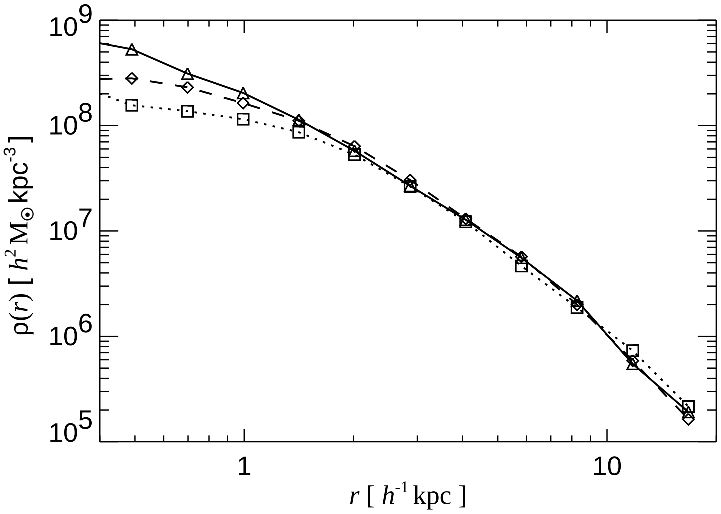


Figure 5.7: Spherically averaged density profile of the stellar component in the merger remnant of two colliding disk galaxies. Triangles show the results obtained using our variable timestep integration, while boxes and diamonds are for fixed timestep integrations with $\Delta t = 0.01$ and $\Delta t = 0.0025$, respectively. Note that the simulation using the adaptive timestep is about as expensive as the one with $\Delta t = 0.01$. In each case, the center of the remnant was defined as the position of the particle with the minimum gravitational potential.

a single merger remnant. At $t = 2.8$, the inner parts of the merger remnant are well relaxed.

This simulation required a total number of 4684 steps and 27795733 force computations, i.e. a computationally equally expansive computation with fixed timesteps could just make 280 full timesteps. The relative error in the total energy was 3.0×10^{-3}, and a Sun Ultrasparc-II workstation did the simulation in 4 hours. Note that the raw force speed with $\sim$2800 force computations per second was really very high – older workstations will achieve somewhat smaller values, of course.

We now consider a simulation of the same system using a fixed timestep. For $\Delta t = 0.01$, the run needs 280 full steps, i.e. it consumes the same amount of CPU time as above. However, in this simulation, the error in the total energy is 2.2%, substantially larger than before. There are also differences in the density profile of the merger remnants. In Figure 5.7, we compare the inner density profile of the simulation with adaptive timesteps (triangles) to the one with a fixed timestep of $\Delta t = 0.01$ (boxes). We here show the spherically averaged profile of the stellar component, with the center of the remnants defined as the position of the particle with the minimum of the gravitational potential. It can be seen that in the innermost $\sim 1\,h^{-1}$kpc, the density obtained with the fixed timestep falls short of the adaptive timestep integration.

To get an idea how small the fixed timestep has to be to achieve similar accuracy as with the adaptive timestep, we have simulated this test a second time, with a fixed timesteps of $\Delta t = 0.0025$. We also show the corresponding profile (diamonds) in Figure 5.7. It can be seen that for smaller Δt, the agreement with the variable timestep result improves. At $\sim 2 \times 0.4\,h^{-1}$kpc, the softening of the dark matter starts to become important. For agreement down to this scale, the fixed timestep needs to be slightly smaller than $\Delta t = 0.0025$, so the overall saving due to the use of individual particle timesteps is at least a factor of $4 - 5$ in this example.

5.7.4 Collapse of a cold gas sphere

The self-gravitating collapse of an initially isothermal, cool gas sphere has been a common test problem of SPH codes (Evrard 1988; Hernquist & Katz 1989; Steinmetz & Müller 1993; Thacker et al. 1998; Carraro et al. 1998). Following these authors, we consider a spherically symmetric gas cloud of total mass M, radius R, and initial density profile

$$\rho(r) = \frac{M}{2\pi R^2}\,\frac{1}{r}. \tag{5.69}$$

We take the gas to be isothermal initially, with an internal energy per unit mass of

$$u = 0.05\,\frac{GM}{R}. \tag{5.70}$$

At the start of the simulation, the gas particles are at rest. We obtain their initial coordinates from a distorted regular grid that reproduces the density profile (5.69), and we use a system of units with $G = M = R = 1$.

In Figure 5.8, we show the evolution of the potential, the thermal, and the kinetic energy of the system from the start of the simulation at $t = 0$ to $t = 3$. We plot results for two simulations, one with 30976 particles (solid), and one with 4224 particles (dashed). During the central bounce between $t \approx 0.8$ and $t \approx 1.2$ most of the kinetic energy is converted into heat, and a strong shock wave travels outward. Later, the system slowly settles to virial equilibrium.

For these runs $N_{\rm s}$ was set to 40, the gravitational softening to $\epsilon = 0.02$, and time integration was controlled by the parameters $\alpha_{\rm tol}\sigma = 0.05$ and $\alpha_{\rm cour} = 0.1$[1], resulting in *very* good energy conservation. The absolute error in the total energy is less than 1.1×10^{-3} *at all times* during the simulation, translating to a relative error of 0.23%. Since we use a time integration scheme with individual timesteps of arbitrary size, this small error is particularly remarkable. The total number of small steps taken by the 4224 particle simulation was 3855, with a total of 2192421 force computations, i.e. the equivalent number of 'full' timesteps was 519. A Sun Ultrasparc-II workstation needed 2300 seconds for the simulation. The larger 30976 particle run took 10668 steps, with an equivalent of 1086 'full' steps, and 12 hours of CPU time. Note that by reducing the

[1] Note that our definition of the smoothing length h differs by a factor of 2 from most previous SPH implementations. As a consequence, corresponding values of $\alpha_{\rm cour}$ are different by a factor of 2, too.

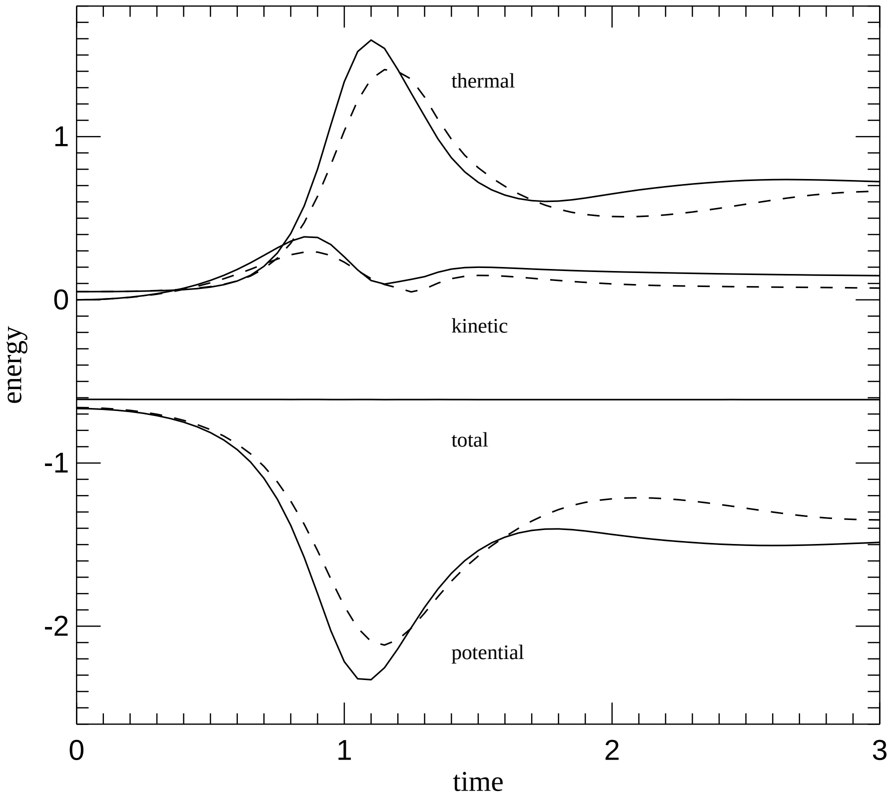


Figure 5.8: Time evolution of the thermal, kinetic, potential, and total energy for the collapse of an initially isothermal gas sphere. Solid lines show results for a simulations with 30976 particles, dashed lines are for a 4224 particle run.

time integration accuracy by a factor of 2, with a corresponding reduction of the CPU time consumption by the same factor, the results do practically not change, however, the maximum error in the energy goes up to 1.2% in this case.

The results of Figure 5.8 agree very well with those of Steinmetz & Müller (1993), and with Thacker et al. (1998), if we compare to their best implementation of artificial viscosity (their version 12). Steinmetz & Müller (1993) have also computed a solution of this problem with a very accurate, one-dimensional, spherically symmetric piecewise parabolic method (PPM). For particle numbers above 10000, our SPH results become very close to the finite difference result. However, even for very small particle numbers, SPH is capable of reproducing the general features of the solution very well. We also note that a three-dimensional PPM calculation of the collapse would require *substantially* more CPU time as our SPH calculations.

5.7.5 Performance and scalability of the parallel code

We here show a simple test of the performance of the parallel version of the code under conditions relevant for real target applications. For this test, we have used a 'stripped-down' version of the initial conditions originally constructed for a high-resolution simulation of a cluster of galaxies. This original set of initial conditions was set-up to follow the cosmological evolution of a large spherical region with comoving radius $70\,h^{-1}$Mpc, within a cosmogony corresponding to a ΛCDM universe with $\Omega_0 = 0.3$, $\Omega_\Lambda = 0.7$, $z_{\rm start} = 50$, and $h = 0.7$. In the center of the simulation sphere, 2 million high-resolution particle were placed in the somewhat enlarged Lagrangian region of the cluster. The rest of the volume was filled with an extended shell of boundary particles of larger mass and larger softening; they are needed for a proper representation of the gravitational tidal field.

To keep our test simple, we have cut out a sphere of comoving radius $12.25\,h^{-1}$Mpc around the origin from these initial conditions, and we only simulated the 500000 high-resolution particles with mass $m_{\rm p} = 1.36 \times 10^9 h^{-1}{\rm M}_\odot$ found within this region. Such a simulation will not be useful for direct scientific analysis because it does not model the tidal field properly. However, this test will show realistic clustering and time-stepping behaviour, and thus allows a reasonable assessment of the expected computational cost of the full problem.

We have run the test problem with GADGET on the Garching T3E from redshift $z = 50$ to redshift $z = 4.3$. We repeated the identical run on partitions of size 4, 8, and 16 processors. In this test, we included quadrupole moments in the tree computation, we used a BH opening criterion with $\theta = 1.0$, and a gravitational softening length of $15\,h^{-1}$kpc.

In Table 5.3 we list in detail the elapsed wall-clock time for various parts of the code for the three simulations. The dominant sink of CPU time is the computation of gravitational forces for the particles. To advance the test simulation from $z = 50$ to $z = 4.3$, GADGET needed 30.0×10^6 force computations in a total of 3350 timesteps. Note that on average only 1.8% of the particles are advanced in a single timestep. Under these conditions it is challenging to eliminate sources of overhead incurred by the time-stepping and to maintain work-load balancing. GADGET solves this task satisfactory. If we would use a fixed timestep, the work-load balancing would essentially be perfect. Note that the use of our adaptive timestep integrator results in a saving of about a factor of 3-5 compared to a fixed timestep scheme with the same accuracy.

We think that the overall performance of GADGET is quite good in this test. The raw gravitational speed is very high, and the algorithm used to parallelize the force computation scales reasonably well, as is seen in the upper left panel of Figure 5.9. Note that the force-speed of the $N_{\rm p} = 8$ run is even higher than that of the $N_{\rm p} = 4$ run. This is because the domain decomposition does exactly one split in the x-, y-, and z-directions in the $N_{\rm p} = 8$ case. The domains are then close to cubes, which reduces the depth of the tree and speeds up the tree-walks.

Also, the force communication does not involve a significant communication over-

Table 5.3: Consumption of CPU-time in various parts of the code for a cosmological run from $z = 50$ to $z = 4.3$. The table gives timings for runs with 4, 8 and 16 processors on the Garching Cray T3E. The computation of the gravitational forces is by far the dominant computational task. We have further split up that time into the actual tree-walk time, the tree-construction time, the time for communication and summation of force contributions, and into the time lost by work-load imbalance. The potential computation is done only once in this test (it can optionally be done in regular intervals to check energy conservation of the code). 'Miscellaneous' refers to time spent in advancing and predicting particles, and in managing the binary tree for the timeline. I/O time for writing a snapshot file (groups of processors can write in parallel) is only 1-2 seconds, and therefore not listed here.

	4 processors		**8 processors**		**16 processors**	
	cumulative time [sec]	relative time	cumulative time [sec]	relative time	cumulative time [sec]	relative time
total	8269.0	100.0 %	4074.0	100.0 %	2887.5	100.0 %
gravity	7574.6	91.6 %	3643.0	89.4 %	2530.3	87.6 %
tree walks	5086.3	61.5 %	2258.4	55.4 %	1322.4	45.8 %
tree construction	1518.4	18.4 %	773.7	19.0%	588.4	20.4 %
communication & summation	24.4	0.3 %	35.4	0.9 %	54.1	1.9 %
work-load imbalance	901.5	10.9 %	535.1	13.1 %	537.4	18.6 %
domain decomposition	209.9	2.5 %	158.1	3.9 %	172.2	6.0 %
potential computation (optional)	46.3	0.2 %	18.1	0.4 %	11.4	0.4 %
miscellaneous	438.3	5.3 %	254.8	6.3 %	173.6	6.0 %

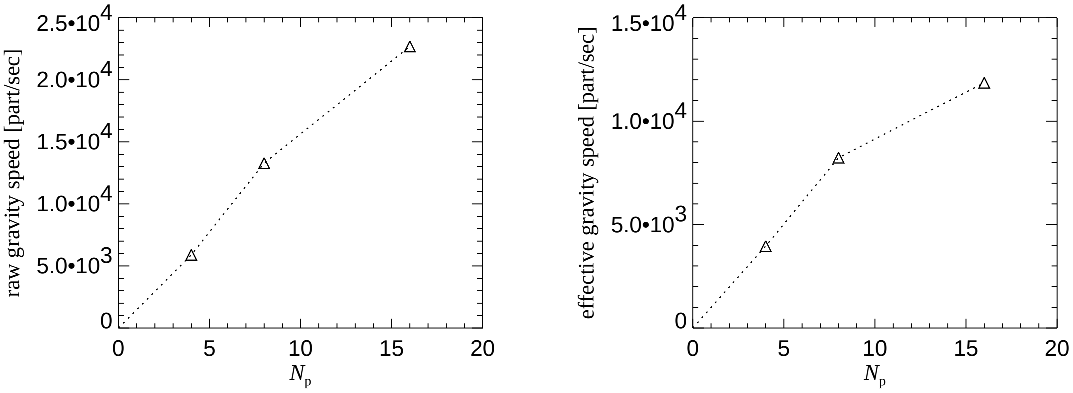


Figure 5.9: Code performance and scalability. The top panel shows the speed of the gravitational force computation as a function of processor number (in particles per second). This is based on the tree walk time alone. In the simulation, additional time is needed for tree construction, work-load imbalance, communication, domain decomposition, etc. This reduces the 'effective' speed, as shown in the bottom panel. This effective speed gives the number of particles advanced by one timestep per second. Note that the only significant source of work-load imbalance in the code occurs in the gravity computation, where some small fraction of time is lost when processors idle wait for others to finish their tree-walks.

head, and the time spent in miscellaneous tasks of the simulation code scales closely with processor number. Most losses of GADGET occur due to work-load imbalance in the force computation. While we think these losses are acceptable in the above test, one should keep in mind that we here kept the problem size *fixed*, and just increased the processor number. If we *also scale up the problem size*, work-load balancing will be significantly easier to achieve, and the efficiency of GADGET will be nearly as high as for small processor number.

5.8 Discussion

We have presented the numerical algorithms of our new code GADGET, designed as a flexible tool to study a wide range of problems in cosmology. Typical applications of GADGET can include interacting and colliding galaxies, star formation and feedback in the interstellar medium, formation of clusters of galaxies, or the formation of large-scale structure in the universe.

In fact, GADGET has already been used successfully in all of these areas. Using our code, the formation of tidal tails in colliding galaxies was studied in Chapter 2, and star formation and feedback in isolated and colliding gas-rich spirals was modeled in Chapter 3. For these simulations, the serial version of the code was employed, both with and without support by the GRAPE special-purpose hardware.

The parallel version of GADGET has been used to compute high-resolution N-body simulations of clusters of galaxies, as will be discussed in Chapter 4. In the largest simulation of this kind, 69 million particles have been employed, with 20 million of them ending up in the virialized region of a single object. The particle mass in the high-resolution zone was just 10^{-10} of the total simulated mass, and the gravitational softening length was $0.7\,h^{-1}$kpc in a simulation volume of diameter $140\,h^{-1}$Mpc, translating to an impressive spatial dynamic range of 2×10^5 in three dimensions.

We have also successfully employed GADGET for two 'constrained-realization' (CR) simulations of the Local Universe. In these simulations, the observed density field as seen by *IRAS* galaxies has been used to constrain the phases of the waves of the initial fluctuation spectrum. As a result, the CR simulations produce large-scale structure with main features in direct correspondence to the most striking objects of the Local Universe. These model universes have a Perseus-Pisces cluster, a Great Attractor, a Cetus-Wall, a Coma Cluster, etc., just like the real Local Universe. The CR-simulations will be particularly useful for generating mock galaxy catalogues that directly mimic all selection biases arising from our special place in the Universe. For each of the two CR simulations, we employed 74 million particles in total, with 53 million high-resolution particles of mass $3.6 \times 10^9 h^{-1}\mathrm{M}_\odot$ (ΛCDM) or $1.2 \times 10^{10} h^{-1}\mathrm{M}_\odot$ (τCDM) in the low-density and critical-density models, respectively.

The main technical features of GADGET are as follows. Gravitational forces are computed with a Barnes & Hut oct-tree, using multipole expansions up to quadrupole order. The cell-opening criterion may be chosen either as the standard BH-criterion, or a new criterion which we have shown to be computationally more efficient and better suited to cosmological simulations starting at high redshift. As an alternative to the tree-algorithm, the serial code can use the special-purpose hardware GRAPE both to compute gravitational forces and for the search of SPH neighbours.

In our SPH implementation, the number of smoothing neighbors is kept exactly constant, leading to very good energy conservation. We also manifestly conserve momentum, i.e. force symmetry is guaranteed for every interacting pair of SPH particles. We use a shear-reduced artificial viscosity that has emerged as a very good parameterization in recent systematic studies that compared several alternative formulations (Thacker et al. 1998; Lombardi et al. 1998).

Parallelization of the code for massively parallel supercomputers is achieved in an explicit message passing approach, using the MPI standard communication library. The simulation volume is spatially split in a recursive orthogonal bisection, and each of the resulting domains is mapped onto one processor. Dynamic work-load balancing is achieved by measuring the computational expense incurred by each particle, and balancing the sum of these weights in the domain decomposition.

The code allows fully adaptive, individual particle timesteps, both for collisionless particles and for SPH particles. The speed-up obtained by the use of individual timesteps depends on the dynamic range of the time scales present in the problem, and on the relative population of these time scales with particles. For a cosmological simulation with a gravitational softening length larger than $\sim 30\,h^{-1}$kpc the overall saving is typically a factor of $3-5$. However, if smaller softening lengths are desired, the use of individual particle timesteps results in much larger savings. For collisionless particles, we use a timestep criterion based on the inverse of the particle acceleration. We have shown that this criterion clearly outperforms simple alternative criteria that are currently in use. Our code can be used both to run simulations in physical and in comoving coordinates. The latter is used for cosmological simulations only. Here, the code employs an integration scheme that can deal with arbitrary cosmological background models, and which is exact in linear theory, i.e. the linear regime can be traversed with maximum efficiency.

All code features are controlled by a parameter file in free text format. Dynamic memory allocation techniques are used throughout, such that no recompilation is necessary, when different problems (collisionless of SPH, vacuum or cosmological background), or varying problem sizes with different numbers of processors, are attacked.

GADGET is an intrinsically Lagrangian code. Both the gravity and the hydrodynamical parts do not impose a restriction on the geometry of the problem, nor any hard limit on the allowable dynamic range. Current and future simulations of structure formation that aim to resolve galaxies in their correct cosmological setting will have to resolve length scales of size $0.1-1\,h^{-1}$kpc in volumes of size $\sim 100\,h^{-1}$Mpc. This range of scales is accompanied by a similarly large dynamic range in mass and time scales. Our new code is essentially free to adapt to these scales naturally, and it invests computational work only where it is needed. It is therefore a good tool to work on these problems.

Since GADGET is written in standard ANSI-C, and the parallelization for massively parallel supercomputers is achieved with the standard MPI library, the code runs on a large variety of platforms, without requiring any change. Having eliminated the dependence on proprietary compiler software and operating systems we thus think that the code will remain usable for the foreseeable future. Since GADGET is highly competitive with other existing codes, we hope that it will become a widely used tool, benefiting continuously from further improvements.

5.9 Appendix: Softened tree nodes

The smoothing kernel we use for SPH calculations is a spline of the form (Monaghan & Lattanzio 1985)

$$W(r;h) = \frac{8}{\pi h^3} \begin{cases} 1 - 6\left(\frac{r}{h}\right)^2 + 6\left(\frac{r}{h}\right)^3, & 0 \le \frac{r}{h} \le \frac{1}{2}, \\ 2\left(1-\frac{r}{h}\right)^3, & \frac{1}{2} < \frac{r}{h} \le 1, \\ 0, & \frac{r}{h} > 1. \end{cases} \tag{5.71}$$

Note that we define the smoothing kernel on the interval $[0,h]$ and not on $[0,2h]$ as it is frequently done in SPH calculations for traditional reasons.

We derive the spline-softened gravitational force from this kernel by taking the force from a point mass m to be the one resulting from a density distribution $\rho(\mathbf{r}) = mW(\mathbf{r};h)$.

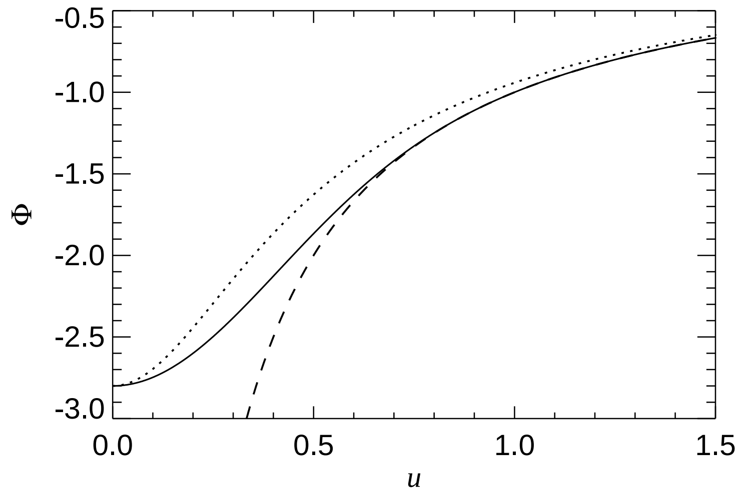


Figure 5.10: Comparison of spline-softened (solid) and Plummer-softened (dotted) potential of a point mass with the Newtonian potential (dashed). Here $h = 1.0$, and $\epsilon = h/2.8$.

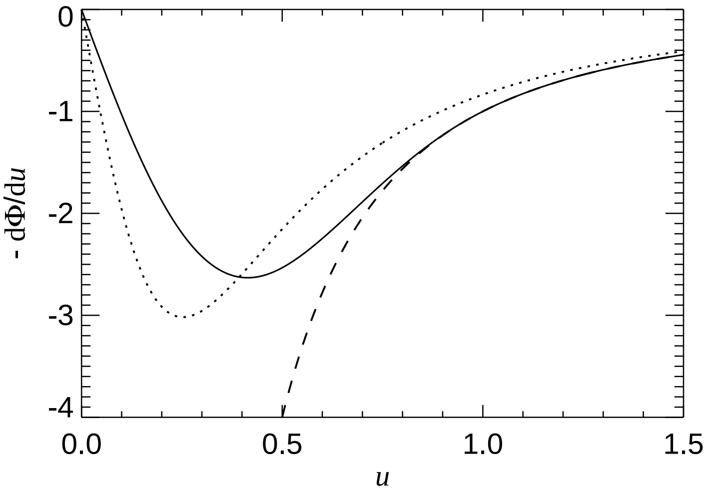


Figure 5.11: Comparison of spline-softened (solid) and Plummer-softened (dotted) force law with Newton's law (dashed). Here $h = 1.0$, and $\epsilon = h/2.8$.

This leads to a potential

$$\Phi(\mathbf{r}) = G\frac{m}{h}W_2\left(\frac{r}{h}\right) \tag{5.72}$$

with a kernel

$$W_2(u) = \begin{cases} \frac{16}{3}u^2 - \frac{48}{5}u^4 + \frac{32}{5}u^5 - \frac{14}{5}, & 0 \le u < \frac{1}{2}, \\ \frac{1}{15u} + \frac{32}{3}u^2 - 16u^3 + \frac{48}{5}u^4 - \frac{32}{15}u^5 - \frac{16}{5}, & \frac{1}{2} \le u < 1, \\ -\frac{1}{u}, & u \ge 1. \end{cases} \tag{5.73}$$

The multipole expansion of a group of particles is discussed in Section 5.3.1. It results in a potential and force given by equations (5.13) and (5.17), respectively. The functions appearing in equation (5.17) are defined as

$$g_1(y) = \frac{g'(y)}{y}, \tag{5.74}$$

$$g_2(y) = \frac{g''(y)}{y^2} - \frac{g'(y)}{y^3}, \tag{5.75}$$

$$g_3(y) = \frac{g_2'(y)}{y}, \tag{5.76}$$

$$g_4(y) = \frac{g_1'(y)}{y}. \tag{5.77}$$

Writing $u = y/h$, the explicit forms of these functions are

$$g_1(y) = \frac{1}{h^3}\begin{cases} -\frac{32}{3} + \frac{192}{5}u^2 - 32u^3, & u \le \frac{1}{2}, \\ \frac{1}{15u^3} - \frac{64}{3} + 48u - \frac{192}{5}u^2 + \frac{32}{3}u^3, & \frac{1}{2} < u < 1, \\ -\frac{1}{u^3}, & u > 1, \end{cases} \tag{5.78}$$

$$g_2(y) = \frac{1}{h^5}\begin{cases} \frac{384}{5} - 96u, & u \le \frac{1}{2}, \\ -\frac{384}{5} - \frac{1}{5u^5} + \frac{48}{u} + 32u, & \frac{1}{2} < u < 1, \\ \frac{3}{u^5}, & u > 1, \end{cases} \tag{5.79}$$

$$g_3(y) = \frac{1}{h^7}\begin{cases} -\frac{96}{u}, & u \le \frac{1}{2}, \\ \frac{32}{u} + \frac{1}{u^7} - \frac{48}{u^3}, & \frac{1}{2} < u < 1, \\ -\frac{15}{u^7}, & u > 1, \end{cases} \tag{5.80}$$

$$g_4(y) = \frac{1}{h^5}\begin{cases} -\frac{96}{5}(5u - 4), & u \le \frac{1}{2}, \\ \frac{48}{u} - \frac{1}{5u^5} - \frac{384}{5} + 32u, & \frac{1}{2} < u < 1, \\ \frac{3}{u^5}, & u > 1. \end{cases} \tag{5.81}$$

In Figures 5.10 and 5.11, we show the spline-softened and Plummer-softened force and potential of a point mass. For a given spline softening length h, we define the 'equivalent' Plummer softening length as $\epsilon = h/2.8$. For this choice, the minimum of the potential at $u = 0$ has the same depth.

–Sie will wissen, was ich für ein Physiker bin. Für sie gibt es zwei Arten von Physikern: Die meisten, das sind die, die nur in Formeln denken, die wenigsten, das sind die, die in Ordnung sind.

Ulrick Woelk, Freigang

6

Genus statistics of the Virgo N-body simulations and the 1.2-Jy redshift survey

Abstract

We study the topology of the Virgo N-body simulations and compare it to the 1.2-Jy redshift survey of *IRAS* galaxies by means of the genus statistic. Four high-resolution simulations of variants of the CDM cosmology are considered: a flat standard model (SCDM), a variant of it with more large-scale power (τCDM), and two low density universes, one open (OCDM, $\Omega_0 = 0.3$) and one flat (ΛCDM, $\Omega_0 = 0.3$, fluctuation amplitudes are chosen so that the simulations approximately reproduce the observed abundance of rich clusters of galaxies at the present day. The fully sampled N-body simulations are examined down to strongly nonlinear scales, both with spatially fixed smoothing, and with an adaptive smoothing technique. While the τCDM, ΛCDM, and OCDM simulations have very similar genus statistics in the regime accessible to fixed smoothing, they can be separated with adaptive smoothing at small mass scales. In order to compare the N-body models with the 1.2-Jy survey, we extract large ensembles of mock catalogues from the simulations. These mock surveys are used to test for various systematic effects in the genus analysis and to establish the distribution of errors of the genus curve. We find that a simple multivariate analysis of the genus measurements is compromised both by non-Gaussian distributed errors and by noise that dominates the covariance matrix. We therefore introduce a principal components analysis of the genus curve. With a likelihood ratio test we find that the 1.2-Jy data favours the ΛCDM, τCDM and OCDM models compared to SCDM. When genus measurements for different smoothing scales are combined, the SCDM model can be excluded at a 99 per cent confidence level, while the other three models fit the 1.2-Jy data well. These results are unlikely to be significantly modified if galaxies are biased tracers of the mass, provided that biasing preserves a monotonic relation between galaxy density and mass density.

6.1 Introduction

The observed large-scale structure of the Universe represents one of the most important constraints for theories of cosmic structure formation. In the past, the clustering of galaxies was mainly studied with statistics like the two-point correlation function, the power spectrum, or the counts-in-cells analysis. These statistical measures have been routinely applied to more and more powerful redshift surveys, leading to significant advances in our understanding of cosmic history. Cosmological N-body simulations have played a vital part in this development, and they are already strongly constrained by the available data.

However, the two-point correlation function and the power spectrum describe the properties of the galaxy distribution only to a limited extent. A full description would involve a hierarchy of three-, four- and higher correlation functions or, alternatively, information on the phase correlations among the different Fourier modes of the density field. Furthermore, the low-order statistics are also quite insensitive to the geometrical aspects of the clustering which the human eye can so easily detect in pictures of the matter distribution. For example, it is still unclear whether the galaxy distribution is best described as filamentary, cellular or sheet-like.

In order to develop a measure of the geometrical and morphological aspects of the galaxy distribution a number of measures have been proposed, among them percolation (Klypin 1988), level-crossing statistics (Ryden 1988), genus statistics (Gott et al. 1986), minimal spanning tree (Pearson & Coles 1995), shape statistic (Luo & Vishniac 1995), and Minkowski functionals (Mecke et al. 1994).

In this chapter we focus on the genus statistic which was first proposed by Gott et al. (1986) and has become a widely accepted statistical tool in cosmology since then. The genus probes the topology of isodensity surfaces of a smoothed mass density field. It is therefore sensitive to global aspects of the density field that manifest themselves in higher order correlations, which reflect the connectedness and morphology of the structure in the Universe. Such features are not revealed by standard measures like the power spectrum or the two-point correlation function. For example, galaxy distributions that are wall-like, bubble-like or filamentary would all lead to different signatures in the genus statistic. Thus the topology has potentially strong discriminative power and might be used to rule out or support certain models for structure formation.

A particularly interesting application of the genus statistic is a test of the random phase hypothesis for the initial density fluctuation field. Because the topology is invariant during the linear growth of structure, the topology of the present galaxy distribution can be related to the topology of the initial density field, which in turn allows a test of the random phase hypothesis. Any departure of the measured topology from the random phase prediction would be evidence for the presence of phase correlations that might reflect non-Gaussian initial conditions.

In contrast to the linear regime, the topology on small, strongly non-linear scales has hardly been studied to date. In particular, there are no theoretical predictions for the genus statistics in this regime. Hence it is presently unclear whether the genus on small scales is a useful statistic to discriminate between different models of structure formation.

In this chapter we examine the topology of a suite of large, high-resolution N-body simulations carried out by the Virgo collaboration (Jenkins et al. 1998). The suite of models consists of a standard cold dark matter model (SCDM), and three variants of the cold dark matter cosmology, which have more power on large scales. These are a flat $\Omega_0 = 1$ model (τCDM), and two low density universes with $\Omega_0 = 0.3$, one open model (OCDM), and one closed by means of a cosmological constant (ΛCDM).

In the first of a series of papers, Jenkins et al. (1998) measured the correlation function, the power spectrum and various statistics of the velocity field of these simulations. As Jenkins et al. (1998) demonstrate, all three models with the power spectrum shape $\Gamma = 0.21$ can fit the APM two-point galaxy correlation function reasonably well, if one allows for a moderate scale-dependent bias. However, the differences *between* these models are rather small, at least when only the distribution at $z = 0$ is examined with these statistics.

In the second paper of the series, Thomas et al. (1997) showed that the internal structure of halos of rich clusters is also similar in all the models and would be difficult to distinguish in practice.

Here we use the genus statistic to analyse these N-body simulations down to strongly non-linear scales. In this regime, differences between the models can be detected. In a second thread we compare the genus of the N-body models to the 1.2-Jy redshift survey of *IRAS* galaxies. Our particular aim is to improve the statistical methodology of such a comparison. For this purpose we work with ensembles of mock catalogues to assess the properties of new smoothing techniques and to derive accurate estimates for errors and systematic effects. This Monte-Carlo technique also allows the derivation of formal exclusion levels for the N-body models. New larger redshift surveys may be used subsequently to further tighten these constraints.

This chapter is organized as follows. In Section 2 we briefly review the genus statistic which we apply in Section 3 to the fully sampled Virgo N-body simulations, both with fixed and adaptive smoothing. Section 4 describes the 1.2-Jy redshift survey and the construction of ensembles of mock catalogues. In Section 5 we introduce different methods to compute smoothed density fields from the redshift survey data, and in Section 6 we discuss various systematic effects that affect the genus statistic. We then turn in Section 7 to the statistical methodology we adopt for the comparison with the 1.2-Jy redshift survey, and we present the results of this comparison in Section 8. Finally we summarize and conclude in Section 9.

6.2 Genus statistic

We first review briefly the genus statistic as introduced by Gott et al. (1986). A number of redshift surveys have been analysed with it, starting with a compilation of relatively small samples examined by Gott et al. (1989). Further surveys that have been studied include the SSRS (Park et al. 1992a), QDOT (Monaghan 1992), Abell Clusters (Rhoads et al. 1994), CfA (Vogeley et al. 1994), and most recently the 1.2-Jy survey (Protogeros & Weinberg 1997) and PSCz (Canavezes et al. 1998). There have also been a number of theoretical studies of the genus in the mildly non-linear regime (Magorrian & Binney 1994; Matsubara & Yokoyama 1996; Matsubara 1996; Matsubara & Suto 1996), and the genus has been applied in two dimensions to the microwave background (Colley et al. 1996) and to slice surveys (Park et al. 1992b; Colley 1997).

Given an isodensity contour of a smoothed mass density field we define the genus as

$$G = -\frac{1}{4\pi} \int \kappa \, \mathrm{d}A, \tag{6.1}$$

where

$$\kappa = \frac{1}{r_1 r_2} \tag{6.2}$$

is the local Gaussian curvature. Here r_1 and r_2 denote the principal radii of curvature, and the integration extends over the whole surface. The Gauss-Bonnet theorem shows that this definition of genus makes G equal to the number of topological holes (like the one in a doughnut) minus the number of isolated regions of the surface.

Assuming an ergodic universe, we define a genus

$$g = \frac{G}{V} \tag{6.3}$$

per unit volume, where V is finite, but large enough to be a representative patch of the universe. The genus depends on the density threshold used to construct the isodensity surface. As a function of threshold we therefore obtain a *genus curve*, which is the central object of this investigation.

We parameterize the genus curve by the fraction f of the volume above the density threshold value or by the quantity

$$\nu = \sqrt{2}\, \mathrm{erf}^{-1}(1 - 2f) \tag{6.4}$$

derived from it. Here erf^{-1} denotes the inverse of the error function $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \mathrm{e}^{-t^2} \mathrm{d}t$. We will stick to the usual convention and present genus curves in the form $g = g(\nu)$.

The definition (6.4) is chosen such that for a Gaussian random field the quantity $\nu = \delta_\mathrm{t}/\sigma$ just measures the threshold value δ_t in units of the dispersion σ. However, we always define ν in terms of the volume fraction via equation (6.4), because this definition has the advantage of making the genus curve invariant under arbitrary monotonic one-to-one transformations of the density field. For example, a simple linear biasing transformation would not affect it. Also, it is insensitive to the skewness of the one-point probability distribution function, that quickly develops in the mildly non-linear regime.

There is a theoretical prediction for the expected genus curve of a Gaussian random field (Hamilton et al. 1986). It is given by

$$g(\nu) = N\,(1 - \nu^2) \exp\left(-\frac{\nu^2}{2}\right), \tag{6.5}$$

where the amplitude

$$N = \frac{1}{(2\pi)^2}\left(\frac{\langle k^2\rangle}{3}\right)^{\frac{3}{2}} \tag{6.6}$$

is determined by the second moment

$$\left\langle k^2 \right\rangle = \frac{\int k^2 P(k)\mathrm{d}^3k}{\int P(k)\mathrm{d}^3k} \tag{6.7}$$

of the (smoothed) power spectrum. Interestingly, only the amplitude of the genus curve depends on the shape of the power spectrum. Apart from that it exhibits a universal, symmetric w-shape that we will use as a benchmark to detect non-Gaussian features of the density field.

For a given threshold value we compute the genus with the algorithm proposed by Gott et al. (1986). The method tesselates space in small cubes that allow the isodensity surface to be defined as the boundary between the volume elements above and below threshold. If the cubes are sufficiently small this approximation does not change the topology of the smooth isodensity surface. The curvature of the resulting polygonal surface is compressed into the vertices of the cubes. This property allows a computer to rapidly sum up the appropriate angle deficits and to compute the genus per unit volume. The method also allows arbitrarily shaped survey volumes. Here one just counts those vertices that are surrounded by eight volume elements that all lie inside the actual survey region.

Based on Weinberg's (1988) code CONTOUR for computing the genus we have written a new version in C that is optimized for a high execution speed, since we need to compute several thousand genus curves in this work. A simple sorting of the density field prior to the genus computation led already to a major speed-up because it is then possible to instantly find the threshold value corresponding to a desired volume fraction. This also allows the efficient computation of high resolution genus curves.

6.3 Fully sampled Virgo simulations

6.3.1 The N-body models

In this Section we compute genus curves for four fully sampled N-body simulations of variants of the CDM cosmology. These simulations, part of the Virgo project (Jenkins et al. 1998) to examine structure formation in the universe at very high resolution, follow 256^3 dark matter particles in boxes of comoving size of $239.5\,h^{-1}$kpc. The simulations have been carried out with the parallel adaptive grid code HYDRA (Pearce & Couchman 1997).

The linear theory power spectrum used to generate the initial conditions of the models is parameterized by the generic fitting form

$$P(k) = \frac{Bk}{\left(1 + [ak + (bk)^{3/2} + (ck)^2]^\nu\right)^{2/\nu}}, \tag{6.8}$$

Table 6.1: Parameters of the examined CDM models. The simulations have been done by the Virgo collaboration.

	SCDM	τCDM	ΛCDM	OCDM
Number of particles	256^3	256^3	256^3	256^3
Box size[h^{-1}kpc]	239.5	239.5	239.5	239.5
z_{start}	50	50	30	119
Ω_0	1.0	1.0	0.3	0.3
Ω_Λ	0.0	0.0	0.7	0.0
Hubble constant h	0.5	0.5	0.7	0.7
Γ	0.5	0.21	0.21	0.21
σ_8	0.60	0.60	0.90	0.85

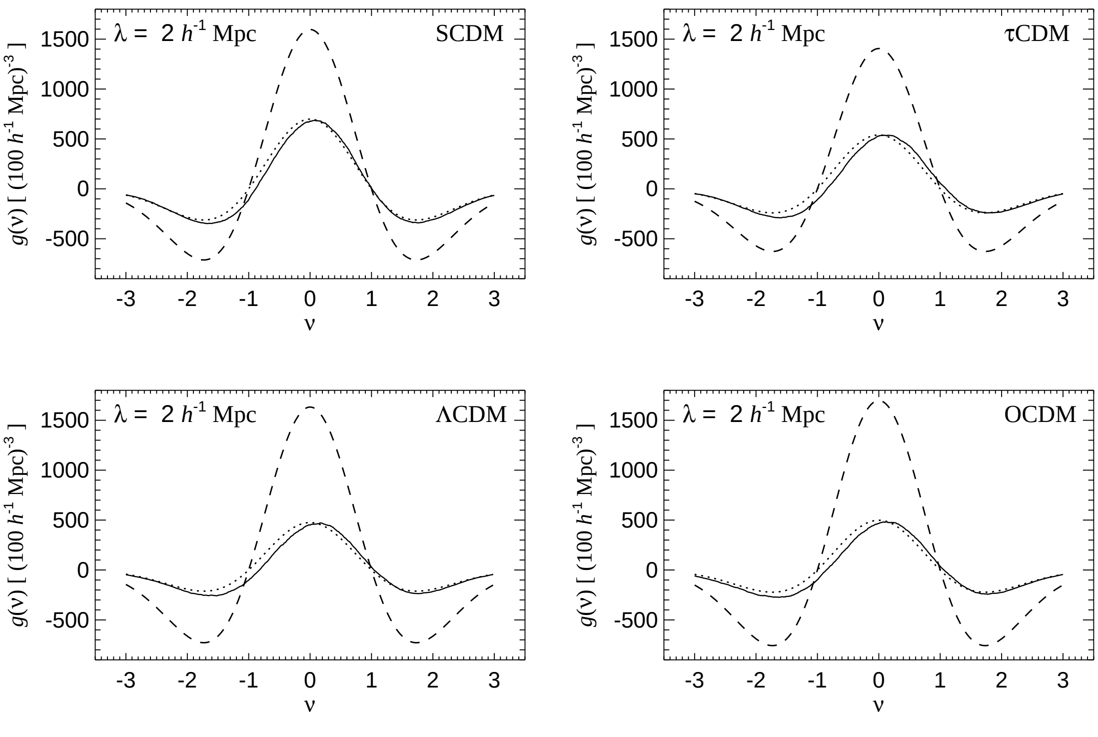


Figure 6.1: Genus curves of four variants of the CDM cosmology. The solid lines in each panel show the genus of the evolved density fields of the Virgo simulations for a smoothing scale of $2\,h^{-1}$kpc. The dotted line is a fit to the random phase genus curve, while the dashed curve gives the genus of the corresponding *Gaussianized* field.

where $a = 6.4\,\Gamma^{-1}\,h^{-1}$kpc, $b = 3.0\,\Gamma^{-1}\,h^{-1}$kpc, $c = 1.7\,\Gamma^{-1}\,h^{-1}$kpc, $\nu = 1.13$, and Γ is a shape parameter (Efstathiou et al. 1992).

While the SCDM model has the shape parameter $\Gamma = 0.5$, the other three models have the same linear power spectrum with $\Gamma = 0.21$. All four models are normalized so as to give the observed abundance of rich clusters of galaxies at the present day. Further simulation parameters are listed in Table 6.1 and may be found in Jenkins et al. (1998).

6.3.2 Fixed smoothing

We start with the ordinary genus statistic, i.e. we assume a spatially constant smoothing kernel. In order to construct a smooth density field we bin the 16.7 million particles of one simulation onto a mesh using the cloud-in-cell assignment, and we smooth the resulting density field with a Fast Fourier convolution. We use a Gaussian kernel of the form

$$W(\mathbf{x}) = \frac{1}{\pi^{3/2}\lambda^3}\,\exp\left(-\frac{\mathbf{x}^2}{\lambda^2}\right). \tag{6.9}$$

Note that this differs from an ordinary normal distribution by a factor of $\sqrt{2}$ in the definition of the smoothing scale. We stick to this convention which is used in the majority of the literature on the subject.

Typically we employ a 128^3 grid to represent the density field. Only for smoothing lengths below $5\,h^{-1}$kpc we do find that a smaller mesh is indicated. We then use a 256^3 grid. For test purposes, we repeated one of our calculations on a 512^3 mesh.

6.3.3 Results

In Figure 6.1 we show the genus curves of the four simulations at a smoothing scale of $2\,h^{-1}$kpc, the smallest scale considered here. We will come back to the question of what happens on still smaller ones later on. It is evident that the genus curves retain their universal w-shape, even at this small smoothing scale where the density fields are already fairly non-linear. Only a small *bubble* shift to the right has developed, which seems somewhat weaker for the SCDM model than for the other cosmologies. Such a bubble shift has been reported by Melott et al. (1988) as well, and it was also found by Vogeley et al. (1994) for the CfA survey.

We obtain qualitatively similar genus curves for larger smoothing scales. For $\lambda > 8\,h^{-1}$kpc the small bubble shift has vanished for all models and the shape of the genus curves is fit perfectly by the random phase form of equation (6.5). In Figure 6.2 we show the measured genus amplitudes as a function of smoothing scale. The SCDM model exhibits a significantly higher amplitude, reflecting the different shape of its power spectrum, while the other three models show very similar amplitudes. This demonstrates that the genus amplitude in the linear and mildly non-linear regime is determined by the shape of the power spectrum alone.

Even if the genus curve is well described by equation (6.5), the underlying density field does not have to be Gaussian. In fact, Vogeley et al. (1994) and Canavezes et al.

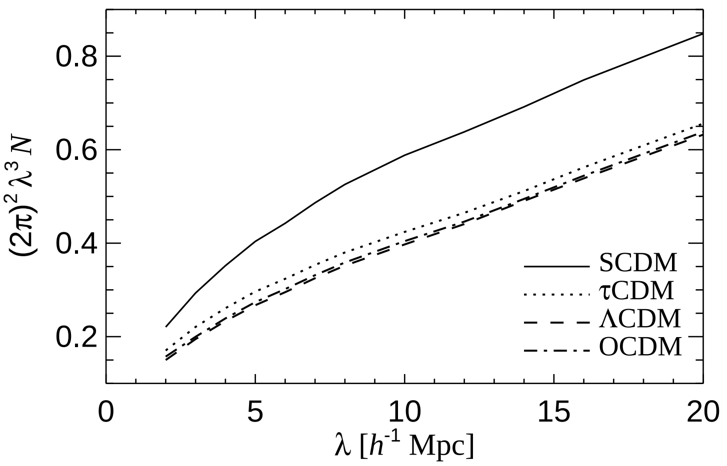


Figure 6.2: Genus amplitude of the Virgo simulations as a function of smoothing scale (fixed smoothing). The dimensionless vertical axis gives the genus amplitude N times the factor $(2\pi)^2\lambda^3$.

(1998) pointed out that the amplitude of the genus curve is suppressed on small scales compared to the expected amplitude based on the power spectrum alone. This amplitude drop was first observed by Melott et al. (1988) and is a direct consequence of phase correlations that develop during the nonlinear growth of density perturbations. The phase drop may be measured for a periodic N-body simulation by *Gaussianizing* the evolved density field, i.e. by taking it to Fourier space, randomizing the phases of all modes constrained by the reality condition $\delta_{\mathbf{k}} = -\delta^{\star}_{\mathbf{k}}$, and transforming back to real space. In this way a Gaussian field with identical power spectrum as the evolved density field is obtained, and a measurement of its genus allows an estimate of the amplitude drop.

We have measured the genus amplitude drop in this way and show the results in Figure 6.3. Here we reach smoothing scales as small as $2\,h^{-1}$kpc, thereby extending the work of Canavezes et al. (1998). Note that the amplitude drop becomes very substantial at small scales, showing that the genus is indeed quite sensitive to higher order correlations in this regime. If we plot the amplitude drop against the variance of the smoothed density fields, we can approximately take out the differences between the models due to their slightly different normalizations. As the lower panel in Figure 6.3 shows, we again find that the models with the same shape of the power spectrum exhibit a nearly degenerate behaviour. However, the horizontal axis of this plot is affected by biasing. Since the bias required to match the observed two-point correlation function (Jenkins et al. 1998) is different for the four models, the amplitude drop may still be a useful measure to discriminate between different CDM variants. Future galaxy redshift sur-

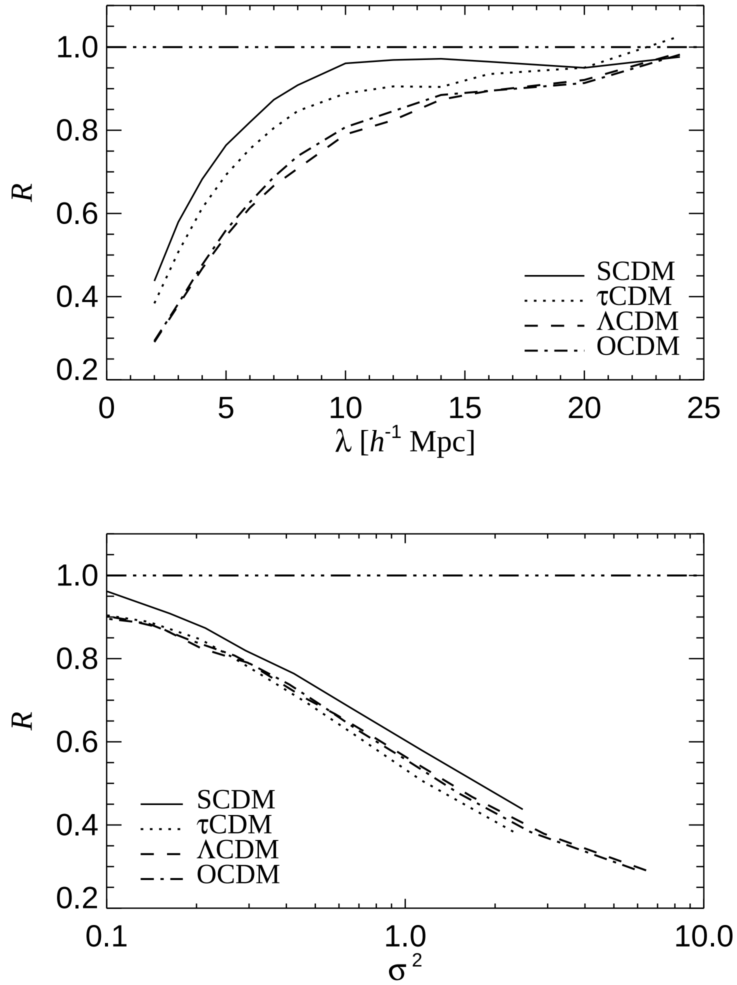


Figure 6.3: Amplitude drop for the Virgo simulations, i.e. the ratio R of the genus amplitude N to the corresponding amplitude of the Gaussianized density field. The top panel shows the amplitude drop versus the smoothing scale, while the lower panel displays it against the variance of the smoothed fields.

veys should allow an accurate measurement of the amplitude drop by combining genus statistics with an independent measure of the power spectrum or clustering strength as outlined by Canavezes et al. (1998). Comparing these measurements with the results obtained in Figure 6.3 for the dark matter may then allow a direct measure of the bias $b^2 = \sigma^2_{\mathrm{gal}}/\sigma^2_{\mathrm{DM}}$.

We now examine systematic limitations of the genus analysis performed above. We will consider finite grid size effects first and then examine the resolution limit of the

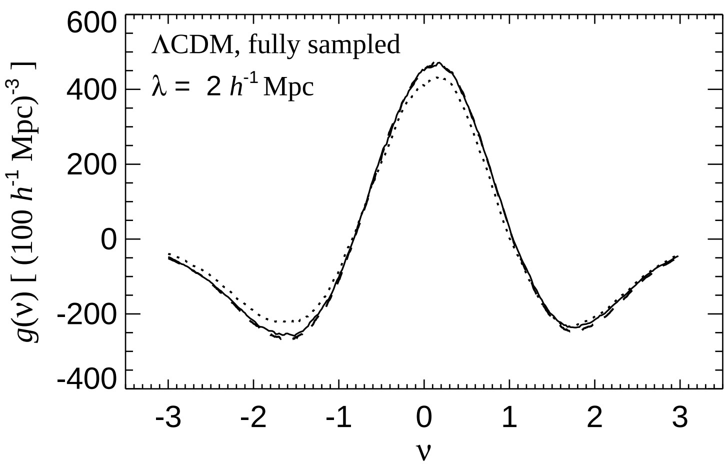


Figure 6.4: Finite grid size effect. The lines are genus curves for the ΛCDM simulation, smoothed with $\lambda = 2\,h^{-1}$kpc. The solid line is computed with a 512^3 grid, the dashed line with 256^3, and the dotted line with 128^3.

Virgo simulations in terms of the genus statistic.

6.3.3.1 Finite grid size

Our method for computing the genus curve relies on an approximate representation of isodensity contours as polygonal surfaces that are made up of faces of small cubes used to tesselate space. As Hamilton et al. (1986) show the genus curve is expected to be unaffected by this approximation, if $d/\lambda \ll 1$, where d is the size of the cubes. Typically we use 128^3-grids, and for $\lambda \leq 5\,h^{-1}$kpc 256^3 grids. We find that for $d \leq 0.5\lambda$ there is hardly any finite grid size effect. For example, in Figure 6.4 we show a comparison of the genus curves for the ΛCDM simulation, smoothed at $\lambda = 2\,h^{-1}$kpc, using 128^3, 256^3, and 512^3 grids. There is only a small depression of the amplitude and the minima of the genus curve, when the 128^3 mesh is used, but for 256^3 the asymptotic behaviour is clearly reached. It is therefore not necessary to use costly computations at 512^3 resolution.

6.3.3.2 Resolution limit

If the smoothing length is reduced below $2\,h^{-1}$kpc one suddenly starts to see features in the genus curve, as exemplified in Figure 6.5, where we show the genus for the ΛCDM simulation at a smoothing scale of $1\,h^{-1}$kpc. However, the ‘ringing’ on the low density side is just an artifact due to the fact that the resolution of the Virgo simulations is limited. We demonstrate that this is indeed the case by analysing yet another simulation of the Virgo consortium, the same ΛCDM model, but with a smaller box size of

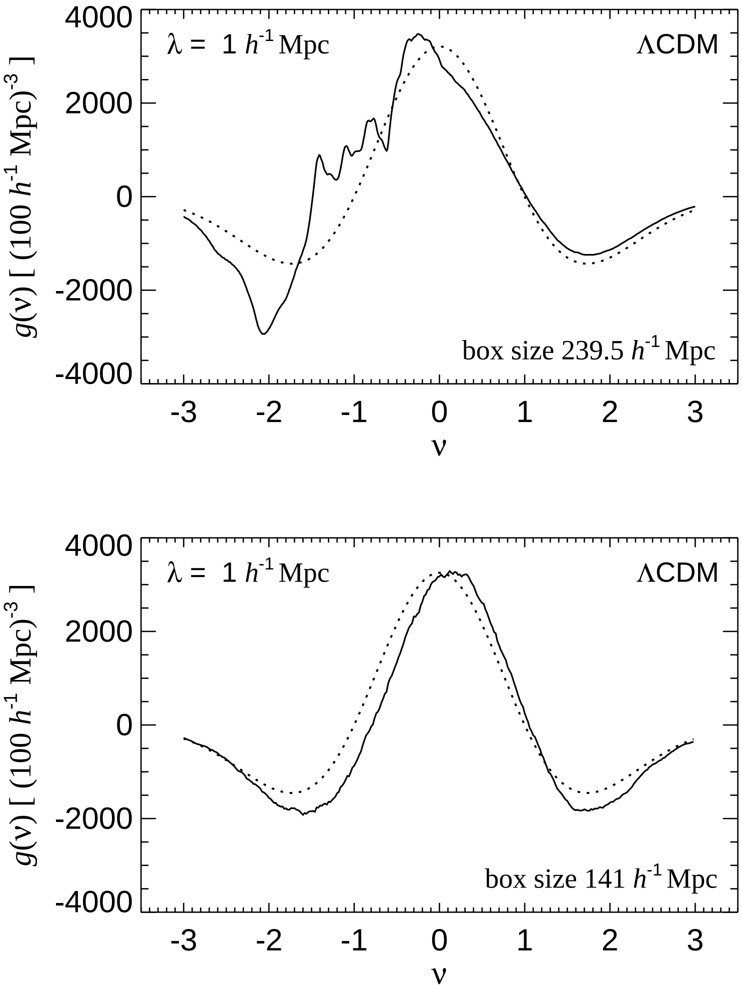


Figure 6.5: Resolution limit of the Virgo simulations. The solid line of the top panel shows the genus of the ΛCDM model smoothed with $\lambda = 1\,h^{-1}$kpc, and the dotted curve shows a best-fit random phase genus curve. In this case the smoothing reaches the level of the inter-particle separation and the features in the genus curve are in fact artifacts. We show that this is the case by computing the genus for a second ΛCDM simulation with smaller box size of $141\,h^{-1}$kpc, but with the same number of particles and hence significantly higher spatial resolution. As seen in the bottom panel, the genus curve remains featureless.

$141\,h^{-1}$kpc. This model has roughly twice the spatial resolution of the original simulation. As the bottom panel of Figure 6.5 demonstrates, it still gives a 'normal' genus curve for $\lambda = 1\,h^{-1}$kpc.

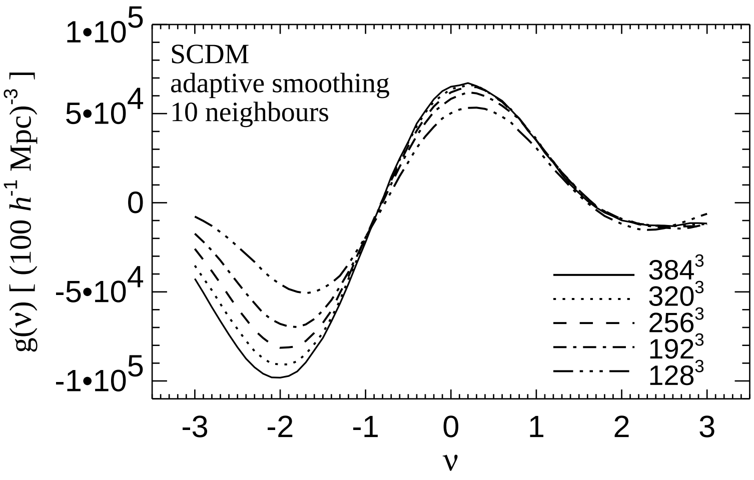


Figure 6.6: Finite grid size effects in the adaptive smoothing technique. Shown are five genus curves for one of the SCDM subvolumes, computed at different grid resolutions. For 10 neighbours, the number considered here, a 256^3 mesh is sufficient to resolve accurately the genus of the density field, at least in the regions of positive genus and for $\nu > 0$. However, the minimum on the negative side is still not fully resolved by a grid as fine as 384^3. Comparing the size distribution of the isolated regions at the two minima of the genus curve, we find that this is due to a larger relative abundance of very small regions at the minimum on the negative side. This population of very small structures is difficult to resolve; those regions with volume smaller than a mesh cell can be lost, leading to a suppression of the genus amplitude.

The smoothing scale of $1\,h^{-1}$kpc is already close to the mean inter-particle separation of $\approx 0.94\,h^{-1}$kpc for the large-box Virgo runs. Due to the strong clustering of matter, the voids apparently contain too few particles to prevent discreteness effects becoming visible when a fixed smoothing kernel with $\lambda = 1\,h^{-1}$kpc is used. On the other hand, the regions with high particle density allow a much higher resolution in principle. In order to take full advantage of this spatially varying resolution an adaptive smoothing technique needs to be developed. We will introduce such a scheme in the next Section.

6.3.4 Adaptive Smoothing

As we saw above, a spatially fixed smoothing is not able to take full advantage of the information content on small scales. Borrowing an idea from smoothed particle hydrodynamics (SPH), we can improve on this by following the particles in a Lagrangian sense and varying the smoothing scale with local density. This application of SPH-smoothing to the dark matter has first been proposed by Thomas & Couchman (1992).

Of course, a difficulty with such an approach is that the resulting density field can-

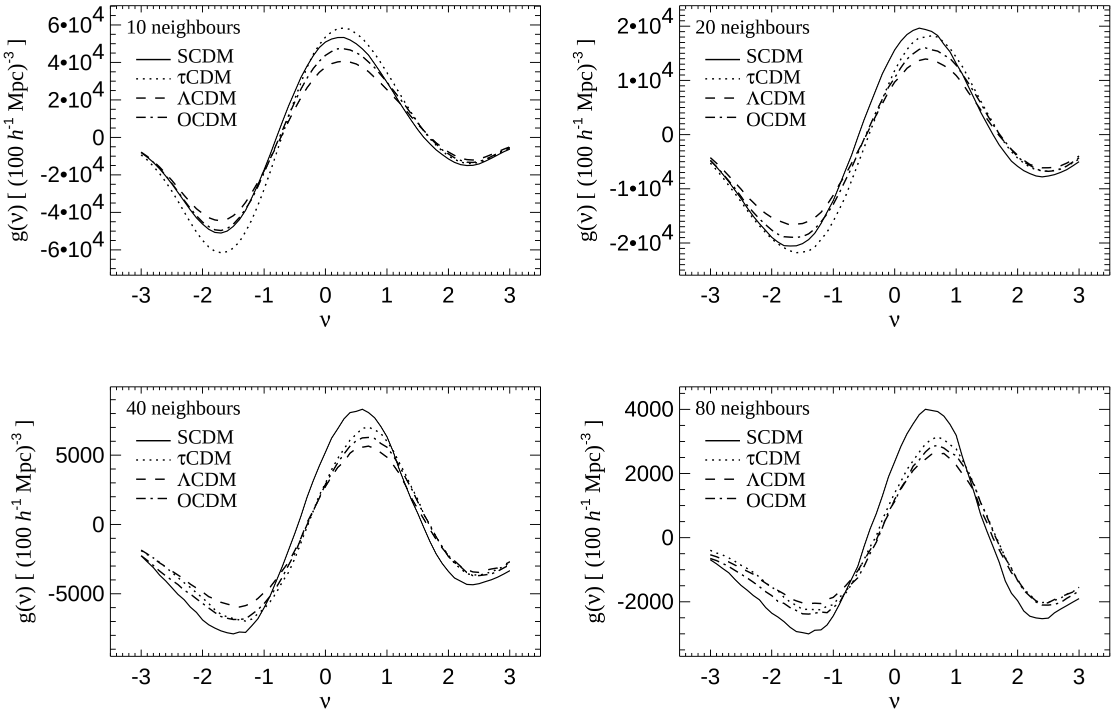


Figure 6.7: Genus curves for the adaptively smoothed Virgo simulations. The four panels show results when 10, 20, 40, and 80 neighbours are used to determine the local smoothing scale. Each curve is the average of 4 subvolumes, each being 64 times smaller than the full simulation volume. Since the variance between these curves is quite small, it is not necessary to extend the computation to a larger fraction of the total volume.

not be studied analytically. In particular, the power spectrum of the smoothed field is not related to that of the underlying field in a simple way. Because of that, it is also not obvious what one can expect for the genus statistic. In particular, adaptive smoothing will not be useful to test the Gaussian random phase hypothesis on linear scales. However, in light of the increased resolution of current redshift surveys and of N-body simulations, the study of the topology at non-linear scales has become possible. In this regime, there are no reliable analytical predictions for the genus even when fixed smoothing is used. Hence a comparison between theory and observation is perhaps best done by using ensembles of mock surveys to calibrate the genus statistic. In this framework one might then as well try an adaptive smoothing scheme in an attempt to maximize the topological information extracted from a given data set.

A potential disadvantage of adaptive smoothing is that it may spoil to some extent the relative independence of the ordinary genus analysis on bias. Further studies will be needed to clarify how much this diminishes the usefulness of adaptive smoothing for a comparison of redshift surveys with simulations. However, the study of the topology of the dark matter at highly non-linear scales is in itself very interesting, and it requires adaptive smoothing to take full advantage of the mass resolution of the simulations.

As Hernquist & Katz (1989) point out there are two different approaches to defining a smoothed density estimate with variable smoothing scale. In the *scatter* approach the mass of each particle is distributed in space, and the density estimate at a particular point $\mathbf{x}$ follows from the overlap of the individual smoothing spheres. Alternatively, one can define a smoothing radius for each point $\mathbf{x}$ and weight all particles in its neighbourhood by the resulting kernel (*gather* approach).

We want to compute the density field on a fine mesh at a large number of points. Because the number of tracer particles is also large, the gather approach is computationally less costly in this case. Hence we will adopt it in the following.

As a smoothing kernel we choose the spherically symmetric spline kernel

$$W(r;h) = \frac{8}{\pi h^3} \begin{cases} 1 - 6\left(\frac{r}{h}\right)^2 + 6\left(\frac{r}{h}\right)^3, & 0 \le \frac{r}{h} \le \frac{1}{2}, \\ 2\left(1 - \frac{r}{h}\right)^3, & \frac{1}{2} < \frac{r}{h} \le 1, \\ 0, & \frac{r}{h} > 1, \end{cases} \tag{6.10}$$

which is frequently used in SPH calculations (Monaghan & Lattanzio 1985). This kernel has the advantage of compact support, which again simplifies the computational task.

For simplicity, we choose the smoothing scale $h(\mathbf{x})$ as the distance to the N-th nearest neighbour. In this way the smoothing scale is set to a fixed fraction of an estimate of the local mean inter-particle separation. We then estimate the density field as

$$\rho(\mathbf{x}) = \sum_i W(\mathbf{r}_i - \mathbf{x}; h(\mathbf{x})), \tag{6.11}$$

where the sum extends over all particles.

The computational cost for adaptive smoothing is much higher than for fixed smoothing, because Fourier techniques can no longer be applied and one has to work in real space. Furthermore, the adaptive smoothing will also be able to resolve very small structures that demand a fine mesh in order to allow a proper resolution of their topology.

Because of this it is not possible to consider the whole simulation box in one pass. Instead, we compute adaptively smoothed density fields for subvolumes of the simulation box of size one quarter of the total box size. In this way, we divide the simulation in 64 subvolumes, each containing still 64^3 particles on average. For the small scales we are considering here, we anticipate that each of these subvolumes will already contain so many structural elements that the cosmic variance between them will be small. Hence we assume that we will have to consider at most a few of the subvolumes in order to accurately reproduce the genus of the whole simulation box.

We compute adaptively smoothed density fields on a regular grid for each of the subvolumes. In the smoothing process, we also consider the particles that lie outside the subvolume, hence boundary smoothing effects are not present. In Figure 6.6 we show the dependence of the result on the chosen grid resolution for the extreme case

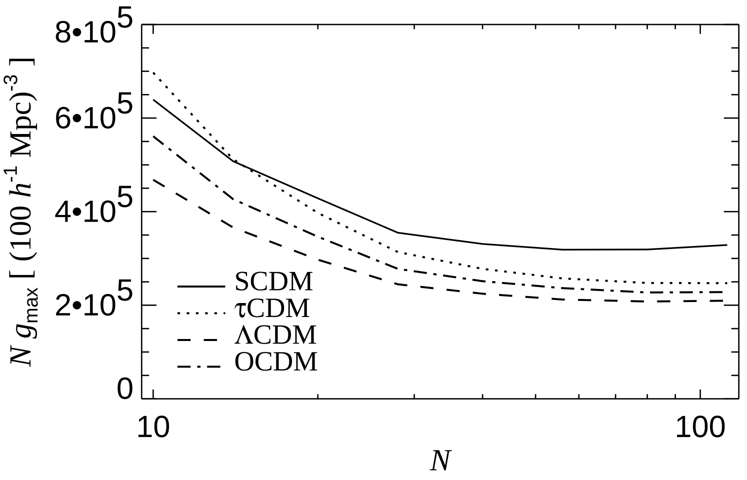


Figure 6.8: Peak genus densities in the Virgo models as a function of the number N of neighbours used in the adaptive smoothing scheme. We have measured the genus for $N = 10$, 14, 20, 28, 40, 56, 80, and 112 neighbours, and plot the maximum genus density times the number N of smoothing neighbours.

of $N = 10$ neighbours. A 128^3 grid clearly leads to a systematic underestimate of the genus, showing that small features are missed due to the coarseness of the grid. On the other hand, the result for the 256^3 grid is already very close to the finer meshes, at least for $\nu \geq -1$. However, the minimum of the genus on the negative side is seen to be very difficult to resolve accurately. By comparing the distributions of the sizes of the isolated regions at the two minima, we find that the number density of very small regions is much larger at the minimum on the negative side than on the positive side. Since these very small voids are so abundant, the loss of the smallest of them due to the finite grid size leads to a suppression of the genus. Fortunately this effect is less severe at larger smoothing scales. As a compromise between computational cost and accuracy, we decided to use 256^3 grids for $N = 10$ neighbours, 192^3 for $14 \leq N \leq 20$, and 128^3 for $N \geq 28$. This ensures that the genus density is accurately resolved for $\nu \geq -1$. However, the depth of the minimum on the negative side may be slightly diminished due to finite grid size effects.

In Figure 6.7 we show genus curves for adaptive smoothing with $N = 10$, 20, 40, and 80 neighbours. In each panel we plot the results for SCDM, τCDM, ΛCDM, and OCDM. Interestingly, there are marked differences between the models, particularly at small smoothing scales. The SCDM model shows usually the highest genus amplitude, as expected for the larger amount of small-scale power in this model. Only on the smallest scale considered here, it is surpassed by the τCDM model. While the genus amplitudes of the three models τCDM, ΛCDM, and OCDM were practically degenerate in the regime

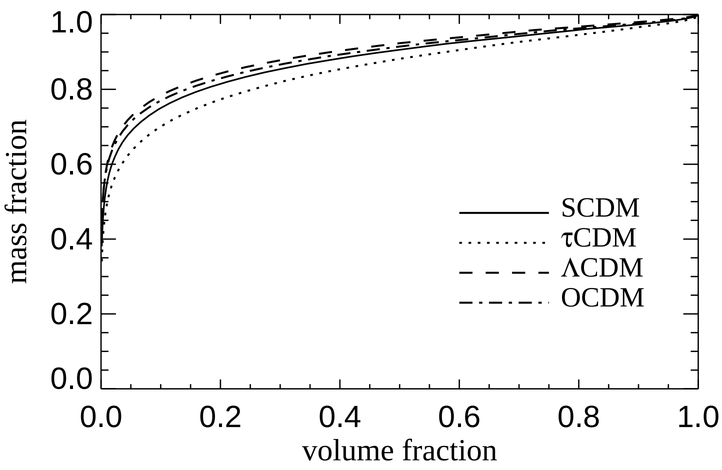


Figure 6.9: Mass fraction above threshold versus volume fraction above threshold for the adaptively smoothed Virgo simulations. Here $N = 10$ neighbours have been used in the smoothing process. At a given volume fraction, the ΛCDM model contains a higher fraction of its mass in the high density regions than the other three models.

accessible to fixed smoothing, they can now be used to discriminate between the models.

Note that the number density of structure elements resolved with adaptive smoothing is really much larger than the one accessible to fixed smoothing. Compared to the $\lambda = 2\,h^{-1}$kpc fixed smoothing, the $N = 10$ adaptive scheme reaches a density of structure elements which is approximately two orders of magnitude larger. This really opens up a new regime for topological analysis.

In Figure 6.8 we plot the peak genus densities as a function of the number of neighbour particles used for the adaptive smoothing. It is interesting that the relative difference between τCDM, ΛCDM, and OCDM grows with decreasing smoothing length. ΛCDM consistently shows the smallest amplitude, which demonstrates that its structure has a higher degree of coherence with fewer small-scale features than the other models.

However, for the $N = 10$ smoothing, all four simulations give roughly the same genus signal at the minimum on the positive side of the genus curves. This minimum occurs at $\nu \simeq 2.4$, corresponding to a volume fraction above threshold of only 0.0082. For a Gaussian random field, the minimum occurs at $\nu = \sqrt{3}$, equivalent to the much larger volume fraction of 0.042. At the minimum the genus is completely dominated by the number density of very dense, isolated clumps. While this number density of roughly $0.012\,h^3\mathrm{Mpc}^{-3}$ is similar for the four models, the mass fraction contained in the clumps is different. For the $N = 10$ smoothing, we plot in Figure 6.9 the mass fraction above threshold as a function of the volume fraction above threshold. Note the very high concentration of the mass in these adaptively smoothed density fields.

In the SCDM model 50 per cent of the mass is contained in only 0.65 per cent of the volume. The τCDM model requires almost twice the volume fraction, 1.23 per cent, to include half of its mass. On the other hand, the ΛCDM and OCDM simulations have half-mass volume fractions of 0.33 and 0.31 per cent, respectively. At high densities a given volume fraction therefore contains a higher fraction of their mass than in the high density models. This reflects the normalization of the models which ensures that they all have the same abundance of virialized clusters of any given mass.

In contrast to the minimum on the positive side, the minimum on the negative side of the genus curve is not shifted significantly. However, here the models show stronger differences in their genus signal. Again, at the minimum the genus is dominated by a high number density of isolated regions. These voids are significantly more abundant in the τCDM model than in the ΛCDM model. This means that the underdense regions are choppier in the τCDM model than in ΛCDM. The latter model has voids which are more coherent and larger on the average. Presently it is unclear, whether this can be understood merely as a consequence of the smaller amount of mass left in the ΛCDM model to fill the voids.

The maxima of the genus densities occur at a smaller volume fraction than 0.5, i.e. the adaptive smoothing results in a substantial bubble-shift to the right. At the maxima, the isodensity surfaces have the topology of a sponge, with an interlocking high density region and a complex system of tunnels and voids. The smaller amplitude of ΛCDM can again be interpreted as a larger degree of coherence; there are not so many topological holes as in τCDM, for example, and the typical size of tunnels and cavities is expected to be larger.

It should be noted that the two low-Ω_0 models show only small differences when their two-point correlation functions or their velocity fields are considered (Jenkins et al. 1998). This demonstrates that the genus can indeed reveal additional information about the morphological properties of the matter distribution. Here we conclude that the voids in the ΛCDM model are in some sense 'emptier' than in the OCDM model.

We also note that only the adaptive smoothing allowed an extension of the genus statistics essentially down to the mass resolution of the Virgo simulations. Future large redshift surveys feature a large number of galaxies, possibly in the range 10^6 for the Sloan and 2dF surveys. Adaptive smoothing techniques should prove very powerful for this kind of data. We now work out a first test of this idea for the 1.2-Jy redshift survey.

6.4 The 1.2-Jy redshift survey and Virgo mock catalogues

6.4.1 The 1.2-Jy redshift survey data

The data of the 1.2-Jy redshift survey (Strauss et al. 1990) of *IRAS* galaxies has been published (Fisher et al. 1995) and can be retrieved electronically from the Astronomical Data Center (ftp://adc.gsfc.nasa.gov). The 5321 galaxies of the survey are selected from the PSC catalogue above a flux limit of 1.2 Jy in the 60 μm band. The sky coverage is

87.6 per cent, excluding only the zone of avoidance for $|b| < 5^{\rm o}$ and a few unobserved or contaminated patches at higher latitude.

There have been numerous studies of the 1.2-Jy survey, essentially covering all standard methods for analysing large-scale structure. Very recently the catalogue has also been examined with topological methods. Yepes et al. (1997) used a percolation analysis, Kersher et al. (1998) employed Minkowski functionals, and Protogeros & Weinberg (1997) used genus statistic.

Our approach is similar to that of the latter authors. We also work with mock catalogues to derive the statistical properties of the genus statistics. However, in contrast to Protogeros & Weinberg (1997) we do not volume-limit our catalogues but use instead a selection function weighting to derive the density fields. Additionally, we introduce adaptive smoothing techniques. Although we agree with their main conclusions, we find no confirmation of their finite volume bias, and we explicitly show that the errors of the genus curve are not multivariate normally distributed.

In our analysis of the 1.2-Jy survey we convert the redshifts to the Local Group frame and use them to infer distances without further corrections for peculiar velocities. Redshift space distortions have only a negligible effect on the genus, as has been shown in a number of studies.

We will assume an Einstein-de-Sitter model for the background cosmology throughout. The results will not be sensitive to this choice because the 1.2-Jy density field maps only the very local Universe.

We define the selection function $S(z) = \langle m(\mathbf{r}) \rangle$ of the survey as the mean expected comoving number density of sources at redshift $z = |\mathbf{r}|$. We employ the fitting form

$$S(z) = \frac{\psi}{z^{\alpha}\left[1 + \left(\frac{z}{z^{\star}}\right)^{\gamma}\right]^{\frac{\beta}{\gamma}}}, \tag{6.12}$$

and adopt the parameters (Table 6.2) determined by Springel & White (1998). Note that the selection function includes a correction for the strong evolution seen in *IRAS* galaxies.

6.4.1.1 Depth of maps

The galaxy density of a flux limited sample falls off quickly with distance. As a consequence, the uncertainty in the density estimate grows rapidly with redshift. It is desirable, of course, to use a survey volume that is as large as possible in order to beat down statistical noise and cosmic variance. According to a useful rule of thumb (Weinberg et al. 1987) discreteness effects are approximately negligible if

$$\lambda \geq d = S^{-\frac{1}{3}}, \tag{6.13}$$

where d is the mean inter-galaxy separation. Adopting this criterion we choose a maximal radius $R_{\rm max}$ by $\lambda = S(R_{\rm max})^{-\frac{1}{3}}$ and use it to delimit the usable survey volume $V_{\rm s}$. This choice ensures that at the far edge of the survey volume the sampling condition is just met, and in the remainder of the volume the sampling is denser.

Table 6.2: Parameters of the selection function of the 1.2-Jy survey.

α	β	γ
$0.741^{+0.128}_{-0.135}$	$4.210^{+0.419}_{-0.344}$	$1.582^{+0.237}_{-0.214}$
$z^\star$	ψ $[h^3\mathrm{Mpc}^{-3}]$	
$0.0184^{+0.00213}_{-0.00167}$	$(486.5 \pm 13.0) \times 10^{-6}$	

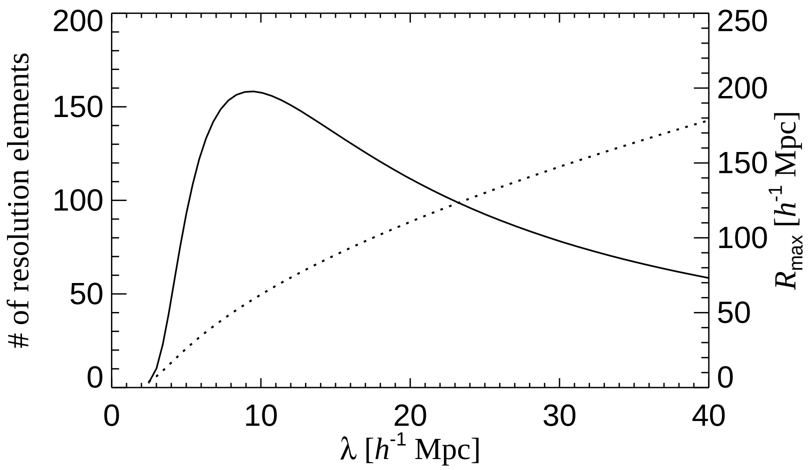


Figure 6.10: The number of resolution elements (solid) for the 1.2-Jy survey when the maximal survey volume is used. Also shown is the radius (dashed) of the usable survey volume.

Table 6.3: The smoothing lengths adopted for the topological analysis of the 1.2-Jy survey. Listed are the chosen survey depth R_{max}, the resulting number N_{res} of resolution elements and the number N_{gal} of galaxies inside the survey volume.

λ $[h^{-1}\mathrm{kpc}]$	R_{max} $[h^{-1}\mathrm{kpc}]$	N_{res}	N_{gal}
5	26.00	92.6	1030
7	42.23	144.7	1783
10	62.07	157.6	2820
14	83.73	140.9	3582
20	110.63	111.5	4199

6.4.1.2 Resolution elements

The notion of number of resolution elements provides a useful way to compare roughly the statistical power of genus measurements. Because the smoothing extends over an effective volume $V_{\rm sm} = \pi^{3/2}\lambda^3$ the number of independent structures that can be present in a finite survey volume is limited. This number is of order

$$N_{\rm res} = \frac{V_{\rm s}}{V_{\rm sm}} = \frac{\omega R_{\rm max}^3}{3\pi^{3/2}\lambda^3}, \tag{6.14}$$

where ω is the solid angle covered by the survey.

The number $N_{\rm res}$ indicates the power of a data set used for topological analysis. With the QDOT survey Monaghan (1992) reached a maximum of about $N_{\rm res} = 80$ whereas the CfA survey allowed Vogeley et al. (1994) to achieve $N_{\rm res} = 260$ for their best subsample. The most powerful dataset examined so far is PSCz, making it possible for Canavezes et al. (1998) to reach $N_{\rm res} = 415$. For the 1.2-Jy redshift survey we have to be content with $N_{\rm res} = 158$ (see Figure 6.10). This already indicates that one can hardly expect very tight constraints from the genus statistics of this data set. Instead, the sparse sampling of the density field can be expected to severely limit the power of the genus test.

For the genus analysis we have examined the smoothing lengths 5, 7, 10, 14, and $20\,h^{-1}$kpc in approximately logarithmic spacing. Table 6.3 lists some relevant parameters for the different cases.

6.4.2 Construction of mock surveys

We use the Virgo N-body simulations to obtain artificial redshift surveys that mimic the statistical properties of the 1.2-Jy survey with respect to sky coverage, selection function and luminosity distribution. The suites of mock catalogues are then analysed in exactly the same way as the 1.2-Jy redshift survey data. In this way sampling noise, cosmic variance, and systematic biases can be reliably modeled, which allows a fair assessment of the viability of the models, even if strong biases in our analysis existed.

To construct a mock catalogue we first select an arbitrary observer position in the periodic simulation box, which we replicate periodically in order to allow the construction of catalogues with sufficient depth. For computational convenience, we adopt a depth of $239.5\,h^{-1}$kpc, corresponding to the size of the simulation box. We do not attempt to restrict the observers to positions that match certain properties of the immediate neighbourhood of the solar system because we want to allow for a realistic degree of cosmic variance. Because we have only one simulation at our disposal, not all the mock catalogues are independent. However, the survey volumes considered are much smaller than the simulation volume itself.

We identify every dark matter particle with a possible galaxy site, i.e. we construct only unbiased galaxy catalogues. We then compute the cosmological redshifts z_i of the potential galaxy sites with respect to the observer and draw uniformly distributed random numbers $x_i \in [0,1]$ for them. Only those galaxies with $x_i \leq S(z_i)$ are kept, where $S(z)$ is the selection function scaled such that $S(z_0) = 1$ for some very small z_0. This selection results in a catalogue with selection function proportional to $S(z)$.

We finally discard the galaxies behind the angular mask of the 1.2-Jy survey and degrade the source density randomly such that the mock surveys contain the same number of sources (5083) as the 1.2-Jy survey in a sphere of radius $239.5\,h^{-1}$kpc.

In a second step (optional for this work) we assign observed fluxes to the galaxies according to the luminosity function that results as a consequence of the adopted selection function (Springel & White 1998). For this purpose we draw a random number $q_i \in [0,1]$ from a uniform distribution, compute a maximum redshift

$$z_{\rm max} = S^{-1}\left[(1-q_i)S(z_i)\right], \tag{6.15}$$

and assign an observed flux

$$f_i = f_{\rm min}\frac{r_{\rm max}^2}{r_i^2}\left(\frac{1+z_{\rm max}}{1+z_i}\right)^{1-\alpha} \tag{6.16}$$

for each source. Here S^{-1} denotes the inverse selection function, and $f_{\rm min} = 1.2$ Jy is the flux limit. Note we here assumed a straight K-correction with $\alpha = -2$, a non-evolving luminosity function, and an Einstein-de-Sitter universe. For $\Omega_0 \neq 1$, the latter may be replaced by the appropriate cosmology.

We usually neglect peculiar velocities in the construction of mock catalogues. As pointed out above, redshift space distortions are generally found to have only a minor effect on the genus statistics, at least on the relatively large scales accessible so far. We test that this is indeed the case by producing an additional set of mock catalogues in redshift space, where we adopt the velocities of the dark matter particles themselves as peculiar velocities of the mock galaxies.

6.4.2.1 The mock samples

We construct two main suites of mock catalogues; one is drawn from the SCDM simulation, the other from the ΛCDM simulation. Each of them contains 500 1.2-Jy mock surveys, that feature 5083 galaxies on 87.6 per cent of the sky, flux limited at 1.2-Jy and volume limited at a depth of $239.5\,h^{-1}$kpc.

These samples are used to develop the statistical methodology for the comparison of the 1.2-Jy data to the N-body models. The large number of mock catalogues is necessary to determine the statistical distribution of the genus reliably. We restrict ourselves to the SCDM and ΛCDM simulations because we do not expect the genus of such a sparse sample as 1.2-Jy to be able to discriminate between ΛCDM, τCDM, and OCDM.

For the adaptive smoothing technique that we outline below, we fill the masked regions of the sky with a small set of fake galaxies. For simplicity, we construct for each catalogue 733 points with random angular positions in the masked regions, and with a redshift distribution sampled from the catalogue itself.

To examine a number of systematic effects we construct additional mock catalogues. For example, we use 100 mock catalogues in redshift space to assess the influence of redshift space distortion on the genus. We further compute additional catalogues that are full-sky 1.2-Jy surveys, i.e. which have no angular mask. These are used to test the influence of the mask. Finally, we compute a number of different survey realizations for fixed observer positions in order to separately determine the relative importance of sampling noise and cosmic variance.

6.5 Smoothing techniques

We now outline our procedures to construct smoothed density fields from the 1.2-Jy redshift survey and the mock catalogues. These fields are then used as input to the genus computation.

6.5.1 Fixed smoothing

Assuming a universal luminosity function an unbiased estimate of the galaxy density field can be obtained by weighting the discrete point distribution $m(\mathbf{r})$ of the observed galaxies with the inverse of the selection function $S(r)$:

$$\rho(\mathbf{r}) \propto \frac{m(\mathbf{r})}{S(r)}. \tag{6.17}$$

We obtain an estimate of the density field smoothed on some scale λ by convolving with a filter $W(\mathbf{r})$, which we choose as the Gaussian of equation (6.9).

However, due to the absence of galaxies in the regions of the angular mask, the density would be systematically underestimated at locations close to unobserved patches of the sky if the smoothing were just done by a straightforward use of the kernel of equation (6.9). In order to avoid this problem we employ the ratio method proposed by Melott & Dominik (1993), who have shown in a systematic study that a smoothing according to

$$\hat{\rho}(\mathbf{r}) = \frac{\int W(\mathbf{r}-\mathbf{r}')\rho(\mathbf{r}')\,\mathrm{d}\mathbf{r}'}{\int W(\mathbf{r}-\mathbf{r}'')M(\mathbf{r}'')\,\mathrm{d}\mathbf{r}''}, \tag{6.18}$$

leads to the smallest loss or distortion of topological information compared to a number of alternative schemes that treat the mask differently. Here $M(\mathbf{r})$ is a mask field defined to be equal to 0 for $\mathbf{r}$ lying behind the angular mask and to be 1 otherwise. For this choice the denominator of equation (6.18) essentially renormalizes the smoothing kernel to the survey volume visible from the reference point $\mathbf{r}$.

In the actual computation of the genus curve we only use the volume with $M(\mathbf{r}) = 1$ which is not hidden by the mask. Additionally, we restrict the genus computation to a sphere carved out of the smoothed density field. Note that there is no boundary smoothing effect due to the outer surface of this sphere since we also include the sources outside this final region in the smoothing process.

We compute the convolutions that appear in the numerator and denominator of equation (6.18) with the help of a Fast Fourier Transform (FFT) on a 128^3 mesh. We choose a grid size of $b = \lambda/8$, which ensures that the genus is free of finite mesh size effects (Hamilton et al. 1986), as we will demonstrate below. The final depth $R_{\rm max}$ of the density field we use for the topological analysis is always small enough to avoid wrap around effects due to the periodic FFT smoothing.

6.5.2 Adaptive smoothing

As we saw in the analysis of the Virgo simulations, the smoothing scale is limited by the poor sampling of underdense regions. To make use of the additional resolution in high density regions a variable smoothing length can be employed. We have already implemented an adaptive smoothing scheme for the fully sampled N-body simulations. We now define such a scheme for the analysis of flux limited redshift surveys as well. Again, the hope is that the overall effect of adaptive smoothing is an increase of the number of structure elements visible in a given density field.

6.5.2.1 Spherically symmetric kernel

We start by making the kernel a function of position, i.e. the smoothed density field is computed as

$$\hat{\rho}(\mathbf{r}) = \int \rho(\mathbf{r}')W(\mathbf{r}-\mathbf{r}';\mathbf{r}')\mathrm{d}\mathbf{r}'. \tag{6.19}$$

This definition corresponds to the *scatter* approach of SPH. Since this scheme strictly conserves 'mass' (if W is properly normalized), it seems more natural to use for a small number of tracer particles than the *gather* formulation, which we employed for the densely sampled N-body simulations.

In a first variant of the adaptive smoothing, we stay with a spherical Gaussian and allow only the smoothing scale to vary with the local density. Later we will generalize the technique to triaxial Gaussians.

For a given survey volume $V_{\rm s}$ and a prescribed smoothing scale λ_0 we first compute the average mass M_0 in a Gaussian sphere of radius λ_0, that is

$$M_0 \equiv \frac{\pi^{\frac{3}{2}}\lambda^3}{V_{\rm s}} \int_{V_{\rm s}} \rho(\mathbf{r})\mathrm{d}\mathbf{r} = \pi^{\frac{3}{2}}\lambda_0^3, \tag{6.20}$$

where the last equality holds (at least on average) due to our normalization of the selection function.

We then compute for every galaxy site $\mathbf{r}_i$ an individual smoothing radius λ_i such that

$$\int \rho(\mathbf{r}) \exp\left[-\frac{(\mathbf{r}-\mathbf{r}_i)^2}{\lambda_i^2}\right] = M_0. \tag{6.21}$$

This definition implies that λ_i will be smaller than λ_0 if the density around $\mathbf{r}_i$ is higher than the mean, and it will be larger if the local density is lower than the mean.

6.5.2.2 Triaxial kernel

Up to now we have only varied the volume of the kernels. In an attempt to improve the flexibility of the smoothing we may also allow the shape of the kernel to vary. For this purpose we adopt triaxial Gaussians

$$W(\mathbf{x};\mathbf{r}) = \frac{1}{\pi^{\frac{3}{2}}(\det\mathbf{\Lambda})^{\frac{1}{2}}} \exp\left(-\mathbf{x}^T\mathbf{\Lambda}^{-1}\mathbf{x}\right) \tag{6.22}$$

as kernels, where the quadratic form $\mathbf{\Lambda}$ is a function of $\mathbf{r}$. We now need to specify the matrices $\mathbf{\Lambda}_i$ for every particle. For simplicity, we set $\mathbf{\Lambda}_i$ proportional to the local moment of inertia tensor around the site $\mathbf{r}_i$, i.e.

$$\mathbf{\Lambda}_i \propto \int (\mathbf{r}-\mathbf{r}_i)(\mathbf{r}-\mathbf{r}_i)^T \rho(\mathbf{r}) \exp\left[-\frac{(\mathbf{r}-\mathbf{r}_i)^2}{\lambda_i^2}\right] \mathrm{d}\mathbf{r}, \tag{6.23}$$

and we keep the original smoothing volume fixed by requiring $(\det\mathbf{\Lambda}_i)^{\frac{1}{2}} = \lambda_i^3$. Once the matrices $\mathbf{\Lambda}_i$ are determined we 'only' need to compute

$$\hat{\rho}(\mathbf{r}) = \sum_i m_i W(\mathbf{r}-\mathbf{r}_i;\mathbf{\Lambda}_i) \tag{6.24}$$

with $m_i = [S(\mathbf{r}_i)]^{-1}$ to arrive at the triaxially smoothed density field.

Because the adaptive smoothing has to be carried out in real space it requires much more CPU time than the fixed smoothing described above. For this reason we construct the density fields only in spheres of radius $R_{\max}$, each of them inscribed in a 64^3 mesh. For simplicity we deal with the mask by filling the empty region with fake galaxies as described above. Note that the masked volume is not used in the calculation of the genus curve.

The adaptive smoothing schemes we employ here are by no means unique and many other variants are conceivable. In fact, we have tried a number of alternatives ourselves. However, the suggested procedure is fairly intuitive and allows a study of the effects of volume and shape adaptivity separately. Currently we see no alternative to Monte-Carlo experiments in determining the performance and properties of such adaptive smoothing techniques.

6.6 Systematic effects

Before turning to the results for the 1.2-Jy survey and the Virgo mock catalogues, we first examine various systematic effects that can affect the genus measurements. In particular, we are interested in the amount of bias present in the genus curve derived from a 1.2-Jy-like sample compared to the genus of the underlying density field.

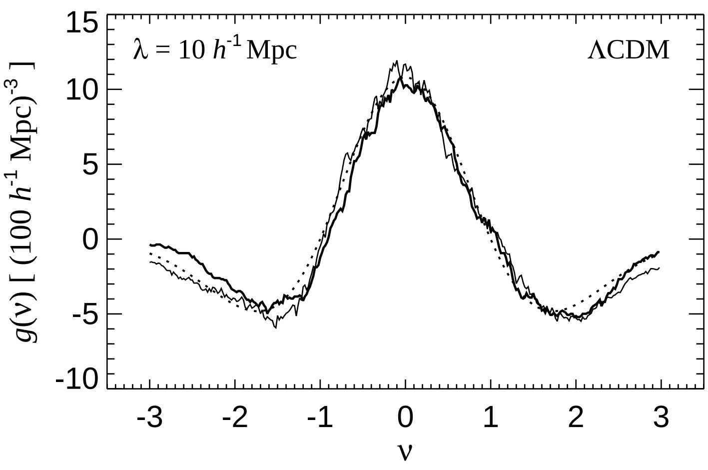


Figure 6.11: Finite volume effect. The thin line shows the average genus curve of 10 spherical subvolumes of radius $60\,h^{-1}$kpc extracted from the SCDM simulation, while the thick line gives the genus of the full simulation volume. The genus curves have been computed for a resolution of $\Delta\nu = 0.01$. Comparing the two curves we find no evidence for the strong amplitude bias claimed by Protogeros & Weinberg (1997). They found an increase of the genus amplitude for the finite volume curves by a factor of 1.5-2.0.

6.6.1 Finite volume effect

Recently, Protogeros & Weinberg (1997) have claimed that the genus curve is severely biased high if it is computed from subvolumes carved out of large N-body simulations or out of Gaussian random fields. In particular, for a roughly spherical volume of radius $R_{\rm max} = 60\,h^{-1}$kpc and smoothing scale $\lambda = 10\,h^{-1}$kpc they detected an increase of the genus amplitude by a factor of 1.5-2. They further found that this volume effect becomes somewhat smaller for larger volumes, yet it seems to be present independently of the shape of the survey volume.

Their result is surprising since the volumes used in their analysis seem large enough that any correlation of the curvature between adjacent points on the surface of the volume is expected to average out, i.e. the mean curvature of the subvolume should allow an unbiased estimate of the genus of the full volume.

We have searched for this volume effect ourselves, but we could not find it. For example, in Figure 6.11 we compare the average genus curve for 10 spherical subvolumes of radius $R_{\rm max} = 60\,h^{-1}$kpc extracted from the SCDM simulation with the genus of the full simulation box. Reassuringly, the average genus curve of the subvolumes appears to be largely unbiased; there is no trace of a strong finite volume effect.

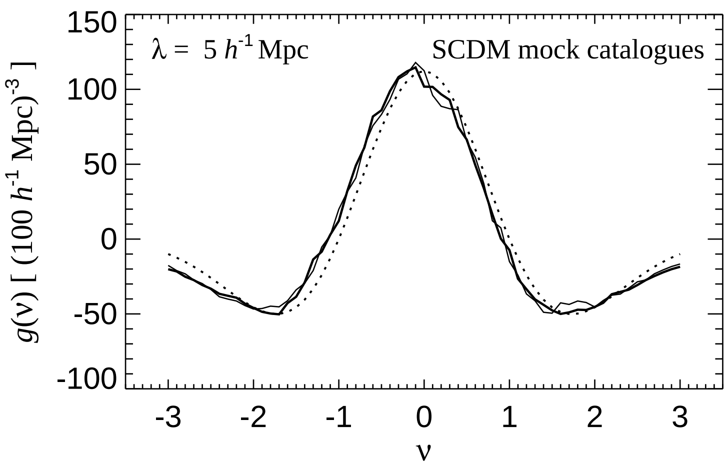


Figure 6.12: Effects due to the angular mask. The thick curve shows the average genus curve for 100 SCDM mock catalogues before they are subjected to the 1.2-Jy mask, i.e. their genus is computed for an all-sky density field. The thin line shows the genus for the masked mock catalogues, where the genus computation involves the ratio method. The curves shown are for fixed smoothing with $\lambda = 5\,h^{-1}$kpc. Larger smoothing scales or the adaptive smoothing techniques result in similar small effects.

6.6.2 Mask effects

We dealt with the limited sky coverage of the 1.2-Jy survey by employing the ratio smoothing method of Melott & Dominik (1993). Here we look for systematic biases inflicted on the genus curve because of that. For this purpose we have constructed an additional suite of 100 mock catalogues for the SCDM model. These catalogues have full sky coverage, with 5816 galaxies to a depth of $239.5\,h^{-1}$kpc.

In Figure 6.12 we compare the average genus curve of these full sky mock catalogues to the average curve of the masked catalogues, where the computation involves the ratio method. The good agreement between the two results gives us confidence that any bias due to our treatment of the mask is very small. The result shown is for fixed smoothing with $\lambda = 5\,h^{-1}$kpc, but we observe a similar small influence of the mask for other smoothing scales and for the adaptive smoothing techniques.

6.6.3 Genus curves

Figure 6.13 shows the average genus curve of the SCDM suite of mock catalogues compared to the genus of the fully sampled simulation. Obviously the genus curve derived from the mock catalogues is biased high compared to full sampling. This can be explained by the influence of shot noise on the genus amplitude, as demonstrated by Canavezes

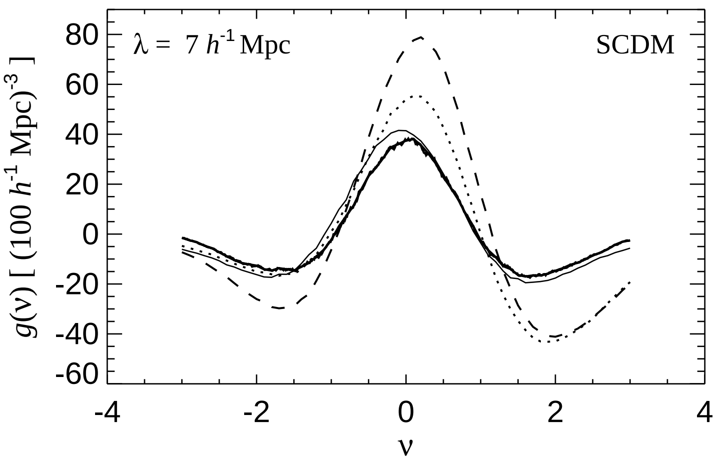


Figure 6.13: Genus curves for the SCDM simulation, smoothed at $7\,h^{-1}$kpc with different smoothing techniques. The thick solid line gives the result for the fully sampled SCDM simulation, while the thin line gives the corresponding average genus curve with a fixed smoothing kernel for the SCDM mock catalogues. The dotted line shows the result for the spherical adaptive smoothing, and the dashed curve is for triaxial adaptive smoothing.

et al. (1998). Furthermore, the average curve exhibits a slight *meatball* bias, that is a shift of the peak to the left. This effect can also be understood as a discreteness error that results from the sparse sampling of the density field.

Also shown in Figure 6.13 are the mean genus curves resulting from the adaptive smoothing techniques we tried. Both of them show a strong enhancement of the genus amplitude. The spherically symmetric smoothing also results in a strong asymmetry between the minima of the genus curve, with the minimum on the positive side being lowered strongly, while the minimum on the negative side remains practically at the level obtained for the fixed smoothing technique. Note that the genus curves for the different smoothing techniques are expected to be different since the smoothing procedures are sensitive to different properties of the density field.

Interestingly, the triaxial smoothing technique gives a genus very close to the spherical smoothing around the minimum on the positive side, while it increases the genus signal for smaller values of ν. Presumably the amplitude of the minimum on the positive side is just set by the number density of individual, isolated clumps. Apparently, the triaxial technique does not change the topology of these isolated regions.

For fixed smoothing, the measured genus is biased compared to the fully sampled simulations, albeit by a small amount. Instead of trying to correct for it as attempted by Canavezes et al. (1998), we compare the 1.2-Jy measurements only to the results obtained for the mock catalogues, and not to the fully sampled simulations themselves. This is a viable procedure, even if strong biases are present.

6.7 Statistical methodology

Ultimately we want to use the genus statistic to compare theory with observation, i.e. to quantify the level of agreement of the 1.2-Jy survey with the Virgo N-body models. A prerequisite to derive formal exclusion levels is a precise understanding of the distribution of errors of the genus measurement.

Perhaps the most general method to assess random and systematic errors is to work with ensembles of mock galaxy surveys that mimic the statistical properties of the actual observed data set. When the mock catalogues and the redshift survey are analysed in the same way, systematic biases that might be present in the adopted analysis enter in the same way.

6.7.1 Distribution of errors

For a suite of n mock catalogues we measure the genus curve at k values $\nu_1, \nu_2, \ldots, \nu_k$ of the filling factor. In what follows, we compute the genus with spacing $\Delta\nu = 0.1$ in the range $[-3.0, 3.0]$, i.e. at $k = 61$ positions. We now use these measurements to estimate the distribution of errors in the genus. The mean genus curve and its covariance matrix may be estimated as

$$\overline{\mathbf{g}} = \frac{1}{n} \sum_{l=1}^{n} \mathbf{g}^{(l)} \tag{6.25}$$

and

$$\mathbf{V} = \mathrm{cov}(g_i, g_j) = \frac{1}{n-1} \sum_{l} (\mathbf{g}^{(l)} - \overline{\mathbf{g}})(\mathbf{g}^{(l)} - \overline{\mathbf{g}})^T, \tag{6.26}$$

where $\mathbf{g}^{(l)} = (g_1^{(l)}, \ldots, g_k^{(l)})$ denotes the measured genus curve for the catalogue l.

Recently Protogeros & Weinberg (1997) conjectured that the distribution of errors is well described by a multivariate Gaussian

$$f(\mathbf{g}) = \frac{1}{(2\pi)^{\frac{k}{2}} |\det \mathbf{V}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2} (\mathbf{g} - \overline{\mathbf{g}})^T \mathbf{V}^{-1} (\mathbf{g} - \overline{\mathbf{g}}) \right]. \tag{6.27}$$

Then the quantity

$$\chi^2(\mathbf{g}) = (\mathbf{g} - \overline{\mathbf{g}})^T \mathbf{V}^{-1} (\mathbf{g} - \overline{\mathbf{g}}) \tag{6.28}$$

would exhibit a χ^2 distribution with k degrees of freedom, and one could estimate the probability

$$p = 1 - \frac{1}{2^{\frac{k}{2}} \Gamma\left(\frac{k}{2}\right)} \int_{\chi^2}^{\infty} x^{\frac{k}{2}-1} \mathrm{e}^{-\frac{x}{2}} \mathrm{d}x \tag{6.29}$$

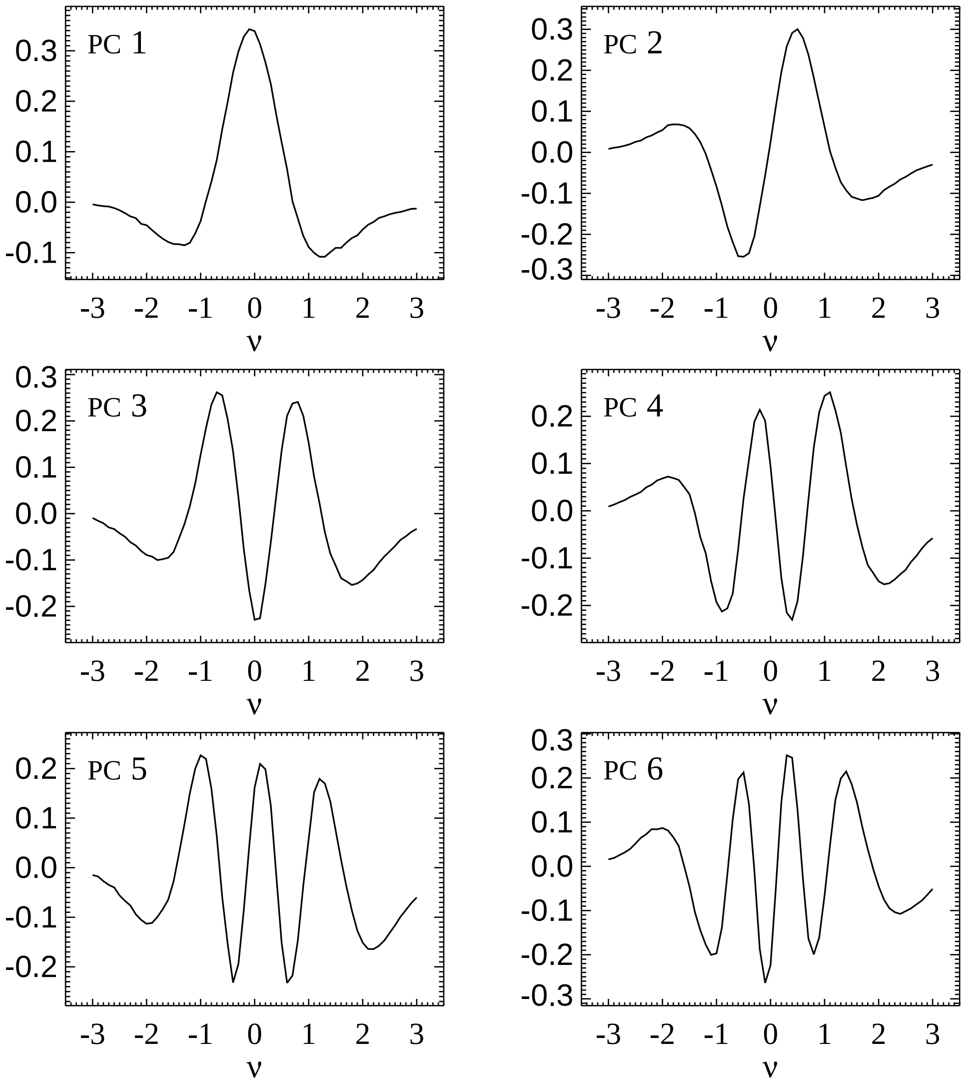


Figure 6.14: Principal components of the genus measurement. The displayed curves are based on an average of covariance matrices corresponding to different smoothing scales. The matrices have been scaled to a common highest eigenvalue before the averaging. The resulting curves are similar to the ones obtained for a single covariance matrix, but they are smoother due to the reduction of noise by the averaging. We use these curves as generic principal components in the PCA of all examined samples. Note that the principal components are normalized to unity and mutually orthogonal to each other.

of finding a mock survey in the ensemble that differs from the mean of the mock catalogues by more than a particular observation with $\chi^2 = \chi^2(\mathbf{g}^{(\mathrm{obs})})$. In this way a formal exclusion level could be derived.

This procedure seems attractive, yet we find that its application requires a good deal of caution. As we show in the appendix of this chapter, the method is partly compromised by the fact that the errors are at best approximately distributed as a multivariate Gaussian. More importantly, we will use a principal components analysis (PCA) to demonstrate that there are only a small number of principal components that can be determined with some confidence. The rest of them are dominated heavily by noise; it is therefore not a good idea to invert the noisy 61×61 covariance matrix of equation (6.26). Instead we will regularize the problem by means of a PCA. Note that a simple smoothing of the genus curve can remove some of the noise. However, as a side effect this will make the covariance matrix close to singular. Of course, this causes trouble if one naively goes ahead and tries to compute V_{ij}^{-1}, so a PCA cannot be avoided in this way.

It should be noted that in the application of the χ^2-method by Protogeros & Weinberg (1997) and Colley (1997) the noise was presumably far less problematic than in our analysis. Especially their smaller number of points on the genus curves, and also the slightly larger number of independent mock catalogues they used, might have reduced the noise to a level that prevented the failure of the χ^2-method that we observe in this study.

6.7.2 Principal components analysis

The principal components analysis (PCA, Murtagh & Heck 1987) is frequently applied in astronomy to extract the most relevant features from data sets that may be strongly contaminated by noise. Since PCA is a linear method it works best for uncorrelated noise, a situation not really valid for the genus. However, we can still expect that it allows the construction of a clean multivariate analysis of the measured genus curves.

Each measured genus curve can be viewed as a point in a 61-dimensional space. The idea of PCA is to transform to a new set of coordinates which correspond to the directions of maximum extension of the cloud of measured genus points. These principal axes are just the eigenvectors of the covariance matrix (6.26). It is convenient to order these components in descending order of their eigenvalues, i.e. the first principal component shows the highest variance. Usually one then considers only the first few principal components, which are the ones that describe the most prominent features of the signal. In this way the method allows to efficiently filter out the noisy contributions to the signal and to concentrate on its essential features.

The principal axes and their eigenvalues can be conveniently found by a singular value decomposition (SVD) of the covariance matrix. Having ordered the principal components in descending order we consider only the first m of them. We can then construct a $m \times 61$ matrix $\mathbf{P}$ that contains the m eigenvectors in its rows and that projects a genus curve $\mathbf{g}$ onto new coordinates $\mathbf{h} = \mathbf{P}\mathbf{g}$. One can also transform back to the original space, giving rise to a PCA-filtered genus curve

$$\mathbf{g}_{\mathrm{PCA}}(m) = \mathbf{P}^T \mathbf{P}\mathbf{g}. \tag{6.30}$$

In Figure 6.14 we display the first six principal components of the genus measurement.

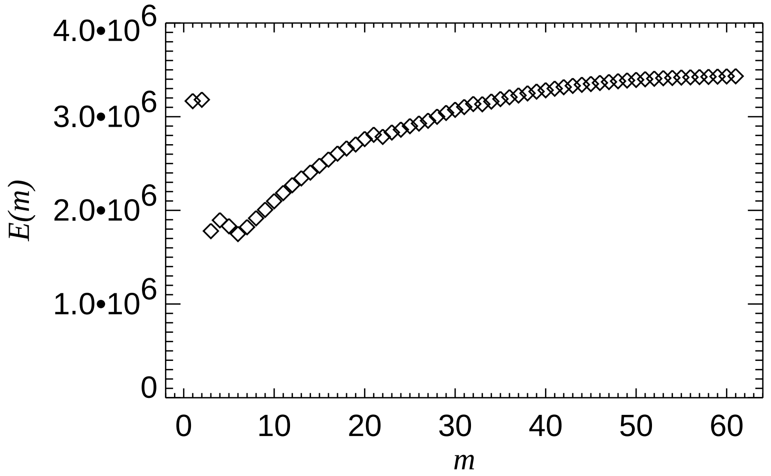


Figure 6.15: Global approximation error E as a function of the number m of included principal components. The points shown are for the SCDM mock catalogues, smoothed with $7\,h^{-1}$kpc. Other smoothing scales show similar results. For the adaptive smoothing techniques the minimum occurs also at $m \approx 6$. However, the higher components lead to a somewhat smaller increase of E.

Actually these principal components have been derived from an averaged covariance matrix, obtained by adding up matrices corresponding to several smoothing scales which we have scaled to a common highest eigenvalue. The curves resulting from just one suite of genus curves look very similar, although they are not quite so smooth. We employ the averaging procedure to establish a generic set of smooth orthogonal principal components that we subsequently apply to all the different samples on an equal footing. Of course, depending on the covariance matrices used in the averaging procedure the derived principal components may differ slightly in detail. However, in all cases the first few components (we will use $m = 6$ of them) span very nearly the same region in the full 61-dimensional space of measured genus curves. Hence the specific choice of covariance matrices is uncritical as long as all the information of these principal components is used in a multivariate analysis.

Some of the principal components shown in Figure 6.14 are easy to interpret. The first clearly measures the amplitude, while the next two can be seen to be sensitive to a horizontal shift and a broadening of the genus curve. Hence the first three principal components are similar in meaning to the genus meta-statistics introduced by Vogeley et al. (1994). However, here these measures are not postulated in an ad-hoc way, but they suggest themselves naturally as the most relevant features of the measured genus curves.

How many principal components should we reasonably take? To answer this question

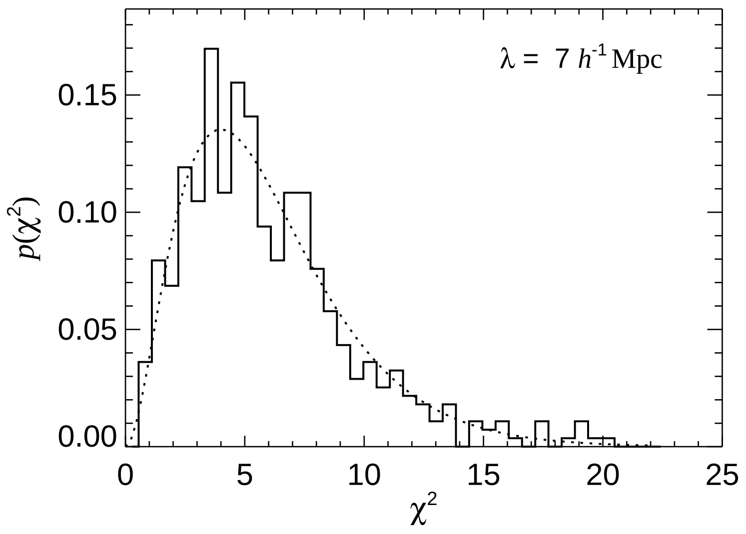


Figure 6.16: Distribution of $\chi^2(\mathbf{h})$, when only the first 6 principal components are included. The variance of the distribution is 12.66, close to the expected value of 12.0 for a χ^2 distribution with $m = 6$ degrees of freedom (dotted line). The data shown here are for the SCDM catalogues with fixed smoothing of $7\,h^{-1}$kpc. All the other cases show equally good agreement with the theoretical χ^2 distribution. Note that the test curve $\mathbf{g} = 2\overline{\mathbf{g}}$, which demonstrates the failure of the method applied in the appendix of this chapter (Figure 6.22), here gives a χ^2 of 32.4, indicating correctly a terrible fit.

we examine a global error function

$$E(m) = \sum_l (\mathbf{g}^{(l)}_{\mathrm{PCA}}(m) - \overline{\mathbf{g}})^2, \tag{6.31}$$

where the sum is over the mock ensemble and $\mathbf{g}^{(l)}_{\mathrm{PCA}}(m)$ denotes the genus curve of catalogue l, treated with a PCA filter of order m. Figure 6.15 shows a minimum of $E(m)$, when $m \approx 6$ principal components are used. Higher principal components lead to additional noise, in the sense that the reconstructed genus curves differ more and more from the ensemble average. For this reason we will restrict ourselves to the first 6 principal components.

As is seen in Figure 6.14 the first six principal components give only small weight on the genus at low and high values of ν, where we know that the errors are not normally distributed. Because of that, one can hope that the distribution of the principal components $\mathbf{h}$ is consistent with a Gaussian, at least if we only use the first $m = 6$ of them. A Kolmogorov-Smirnov test (see the appendix of this chapter) reveals that this is indeed the case. We will therefore attempt an ordinary multivariate analysis based on the fixed set of principal components displayed in Figure 6.14. Since we keep them fixed for all suites of mock catalogues, we do not expect that the correlations between

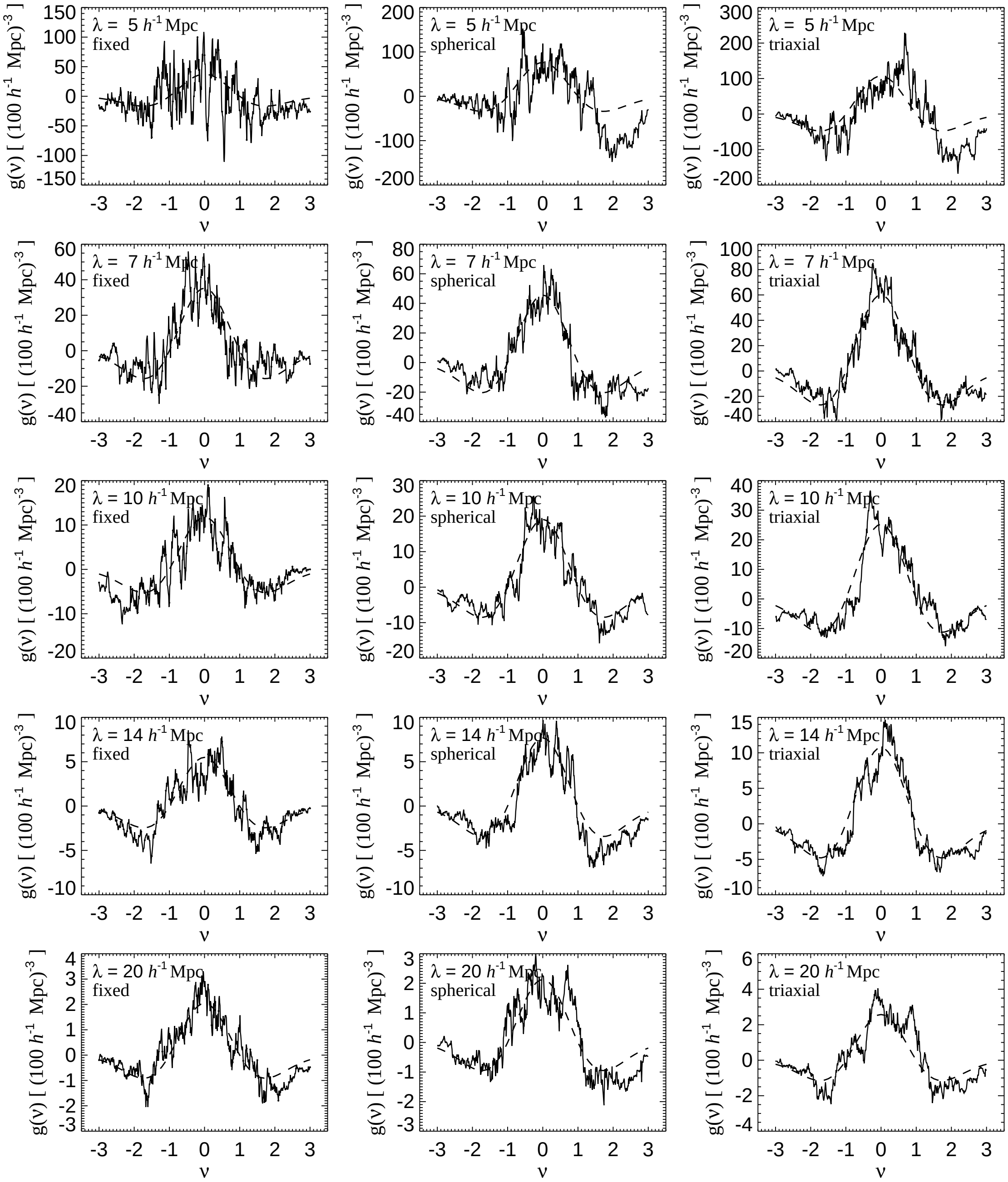


Figure 6.17: Genus curves of the 1.2-Jy redshift survey of *IRAS* galaxies. The left column shows high resolution genus curves obtained with fixed smoothing, while the middle and right columns give the results for the two adaptive smoothing schemes. In each row the same smoothing scale is considered, rising from $5\,h^{-1}$kpc to $20\,h^{-1}$kpc. Note however that the vertical scale is different for individual panels. The adaptively smoothed genus curves exhibit a considerably higher genus amplitude. The dashed lines are random phase genus curves with amplitudes determined from the first principal component of the measured genus curves. In the case of adaptive smoothing, λ sets the mass per smoothing volume via equation (6.20).

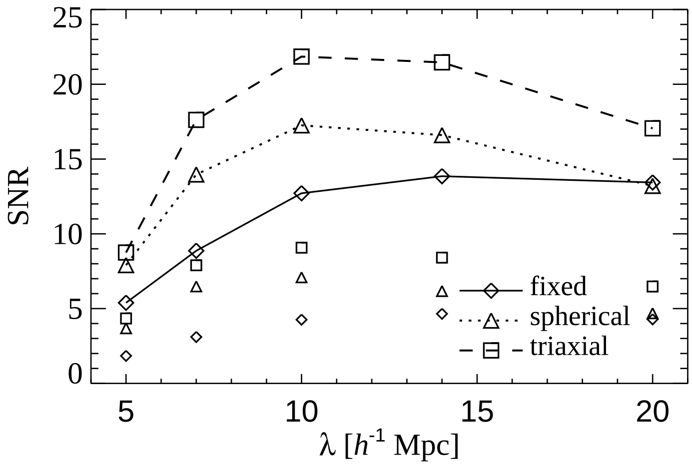


Figure 6.18: Average signal-to-noise ratio of the measured genus curves for different smoothing techniques as a function of smoothing scale. The results are for the SCDM mock catalogues and PCA-filtered ($m = 6$) genus curves. The small symbols show the corresponding SNR for the raw genus curves. The ΛCDM catalogues give a similar result.

the different principal components will exactly vanish, although they should be small. For this reason we compute the covariance matrix **C** of the measured values of **h**, and consider the statistic

$$\chi^2(\mathbf{h}) = (\mathbf{h} - \overline{\mathbf{h}})^T \mathbf{C}^{-1} (\mathbf{h} - \overline{\mathbf{h}}). \tag{6.32}$$

In Figure 6.16 we show an example of the distribution of this quantity. It is indeed very well fit by a χ^2 distribution with $m = 6$ degrees of freedom. Note that for this statistic the pathological result obtained in the appendix for the curve $\mathbf{g} = 2\overline{\mathbf{g}}$ is gone; here this curve gives $\chi^2 = 32.4$, implying a terrible disagreement, as it should.

6.8 Comparison of the 1.2-Jy survey with Virgo

6.8.1 1.2-Jy genus curves

Figure 6.17 shows the genus curves we obtain for the 1.2-Jy redshift survey with three different smoothing schemes. The left column displays results for fixed smoothing, while the middle and right columns give the corresponding curves for the two variants of adaptive smoothing. The genus curves in the Figure are computed with a uniform spacing of $\Delta\nu = 0.01$ in the interval $-3 \leq \nu \leq 3$. Note that the genus curves that are used for the comparison with the Virgo simulations represent only a subsampling of this data with a spacing of $\Delta\nu = 0.1$.

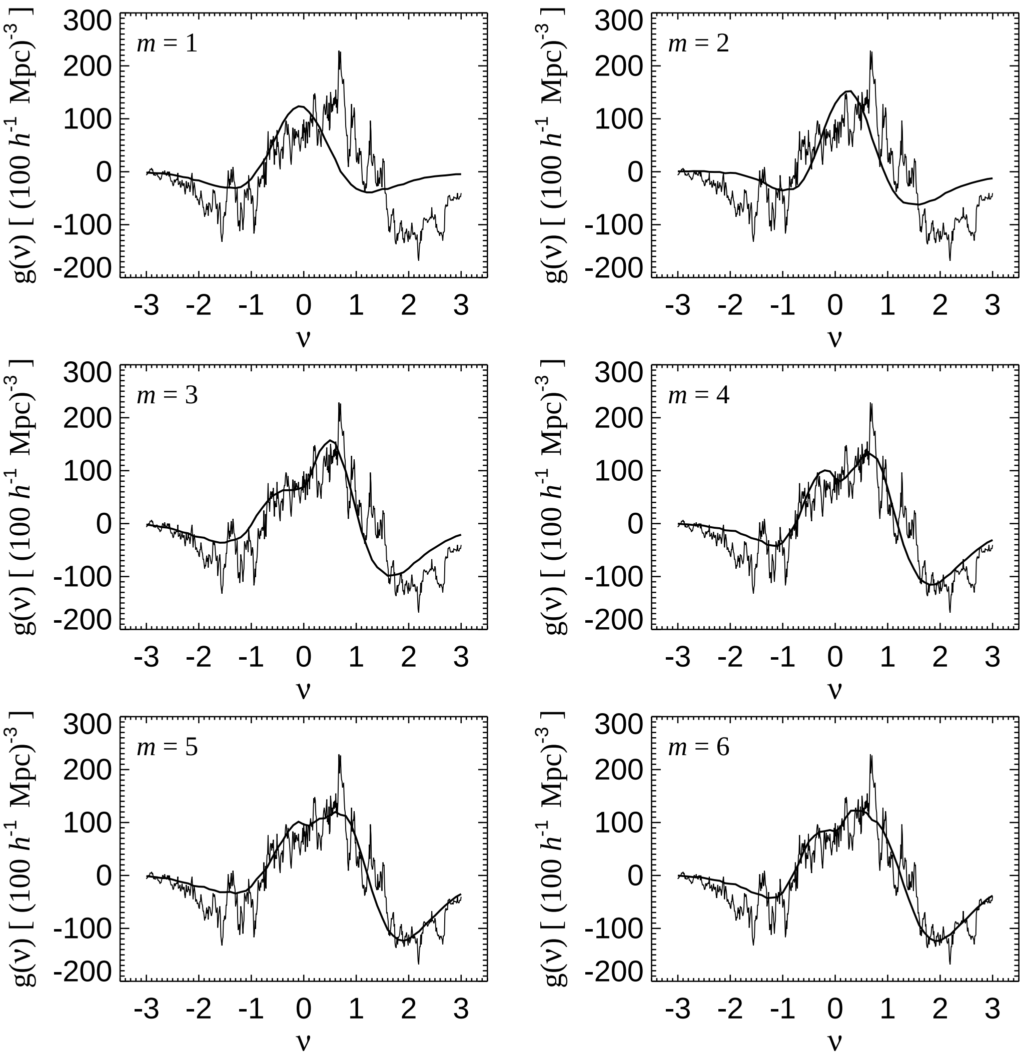


Figure 6.19: Example for a PCA-filtered genus curve as a function of the number m of included principal components. In each panel the thin line shows the 1.2-Jy genus curve resulting for triaxial adaptive smoothing with $\lambda = 5\,h^{-1}$kpc. The thick line gives the corresponding PCA-filtered curve for the specified order m of the filter.

The large amount of jitter in the curves indicates that there is substantial noise in the measurement. Interestingly, the adaptively smoothed genus curves show less scatter and seem to be smoother than with fixed smoothing. Clearly, the triaxial smoothing technique produces the smoothest genus curves. It might therefore be hoped that this reflects an increase of the number of resolved structural elements which in turn allows a measurement of the genus curve with less error.

In order to test this expectation we define an average signal-to-noise ratio (SNR)

$$\langle \mathrm{SNR} \rangle \equiv \left\langle \frac{\overline{g}^2}{\sigma^2} \right\rangle = \frac{1}{k} \sum_{i=1}^{k} \frac{\overline{g}_i^2}{\sigma^2(g_i)} \tag{6.33}$$

of the measured genus curves. Here $\overline{g}_i$ denotes the mean genus density at each of the $k = 61$ measured positions, and $\sigma^2(g_i)$ is the variance of the corresponding measurements.

In Figure 6.18 we show the SNR for the three different smoothing techniques as a function of smoothing scale. As expected, the adaptive smoothing techniques can significantly improve the SNR compared to the fixed smoothing scheme. Also, the triaxial method clearly performs better than the spherical adaptive smoothing. Note that Figure 6.18 shows the SNR for genus curves that are filtered with a PCA-filter of order $m = 6$. The SNR for the raw genus curves is also plotted; it is generally smaller, showing that the PCA-filter can indeed reduce the contamination with noise.

A further illustration of the effect of the PCA-filtering is given in Figure 6.19. Here we show the 1.2-Jy genus curve resulting for the triaxial adaptive smoothing technique with $\lambda = 5\,h^{-1}$kpc together with PCA-filtered versions of it. As the number m of included principal components is increased, more features of the measured genus curve can be faithfully reproduced by the filtered curve. However, adding in more principal components will eventually only lead to a reproduction of the noise inherent in the measured curve.

6.8.2 Comparison with Virgo

As outlined in Section 7 we base our statistical comparison between Virgo and the 1.2-Jy survey on 6 principal components of the measured genus curves. The amplitudes of the principal components are just the projections of the measured genus on the curves shown in Figure 6.14.

In Figure 6.20 we show examples of the distribution of the principal components for fixed smoothing with $\lambda = 10\,h^{-1}$kpc. In each panel, the solid histogram shows the result for the SCDM suite of mock catalogues, while the thin line gives the ΛCDM distribution. The dotted lines show the normal distributions that result from the mean and the variances of these histograms, and the dashed vertical line marks the measurement for the 1.2-Jy survey itself. The panels of the bottom half of Figure 6.20 show the corresponding genusplots for the triaxial adaptive smoothing technique.

It is immediately apparent, that the discrimination between SCDM and ΛCDM is poor with a dataset as sparse as 1.2-Jy. The amplitude h_1 proves to be the most sensitive measure of differences between the models. However, the large uncertainties in the genus measurement manifest themselves in broad, overlapping distributions of the principal components. Nevertheless we can still compute the exclusion probability of the 1.2-Jy measurements for each of the models. For this purpose we calculate the $\chi^2(\mathbf{h})$ value of the 1.2-Jy measurement with respect to the distributions of the SCDM and ΛCDM models, and the probability p that a mock catalogue shows a higher χ^2 than the

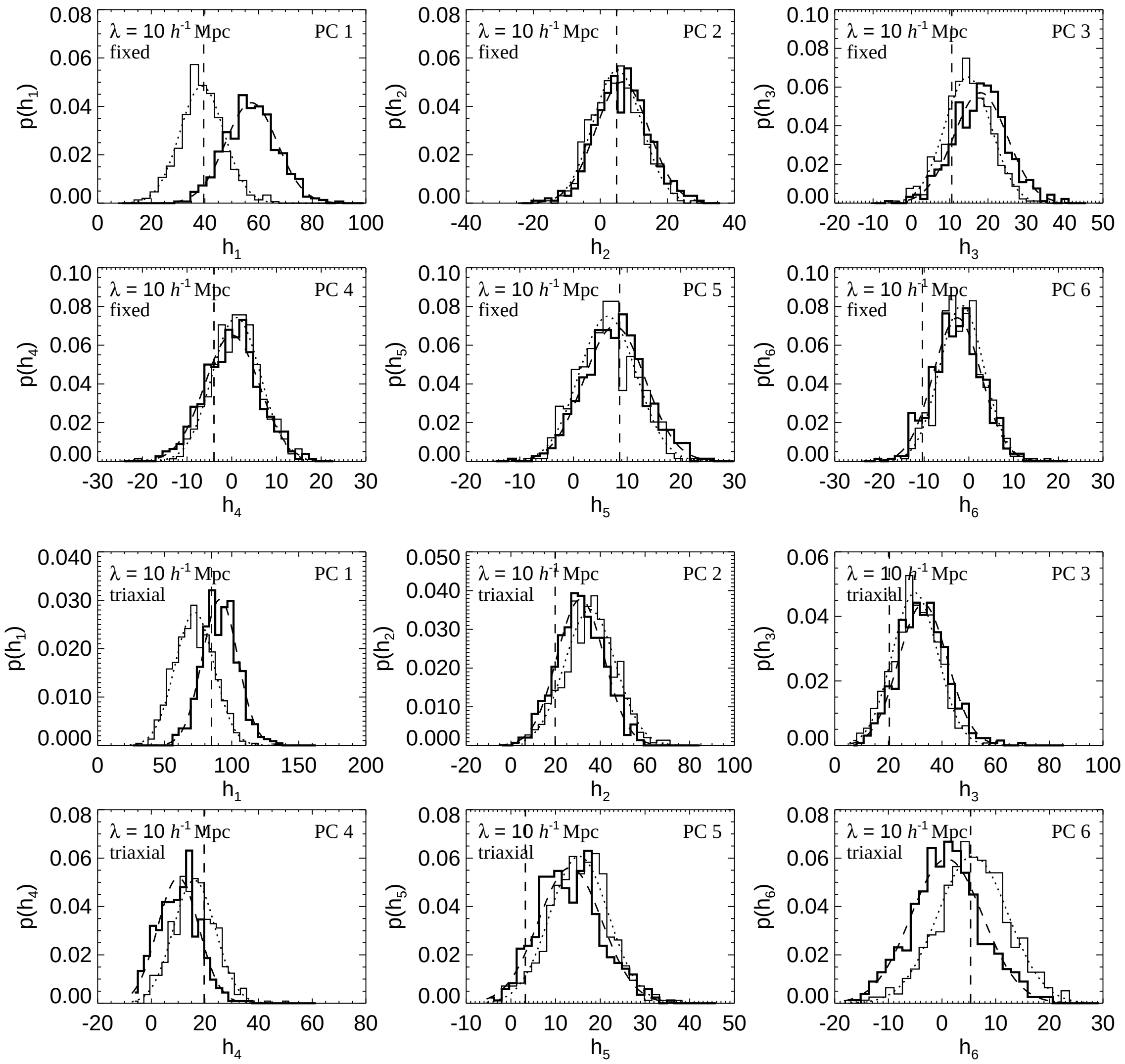


Figure 6.20: Distribution of the measured principal components for the SCDM (thick histogram) and ΛCDM (thin histogram) models. In each panel, the dashed vertical line marks the result for the 1.2-Jy survey. The dotted lines are normal distributions with the mean values and variances of the individual histograms. The data shown compares the distribution of the principal components for the fixed and triaxial adaptive smoothing techniques with $\lambda = 10\,h^{-1}$kpc. Other smoothing scales show qualitatively similar results.

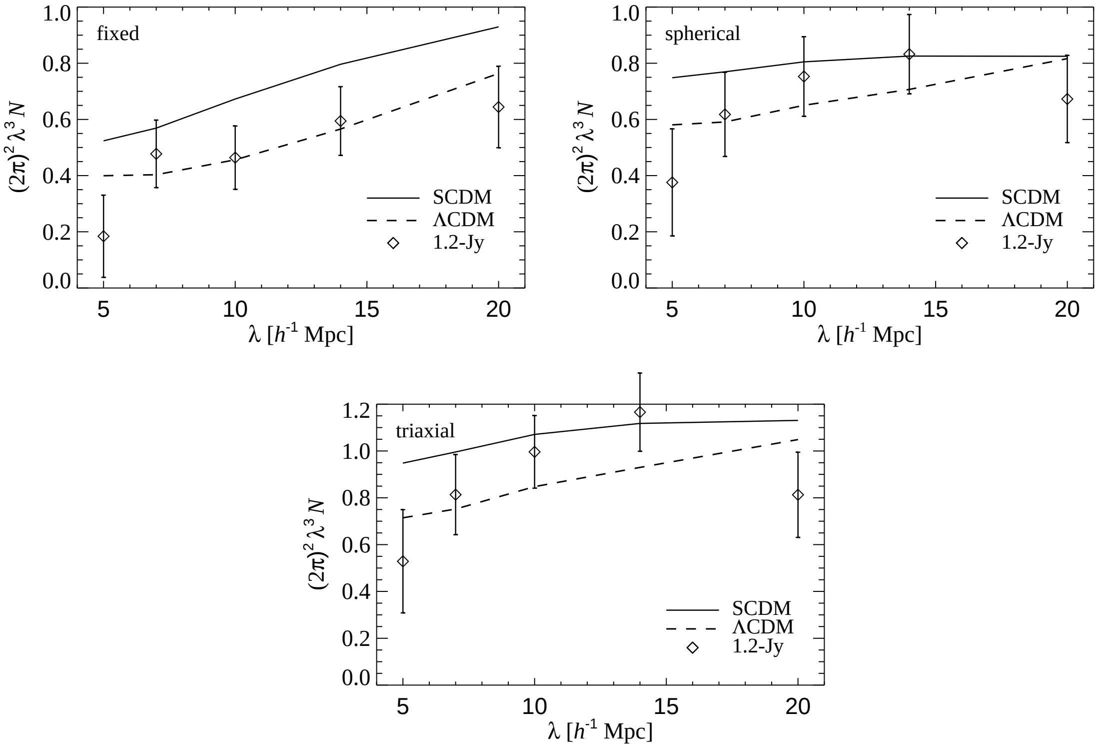


Figure 6.21: Measured genus amplitudes for the SCDM and ΛCDM mock catalogues and the 1.2-Jy survey. Each panel shows results for one of the three different smoothing schemes employed. The mean of the SCDM mock catalogues is displayed as solid line, the ΛCDM result is the dashed curve, and the diamonds mark the 1.2-Jy measurements. The attached error bars indicate the rms scatter of the SCDM sample. Note that we have normalized the first principal component such that the measured amplitude N corresponds to that of a best-fitting Gaussian curve (equation 6.5).

1.2-Jy survey. Here we assume that the principal components are multivariate normally distributed.

Additionally we give the likelihood ratio (Protogeros & Weinberg 1997)

$$\mathcal{L}\left(\tfrac{\Lambda\mathrm{CDM}}{\mathrm{SCDM}}\right) = \frac{|\det \mathbf{C}_{\mathrm{S}}|^{\frac{1}{2}}}{|\det \mathbf{C}_{\Lambda}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\left(\chi^2_{\Lambda} - \chi^2_{\mathrm{S}}\right)\right] \tag{6.34}$$

between the ΛCDM and SCDM models. Here χ^2_{S} and χ^2_{Λ} denote the χ^2-values of the 1.2-Jy data with respect to the SCDM and ΛCDM samples, and $\mathbf{C}_{\mathrm{S}}$ and $\mathbf{C}_{\Lambda}$ are the 6×6 covariance matrices of the measured principal components. We list a summary of our results in Table 6.4.

All the measured probabilities p are so high, that the 1.2-Jy genus data are consistent

Table 6.4: Comparison of SCDM and ΛCDM with the 1.2-Jy survey assuming a multivariate normal distribution for the principal components. Listed are the χ^2 values ($m = 6$ degrees of freedom) for the 1.2-Jy survey when it is compared to either the SCDM or the ΛCDM suite of mock catalogues, and the resulting exclusion levels p. We also compute the relative likelihood of ΛCDM compared to SCDM.

Fixed smoothing					
	SCDM		ΛCDM		
λ [h^{-1}kpc]	χ^2	p	χ^2	p	$\mathcal{L}\left(\frac{\Lambda\mathrm{CDM}}{\mathrm{SCDM}}\right)$
5	7.22	0.30	3.91	0.69	8.122
7	7.06	0.32	5.43	0.49	4.316
10	8.05	0.23	4.64	0.59	10.790
14	5.46	0.49	5.15	0.52	2.429
20	9.04	0.17	7.30	0.29	4.465

Adaptive smoothing (spherical)					
	SCDM		ΛCDM		
λ [h^{-1}kpc]	χ^2	p	χ^2	p	$\mathcal{L}\left(\frac{\Lambda\mathrm{CDM}}{\mathrm{SCDM}}\right)$
5	5.52	0.48	2.10	0.91	5.208
7	12.97	0.04	10.72	0.10	2.860
10	5.76	0.45	6.98	0.32	0.637
14	4.98	0.55	6.82	0.34	0.382
20	3.62	0.73	3.90	0.69	0.911

Adaptive smoothing (triaxial)					
	SCDM		ΛCDM		
λ [h^{-1}kpc]	χ^2	p	χ^2	p	$\mathcal{L}\left(\frac{\Lambda\mathrm{CDM}}{\mathrm{SCDM}}\right)$
5	4.87	0.56	2.25	0.90	3.478
7	11.31	0.08	8.97	0.18	2.852
10	7.26	0.30	8.14	0.23	0.633
14	7.34	0.29	8.67	0.19	0.524
20	11.88	0.06	10.40	0.11	2.545

with being drawn from either of the two CDM models, i.e. the genus test cannot rule out the SCDM or ΛCDM model with high significance when only one smoothing scale is considered. However, we can still make a statement about their relative likelihoods given the measured 1.2-Jy data. For the fixed smoothing, the ΛCDM model is more likely than the SCDM model for all smoothing scales. This preference of ΛCDM is also

Table 6.5: Comparison of the SCDM and ΛCDM models with the 1.2-Jy survey using a combination of genus amplitudes for different smoothing scales. Listed are the fractions $p_{\rm Q}$, $p_{\rm R}$ of mock catalogues that fit the ensemble mean worse than the 1.2-Jy measurements. The $p_{\rm Q}$ statistic is just a sum over the χ^2-deviations at individual smoothing scales, and the $p_{\rm R}$ statistic is based on a fit of the run of genus amplitude with smoothing scale.

		SCDM	ΛCDM
fixed	$p_{\rm Q}$	0.012	0.522
	$p_{\rm R}$	0.002	0.696
spherical	$p_{\rm Q}$	0.330	0.590
	$p_{\rm R}$	0.242	0.992
triaxial	$p_{\rm Q}$	0.174	0.384
	$p_{\rm R}$	0.132	0.944

found with the adaptive smoothing, however at a weaker level since the results actually favour SCDM at $10\,h^{-1}$kpc and $14\,h^{-1}$kpc.

To improve the discriminative power of the genus test we can try to combine the measurements corresponding to different smoothing scales. As we have demonstrated above, the genus amplitude (i.e. the first principal component h_1) is most sensitive to differences between the models, essentially because it measures the shape of the power spectrum. Hence we will focus on it in the following.

In Figure 6.21 we show the average genus amplitudes of the SCDM and ΛCDM mock catalogues together with the measurements for the 1.2-Jy survey. The models follow approximately parallel lines, albeit at different heights, when we plot dimensionless genus densities $(2\pi)^2\lambda^3 N$. To quantify the overall agreement between the 1.2-Jy survey and the models we consider two simple statistics. First we compute a χ^2-like quantity by adding up the amplitude measurements corresponding to the different smoothing scales, viz.

$$Q = \sum_i \frac{(A_i^{\rm m} - \overline{A}_i)^2}{\sigma_i^2}, \tag{6.35}$$

where $A_i^{\rm m}$ is the measured 1.2-Jy amplitude at smoothing scale i, and $\overline{A}_i$ and σ_i give the mean and dispersion of one of the two mock ensembles. Because the survey volumes corresponding to different smoothing scales are not independent, the quantity Q cannot be expected to follow a χ^2 distribution. Instead, we calibrate its distribution with the mock ensembles themselves. In particular, we compute the fraction $p_{\rm Q}$ of mock catalogues that give a higher value for Q than 1.2-Jy. This quantity can be interpreted as an exclusion probability.

The above test makes no assumption about the run of genus amplitude with smoothing scale. We can further strengthen the constraints by assuming that the power spectrum of the 1.2-Jy survey is reasonably close to a CDM spectrum. Then the measured genus amplitudes should lead to an approximately straight line as well. Using linear regression we therefore fit a straight line to the measured amplitudes in the $(2\pi)^3\lambda^3 N$–λ diagram and consider the distribution of the amplitude of this fit at the intermediate smoothing scale of $\lambda = 12.5\,h^{-1}$kpc. Again we calibrate this statistic with the mock catalogues. In particular, we compute for the measured 1.2-Jy value the fraction p_{R} of mock catalogues that deviate more from the mean than 1.2-Jy itself.

Table 6.5 summarizes the results of these tests. As already suggested by Figure 6.21 the fixed smoothing technique clearly favours the ΛCDM model, and rules out the SCDM model with high significance. In fact, the combined genus test excludes the SCDM model at a 99 per cent confidence level.

The adaptive smoothing techniques, however, disappoint the hopes for stronger constraints. While they also clearly favour the ΛCDM model, they cannot exclude SCDM with reasonable significance. This is mainly due to 1.2-Jy measurements that are quite high at intermediate smoothing scales and which fit the SCDM model better than ΛCDM in this regime. This may well be a fluctuation due to cosmic variance. We also want to argue that this should not be viewed as a failure of the adaptive smoothing techniques. As we have shown, adaptive smoothing does increase the signal-to-noise ratio of the measured genus curves. Hence it is able to measure more properties of the examined density fields. The results we obtain just mean that these additional properties of the 1.2-Jy density field agree with both models and are not able to discriminate more strongly between them.

In summary the genus statistic of the 1.2-Jy survey is consistent with the ΛCDM model, while the SCDM cosmology is ruled out with high significance. This is in accordance with the expectation that a model with power spectrum shape $\Gamma \approx 0.2$ should do substantially better than SCDM, since the observed galaxy distribution has been repeatedly shown to exhibit more large-scale power than SCDM. Our results are also consistent with the work of Protogeros & Weinberg (1997), who found that the 1.2-Jy topology can be explained by an open CDM cosmology with $\Gamma = 0.25$, although a scale-free $n = -1$ model gave even a somewhat better fit.

6.9 Conclusions

In this work we have examined the genus statistics of N-body simulations down to the smallest scales examined so far. With the conventional smoothing technique of a spatially fixed kernel we showed that the genus curves of CDM cosmologies retain their random phase shape far into the non-linear regime. However, the genus amplitude is strongly reduced by phase correlations in the density field on scales below $10\,h^{-1}$kpc. At $2\,h^{-1}$kpc the suppression reaches a factor of 4. While it is not obvious at the moment how this amplitude drop is related to more traditional measures of higher order correlations, it might be an interesting quantity to characterize non-linearity in future investigations.

The genus in the fixed smoothing regime fails to show strong differences between the τCDM, ΛCDM, and OCDM models. This suggests that the genus is only sensitive to the shape of the power spectrum in this regime.

We have shown that an adaptive smoothing is required to use the smallest resolved mass scales of the Virgo simulations. Because of that we have computed genus curves with a novel adaptive smoothing technique for the fully sampled simulations. With the adaptive scheme we can clearly separate the four models at the smallest scales we examined, i.e. the 'degeneracy' between the three models τCDM, ΛCDM, and OCDM can be lifted. In addition, on these scales the genus statistics show very strong departures from 'quasi-Gaussian' behaviour.

We have also performed a large Monte-Carlo experiment in order to establish the statistical properties of the genus statistic when it is applied to a redshift survey like the *IRAS* 1.2-Jy catalogue. For this purpose we extracted a large number of 1.2-Jy mock catalogues from the simulations.

We found that the genus statistic of the 1.2-Jy survey is well consistent with the ΛCDM simulation, while the SCDM model is ruled out at a 99 per cent confidence level. We have not explicitly examined τCDM and OCDM, since we expect them to show at most marginal differences from ΛCDM at the resolution of the 1.2-Jy data.

In this work we also proposed two variants of adaptive smoothing techniques for flux limited redshift surveys. We demonstrated that these techniques can improve the signal-to-noise ratio of the measured genus curves. Hence they are able to extract more topological information from a given redshift survey. Using the 1.2-Jy catalogue, however, we could not achieve a stronger discrimination between the SCDM and ΛCDM models. Since adaptive smoothing is sensitive to additional properties of the density field, we conclude that these properties of the 1.2-Jy survey are consistent with both models. We remain convinced that adaptive smoothing will exhibit a clear advantage, if redshift surveys are used that allow a reconstruction of the density field on strongly clustered scales. Since very large redshift surveys like the Sloan survey are underway, this regime will be accessible in the next few years.

Due to its sensitivity to higher order correlations, the genus statistic remains a useful tool to test the random phase hypothesis, and to compare theoretical models with observations. In this work we confirm that the topology of the 1.2-Jy redshift survey is consistent with current models of cold dark matter universes that grow structure out of random phase initial conditions. A particular advantage of the genus is that these results should be largely independent of a possible bias between the galaxy density and the mass density, at least if this bias relation is monotonic.

6.10 Appendix: Distribution of errors

In Section 6.7.1 we defined a multivariate analysis of the measured genus curves using the full 61×61 covariance matrix of equation (6.26). Here we demonstrate that this method can fail in practice, and we examine reasons for this failure.

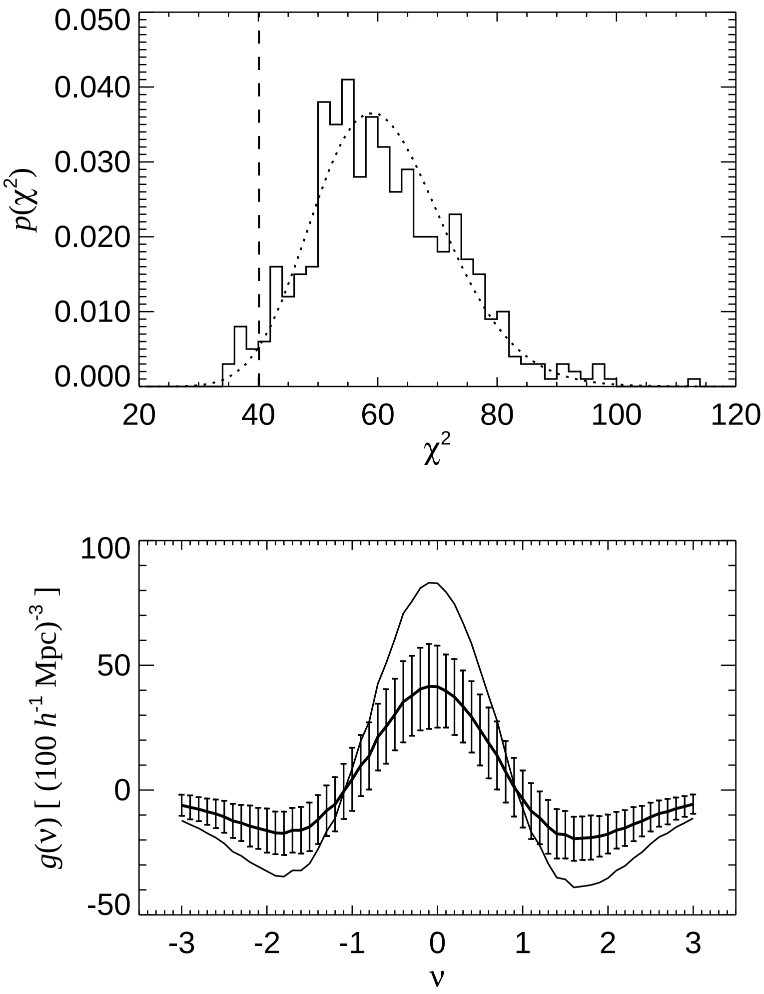


Figure 6.22: The top panel shows the distribution of χ^2 defined by equation (6.28) for $k = 61$ measured points of the genus curve (SCDM mock catalogues, fixed smoothing of $7\,h^{-1}$kpc). The dotted curve is the theoretical χ^2 distribution for 61 degrees of freedom. The dashed line marks the χ^2-value obtained if one checks the fit of a genus curve that is everywhere twice the mean, i.e. $\mathbf{g} = 2\overline{\mathbf{g}}$. The resulting low value of χ^2 suggests a perfect fit; yet this curve is highly inconsistent with the mock ensemble, as we demonstrate in the bottom panel. Here the thick line gives $\overline{\mathbf{g}}$, while the thin line shows $\mathbf{g} = 2\overline{\mathbf{g}}$. The error bars are the rms deviations of the genus at individual points ν_i of the curve.

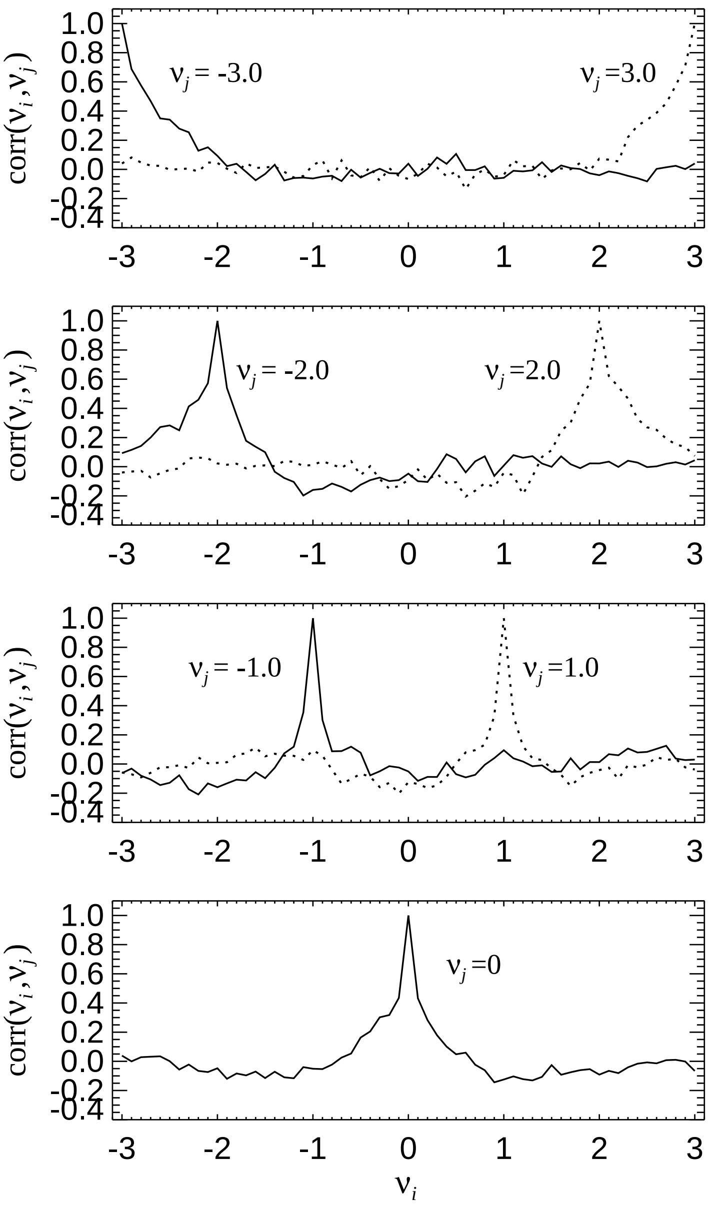


Figure 6.23: Pearson's correlation coefficient for the genus measurement. The labelled curves show the correlation coefficient for a number of different points on the genus curve. Adjacent points are quite strongly correlated over a range $\Delta\nu \approx 1$. The result shown here is for the SCDM catalogues with fixed smoothing of $5\,h^{-1}$kpc; it is very similar for all other smoothing scales we consider in this work.

In Figure 6.22 we show the distribution of χ^2 as defined in equation (6.28) for the SCDM mock catalogues, smoothed with $\lambda = 7\,h^{-1}$kpc. The dotted line shows a χ^2 distribution for $k = 61$ degrees of freedom, which is apparently able to fit the observed distribution reasonably well. However, this does not necessarily mean that the distribution of errors is in fact well approximated by a multivariate Gaussian. A first hint that a problem is lurking here may be obtained by computing the χ^2 value for a genus curve with $\mathbf{g} = 2\overline{\mathbf{g}}$, i.e. one that deviates by 100 per cent from the mean. For this curve the χ^2 comes out as 40.1, suggesting a perfect fit, although the curve $\mathbf{g}$ is actually discrepant at individual points on the genus curve with high significance level, as is shown in the bottom panel of Figure 6.22.

Clearly, this peculiar result needs to be understood. In Figure 6.23 we show Pearson's correlation coefficient for a number of places on the genus curve. In general, adjacent points on the genus curve are correlated over roughly a range $\Delta\nu \approx 1$. The main effect of these correlations is to introduce negative values in the inverse V_{ij}^{-1} of the covariance matrix in the elements just off the diagonal. These terms reduce the weight of deviations from the mean if adjacent points exhibit deviations of the same sign.

As a result, coherent deviations (as in $\mathbf{g} = 2\overline{\mathbf{g}}$) are hardly penalized at all. This partly explains the low χ^2 we obtained for the test curve $\mathbf{g} = 2\overline{\mathbf{g}}$; yet it does not fully account for why the test fails so badly. We think that there are two reasons for this. First, the statistical properties of the genus errors cannot be fully described by the covariance matrix alone, because the distribution of errors is not consistent with a multivariate Gaussian. This will be shown below. Second, and more importantly, the noise in the covariance matrix compromises its inversion; although the inversion is mathematically possible and stable, the result is not necessarily meaningful, because it is strongly affected by the noise. The principal components analysis developed in Section 6.7.2 provides a solution to this problem. The noise may also be tamed by using a smaller number of points to sample the genus curves, or by using a very large ensemble of mock catalogues.

We now demonstrate that the distribution of the genus measurements is in fact not a multivariate Gaussian. It is sufficient to show that even the distribution of just the genus at a single value of ν_i is not normal. For this purpose we employ a Kolmogorov-Smirnov (KS) test. Based on the $n = 500$ genus measurements for a mock ensemble we can estimate the mean and the variance at individual points on the genus curve from the sample itself. We can then evaluate the KS test

$$D_n = \sup\left|S_n(g) - F(g)\right|, \tag{6.36}$$

where $S_n(g)$ denotes the cumulative distribution function of the measurements and $F(g)$ is the cumulative probability distribution function of the presumed Gaussian.

In general the KS test is fully distribution-free only if the test distribution F is known beforehand. However, here we estimate the parameters of the normal distribution from the sample itself. Because mean and variance are only scale factors of the normal distribution, the KS test remains applicable, although the distribution of D_n is changed (Kendell & Stuart 1973). We calibrate the latter with a Monte-Carlo experiment.

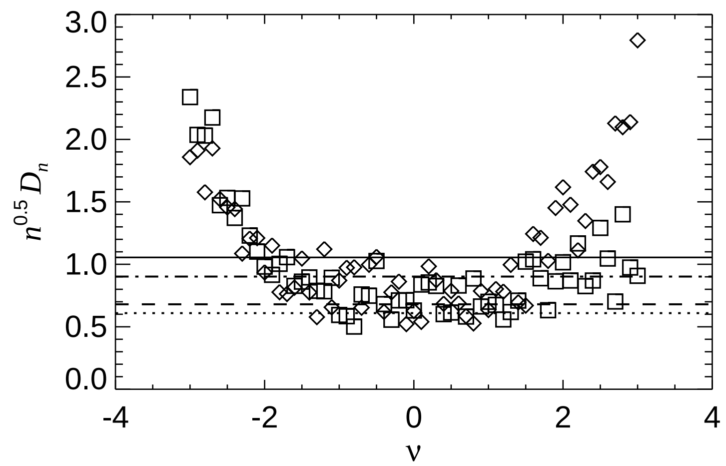


Figure 6.24: Kolmogorov-Smirnov test of normality of the genus measurements. The symbols give the test measure $\sqrt{n}D_n$ for 61 points on the genus curve of the SCDM mock sample. Diamonds refer to fixed smoothing with $5\,h^{-1}$kpc, boxes are for adaptive smoothing. The horizontal lines delimit probability regions for various exclusion levels. If the distribution were normal, a point should be found with probability 50% below the lowest line, with 68% below the dashed line, with 95% below the dot-dashed line, and with 99% below the solid line. We obtain a qualitatively similar result for other smoothing scales and mock ensembles.

In Figure 6.24 we show measurements of $\sqrt{n}D_n$ for the SCDM mock ensemble, smoothed at $5\,h^{-1}$kpc with the spherical and, alternatively, the triaxially adaptive method. If the distribution of the g_i were Gaussian, in half of the cases we should measure $\sqrt{n}D_n > 0.609$, and in only 1 per cent of the cases $\sqrt{n}D_n > 1.055$. Hence the KS test shows that the genus distribution is inconsistent with a Gaussian at low and high values of ν. Close to $\nu = 0$ the points are consistent with a normal distribution, although there seems to be a lack of points with low values of $\sqrt{n}D_n$.

We think that it is not too surprising that the distribution of errors is not well described by a multivariate Gaussian. For example, sampling fluctuations can affect the whole volume-fraction/density-threshold relation and thus result in coherent shifts of parts of the genus curve along the x-axis. Similarly, the small survey volumes examined here can lead to large irregular fluctuations in the genus due to particular density features of the patch under examination. As a result, we expect that the genus curve exhibits complicated higher order correlations which make it rather difficult to take full advantage of its information content.

–Diesem von Reichtum, Macht und Senilität, von Gleichgültigkeit, Puritanismus und geistiger Hygiene, von Armut und Verschwendung, von technologischer Eitelkeit und nutzloser Gewalttätigkeit total verdorbenem Universum muß ich unweigerlich den Hoffnungsschimmer des neuen Tages zuerkennen. Vielleicht wird weiterhin die ganze Welt von ihm schwärmen, während sie von ihm beherrscht und ausgebeutet wird.

Jean Baudrillard, Amerika

7 Concluding remarks

Abstract

In each of the preceding chapters, I have given a summary and a discussion of the main results found in the corresponding sections. I will here briefly try to reiterate some of my findings and to put them into perspective within cosmology.

7.1 Results and conclusions

In this thesis, I have studied a number of related aspects of the theory of structure formation occurring via gravitational instability of Gaussian fluctuations in universes dominated by cold dark matter. This class of theoretical models currently provides the most successful description of structure formation, and it is also the most widely investigated up to now.

An important paradigm of these theoretical models is the concept of hierarchical galaxy formation. In CDM universes, galaxies form bottom-up, with small galaxies being born first. A hierarchy of merging processes then leads to the build-up of ever more massive galaxies. In these scenarios, mergers of galaxies are common events that occur frequently. In fact, according to the merger hypothesis, most ellipticals are thought to be formed by mergers of large disk galaxies.

In this work, I have addressed two main issues related to galaxy mergers. In Chapter 2, I have focused on the formation of tidal tails in collisions of disk galaxies. In previous work, it had been suggested that the massive dark haloes expected in CDM models around disk galaxies could prevent them from forming tails as massive and extended as those observed in many prominent interacting systems. However, with a series of self-consistent collisionless simulations of galaxy-galaxy mergers I have shown that the halo-to-disk mass ratio is not a good indicator of the ability to produce tails. Instead, the relative size of disk and halo, or alternatively, the ratio of circular velocity to local escape speed at the half mass radius of the disk decide whether strong tails can be formed. Using a detailed modeling of the structure of galaxies expected in the currently favoured CDM models, I have shown that these models are not compromised by an inability to form tails, because there is a sufficiently large population of disk galaxies that is susceptible to tail formation. This conclusion is almost independent of the cosmological parameters employed for the CDM model. Tidal tails thus provide no strong constraints on them, but they remain a highly useful probe for the structure of individual interacting systems.

A second issue related to galaxy mergers was addressed in Chapter 3. I here developed a hydrodynamical model for the regulation of star formation in disk galaxies by feedback processes. Because the complicated physics of feedback is neither understood in detail nor can it directly be simulated, I have adopted a heuristic approach and introduced a fiducial second reservoir of internal energy, accounting for the kinetic energy content of the gas on unresolved scales. Feedback by supernova explosions is then thought to give rise to an effective turbulent pressure described by this reservoir.

Using this model, I have shown that a self-regulation of star formation can be achieved, with star formation rates in agreement with the observed global Schmidt-law for disk galaxies. The scheme allows converged numerical results to be reached with moderate numerical resolution, which is of high relevance for future applications to cosmological simulations of galaxy formation. I have also applied the model to mergers of gas-rich disk galaxies. These dissipative galaxy collisions lead to a strong central influx of gas, where it feeds a central starburst. The presence of a bulge can stabilize the disk such that the onset of the starburst is delayed from the first encounter of the galaxies to their final merger. By studying the isophotal shapes of the merger remnants and comparing them to equivalent collisionless merger remnants, I showed that dissipation favours disky isophotal shapes. This supports a long-standing hypothesis raised in many observational studies of the formation path of elliptical galaxies. In general, the density profiles of the merger remnants are well consistent with de Vaucouleurs profiles, except for the innermost region in the strongly dissipative simulations. Here, the newly created stars give rise to a luminous core with luminosity density that may be higher than those observed in the centers of ordinary ellipticals.

Hierarchical galaxy formation is a process that continuously acts over the course of cosmic evolution. In Chapter 4, I have therefore addressed the problem of galaxy formation within the proper cosmological background using simulations of the formation of a rich cluster of galaxies. These simulations have presumably the largest numerical resolution ever achieved in this field so far. The large resolution allowed an extension of semi-analytic models for galaxy formation to a regime where one can actually track the merging processes of dark matter subhaloes within larger virialized systems. I developed new techniques to find these subhaloes, and to trace their evolution from simulation output to output. In this new methodology I have been able to directly follow the merging processes of galaxies, thereby eliminating uncertainties due to the previously required estimates of merger timescales. I have shown that the resulting semi-analytic models produce remarkably successful results for the population of cluster galaxies. The luminosity function of cluster galaxies is very well fit by a Schechter function with a relatively flat faint-end, and the overall mass-to-light ratio is consistent with observations. Simultaneously, the Tully-Fisher relation of spiral galaxies in the field is very well fit. Within the cluster, elliptical galaxies are more concentrated towards the center than spirals, i.e. a pronounced density-morphology relation results, spatially resolved within the cluster. The star formation histories of cluster and field galaxies reveal that the cluster galaxies form their stars substantially earlier than the galaxies in the field. Already at redshift $z \simeq 4$ half of the stars in the cluster have been formed, while this is not the case for the field before $z \simeq 2$.

In order to study resolution effects in my analysis, I have simulated the same cluster of galaxies several times using different numerical resolution. Reassuringly, the number of subhaloes I find is consistent between simulations of different numerical resolution. Each of the simulations exhibits the right abundance of subhaloes down to its resolution limit. Note that the total mass attached to subhaloes in the cluster is about 10-11%.

The formalism I developed also allows the detection of mergers between galaxies inside the cluster. Such events do indeed occur, albeit at a small rate. Note that despite the survival of hundreds to thousands of subhaloes in the final cluster, there is still plenty of merging going on. In the semi-analytic modeling bulges of galaxies form in mergers of galaxies while most stars are born in disks. However, the majority of stars in the cluster are found in bulges, showing that their parent galaxies have undergone mergers since most of the stars had been born.

A common theme through the work discussed above has been the extensive use of computer simulations. These simulations have been carried out with a new N-body code that combines a gravitational tree-algorithm with smoothed particle hydrodynamics, as discussed in Chapter 5. The code uses a new time integration scheme allowing individual, fully adaptive timesteps for all particles. With respect to the tree algorithm, I have introduced a new cell-opening criterion, which is more efficient than the standard BH-criterion, and has also much better properties when applied to cosmological simulations at high redshift. I have also introduced a new concept of dynamical tree updates, eliminating a possible bottleneck caused by the tree construction time in a regime where the dynamic range in particle timesteps becomes very large. In the hydrodynamical part, the code keeps the number of SPH smoothing neighbours exactly constant, and it guarantees force symmetry between all hydrodynamically interacting particle pairs. Together with a sheer-reduced formulation of the artificial viscosity, the hydrodynamical part performs very well in test problems, in particular with respect to observing the conservation laws. The code is essentially free in its ability to adapt to the geometry and the dynamic range of the problem under consideration, and it only invests computational work where it is needed.

I have also ported the code to massively parallel supercomputers with distributed memory using an explicit communication paradigm. In principle, gravitational tree algorithms are very difficult to parallelize on distributed machines, because at any given time, every particle interacts with every other particle. Despite this problem, it is possible to design a decomposition of the problem, such that high degrees of parallelization and resulting speed-up factors can be achieved. I have used a recursive orthogonal domain decomposition to partition the problem onto the processors, where the domain decomposition is dynamically determined by work-load balance considerations. Each domain then acts as an independent computational entity that can provide the force exerted by its mass content on arbitrary locations in space. Using a collective communication and summation scheme, I then obtain the total force on the subset of particles that is advanced at the current timestep. This parallelization strategy allows to fully benefit from the substantial gain in performance resulting from the use of adaptive timesteps. The parallel code developed in this work scales well to high processor numbers and has been successfully used to run several cosmological simulations with up to 74 million particles on 512 processors. It is fully written in the C/MPI language standard, allowing its use on a large variety of systems, without requiring any substantial change.

One of the fundamental assumptions of current theories of galaxy formation is whether the primordial fluctuations that seed structure form a Gaussian random field. It is thus an important task to test observationally whether this Gaussian random phase hypothesis holds, or whether phase correlations in the primordial density field existed. In Chapter 6, I have used the genus statistic to study the density field of the Local Universe, as observed with the 1.2-Jy redshift survey of *IRAS* galaxies. Unlike the basic statistical measurements of the power spectrum or the two-point correlation function, the genus statistic is sensitive to correlations of higher order, and it can be used to test whether the primordial density field is consistent with being a Gaussian random field.

In order to compare N-body models for the growth of structure with the 1.2-Jy survey, I have extracted large ensembles of mock galaxy catalogues from the simulations. These mock surveys, which mimic the selection properties of the 1.2-Jy survey, have then be used to find the distribution of errors of the measured genus curves, and to establish a principal components analysis of these curves. Using a likelihood ratio test and a combination of genus measurements at different smoothing scales, I found that the 1.2-Jy survey is inconsistent with the SCDM model at a 99% confidence level, while CDM models with a shape parameter $\Gamma \simeq 0.2$ for their power spectrum are in good agreement with the data. On linear scales, the *IRAS* density field is consistent with the Gaussian random field hypothesis. I also introduced a new adaptive smoothing technique, and I showed that it increases the signal-to-noise ratio of the genus test. Adaptive smoothing can thus be used to extract a larger amount of information from a given data set. For the numerical N-body models, I have analysed the genus also in the strongly non-linear regime, where large phase correlations due to gravitational clustering develop. The amplitude drop of the genus curve with respect to the Gaussian value provides an interesting quantitative measure of phase correlations and may in principle be used to differentiate between models.

7.2 Theoretical prospects

If one compares the current status of the field of galaxy formation with the one just a few years ago, one is certainly impressed by the very rapid progress that has been achieved in this area. Over the last decade or so, the paradigm of hierarchical galaxy formation has emerged as a highly successful theory, able to provide explanations for a large variety of observational facts. Similarly, for a suitable choice of the cosmological parameters, the cold dark matter models are in reasonable agreement with all the basic aspects of observational data. At present, there is no alternative theory that comes even close to these successes, so one might be tempted to start talking of something like a 'standard model' for cosmic structure formation.

However, galaxy formation is still far from being 'solved', and it is too early to consider hierarchical models as being proven. While the general picture of hierarchical galaxy formation seems to work, this picture is still very crude in many respects. For example, the details of star formation in galaxies or the chemical enrichment of the interstellar and intergalactic media are only poorly understood. There are also many well-established observational facts that are as of yet unexplained, or it is unsettled which one out of several competing scenarios offers the correct explanation. For example, the tilt of the fundamental plane of elliptical galaxies, or the physical origin of gamma ray bursts, may be considered to fall into this category.

At a more fundamental level, we should not forget that cosmological parameters as basic as the mass density of the Universe, or the Hubble parameter are still not known precisely. On top of that lingers one of the most enigmatic questions of cosmology: What is the physical nature of the dark matter?

However, cosmology will undoubtedly see substantial further progress in the next few years. It can be expected that new observational data will be the catalyst allowing such advances. In recent years, the exploration of the high-redshift Universe has become possible with the *Hubble Space Telescope*, and the new huge 8-10 meter class telescopes like *VLT* and *Keck* will continue to rapidly advance our understanding of the birth of galaxies. Numerous new satellite missions will be launched in the near future. Projects like *Planck-Surveyor* and *MAP* will provide measurements of the cosmic microwave background with very large angular resolution. If successful, these missions will provide precision measurements of the cosmological parameters. A host of new X-ray satellites like *XMM* or *Chandra* will follow up on the successful exploration of the X-ray sky. New huge redshift surveys like the *Sloan* or *2dF* surveys are about to provide databases of unseen quality and extent.

The sheer amount of cosmological data that will be provided by these projects is awe inspiring, and it will likely transform our view of galaxy formation by constraining many basic properties of the galaxy population to unprecedented accuracy. As a result, cosmology is probably about to enter the 'precision era'. Theoretical models that try to keep up with the ever more detailed observational picture of the Universe will likely require ever more complex modeling, too. Therefore it appears that numerical astrophysics might become even more important than it is today.

References

Abel, T., Anninos, P., Zhang, Y., Norman, M. L. (1997), *Modeling primordial gas in numerical cosmology*, New Astronomy, 2, 181

Abel, T., Bryan, G. L., Norman, M. L. (1998), *Formation and fragmentation of primordial molecular clouds*, in *Evolution of Large-Scale Structure: From Recombination to Garching*, preprint, astro-ph/9810215

Athanassoula, E., Bosma, A., Lambert, J. C., Makino, J. (1997), *Performance and accuracy of a GRAPE-3 system for collisionless N-body simulations*, preprint, astro-ph/9709246

Balsara, D. W. (1995), *von Neumann stability analysis of smoothed particle hydrodynamics – suggestions for optimal algorithms*, J. Comp. Phys., 121, 357

Barnes, J., Hut, P. (1986), *A hierarchical $\mathcal{O}(N \log N)$ force-calculation algorithm*, Nature, 324, 446

Barnes, J. E. (1988), *Encounters of disk/halo galaxies*, ApJ, 331, 699

Barnes, J. E. (1989), *Evolution of compact groups and the formation of elliptical galaxies*, Nature, 338, 123

Barnes, J. E. (1992), *Transformations of galaxies. I. Mergers of equal-mass stellar disks*, ApJ, 393, 484

Barnes, J. E., Hernquist, L. (1996), *Transformations of galaxies. II. Gasdynamics in merging disk galaxies*, ApJ, 471, 115

Barnes, J. E., Hernquist, L. E. (1991), *Fueling starburst galaxies with gas-rich mergers*, ApJ, 370, L65

Baugh, C. M., Cole, S., Frenk, C. S. (1996a), *Evolution of the Hubble sequence in hierarchical models for galaxy formation*, MNRAS, 283, 1361

Baugh, C. M., Cole, S., Frenk, C. S. (1996b), *Faint galaxy counts as a function of morphological type in a hierarchical merger model*, MNRAS, 282, L27

Baugh, C. M., Cole, S., Frenk, C. S., Lacey, C. G. (1998), *The epoch of galaxy formation*, ApJ, 498, 504

Bekki, K., Shioya, Y. (1997), *Formation of boxy and disky elliptical galaxies in early dissipative mergers*, ApJ, 478, L17

Bender, R., Burstein, D., Faber, S. M. (1992), *Dynamically hot galaxies. I - Structural properties*, ApJ, 399, 462

Bender, R., Burstein, D., Faber, S. M. (1993), *Dynamically hot galaxies. II - Global stellar populations*, ApJ, 411, 153

Bender, R., Döbereiner, S., Möllenhoff, C. (1987), *Radio activity and the shape of elliptical galaxies*, A&A, 177, L53

Bender, R., Möllenhoff, C. (1987), *Morphological analysis of massive early-type galaxies in the Virgo Cluster*, A&A, 177, 71

Benson, A. J., Cole, S., Frenk, C. S., Baugh, C. M., Lacey, C. G. (1999), *The nature of galaxy bias and clustering*, preprint, astro-ph/9903343

Binney, J., Tremaine, S. (1987), *Galactic Dynamics*, Princeton University Press

Blain, A. W., Smail, I., Ivison, R. J., Kneip, J.-P. (1999), *The history of star formation in dusty galaxies*, MNRAS, 302, 632

Bloemen, J. B. G. M. (1987), *On stable hydrostatic equilibrium configurations of the galaxy and implications for its halo*, ApJ, 322, 694

Bond, J. R., Kaiser, N., Cole, S., Efstathiou, G. (1991), *Excursion set mass functions for hierarchical Gaussian fluctuations*, ApJ, 379, 440

Bruzual, A. G., Charlot, S. (1993), *Spectral evolution of stellar populations using isochrone synthesis*, ApJ, 405, 538

Byun, Y.-I., Grillmair, C. J., Faber, S. M., Ajhar, E. A., Dressler, A., Kormendy, J., Lauer, T. R., Richstone, D., Tremaine, S. (1996), *The centers of early-type galaxies with HST. II. Empirical models and structural parameters*, AJ, 111, 1889

Canavezes, A., Springel, V., Oliver, S. J., Rowan-Robinson, M., Keeble, O., White, S. D. M., Saunders, W., Efstathiou, G., Frenk, C., McMahon, R. G., Maddox, S., Sutherland, W., Tadros, H. (1998), *The topology of the IRAS Point Source Catalogue Redshift Survey*, MNRAS, 297, 777

Carraro, G., Lia, C., Chiosi, C. (1998), *Galaxy formation and evolution - I. The Padua TREE-SPH code (PD-SPH)*, MNRAS, 297, 1021

Cole, S. (1991), *Modeling galaxy formation in evolving dark matter halos*, ApJ, 367, 45

Cole, S., Aragon-Salamanca, A., Frenk, C. S., Navarro, J. F., Zepf, S. E. (1994), *A recipe for galaxy formation*, MNRAS, 271, 781

Colina, L., Lipari, S., Macchetto, F. (1991), *Massive star formation and superwinds in IRAS 19254 - 7245 (the 'Superantennae')*, ApJ, 379, 113

Colley, W. N. (1997), *Two-dimensional topology of large-scale structure in the Las Campanas Redshift Survey*, ApJ, 489, 471

Colley, W. N., Gott, J. R., Park, C. (1996), *Topology of COBE microwave background fluctuations*, MNRAS, 281, L82

Copi, C. J., Schramm, D. N., Turner, M. S. (1995), *Big-bang nucleosynthesis and the baryon density of the Universe*, Science, 267, 192

Couchman, H. M. P. (1991), *Mesh-refined P^3M: A fast adaptive N-body algorithm*, ApJ, 368, 23

Couchman, H. M. P., Thomas, P., Pearce, F. (1995), *Hydra: Adaptive-mesh implementation of P^3M-SPH*, ApJ, 452, 797

Davé, R., Dubinski, J., Hernquist, L. (1997), *Parallel TreeSPH*, New Astronomy, 2, 277

de Zeeuw, T., Franx, M. (1991), *Structure and dynamics of elliptical galaxies*, Ann. Rev. Astron. Astrophys., 39, 239

Debattista, V. P., Sellwood, J. A. (1998), *Dynamical friction and the dstribution of dark matter in barred galaxies*, ApJ, 493, L5

Diaferio, A., Kauffmann, G., Colberg, J. M., White, S. D. M. (1999), *Clustering of galaxies in a hierarchical universe - III. Evolution to high redshift*, preprint, astro-ph/9812009

Dikaiakos, M. D., Stadel, J. (1995), *PKDGRAV*, in preparation

Dressler, A. (1980), *Galaxy morphology in rich clusters - Implications for the formation and evolution of galaxies*, ApJ, 236, 351

Dubinski, J. (1996), *A parallel tree code*, New Astronomy, 1, 133

Dubinski, J., Hernquist, L., Mihos, J. C. (1997), *Constraining dark halos with tidal tails*, preprint, astro-ph/9712110

Dubinski, J., Mihos, J. C., Hernquist, L. (1996), *Using tidal tails to probe dark matter halos*, ApJ, 462, 576

Dubinski, J., Mihos, J. C., Hernquist, L. (1999), *Constraining dark halo potentials with tidal tails*, preprint, astro-ph/9902217

Efstathiou, G., Bond, J. R., White, S. D. M. (1992), *COBE background radiation anisotropies and large-scale structure in the Universe*, MNRAS, 258, 1

Efstathiou, G., Davis, M., Frenk, C. S., White, S. D. M. (1985), *Numerical techniques for large cosmological N-body simulations*, ApJS, 57, 241

Efstathiou, G., Lake, G., Negroponte, J. (1982), *The stability and masses of disc galaxies*, MNRAS, 199, 1069

Eisenstein, D. J., Hut, P. (1998), *HOP: A new group-finding algorithm for N-body simulations*, ApJ, 498, 137

Evrard, A. E. (1988), *Beyond N-body: 3D cosmological gas dynamics*, MNRAS, 235, 911

Farouki, R. T., Shapiro, S. L. (1982), *Simulations of merging disk galaxies*, ApJ, 259, 103

Farouki, R. T., Shapiro, S. L., Duncan, M. J. (1983), *Hierarchical merging and the structure of elliptical galaxies*, ApJ, 265, 597

Fisher, K. B., Huchra, J. P., Strauss, M. A., Davis, M., Yahil, A., Schlegel, D. (1995), *The IRAS 1.2 Jy Survey: Redshift data*, ApJS, 100, 69

Frenk, C. S., White, S. D. M., Bode, P., Bond, J. R., Bryan, G. L., Cen, R., Couchman, H. M. P., Evrard, A. E., Gnedin, N., Jenkins, A., Khokhlov, A. M., Klypin, A., Navarro, J. F., Norman, M. L., Ostriker, J. P., Owen, J. M., Pearce, F. R., Pen, U.-L., Steinmetz, M., Thomas, P. A., Villumsen, J. V., Wadsley, J. W., Warren, M. S., Xu, G., Yepes, G. (1999), *The Santa Barbara cluster comparison projekt*, preprint, astro-ph/9906160

Gebhardt, K., Richstone, D., Ajhar, E. A., Lauer, T. R., Byun, Y., Kormendy, J., Dressler, A., Faber, S. M., Grillmair, C., Tremaine, S. (1996), *The centers of early-type galaxies with HST. III. Non-parametric recovery of stellar luminosity distributions*, AJ, 112, 105

Geiger, B., Schneider, P. (1998), *Constraining the mass distribution of cluster galaxies by weak lensing*, MNRAS, 295, 497

Geiger, B., Schneider, P. (1999), *A simultaneous maximum likelihood approach for galaxy–galaxy lensing and cluster lens reconstruction*, MNRAS, 302, 119

Gelb, J. M., Bertschinger, E. (1994), *Cold dark matter. I. The formation of dark halos*, ApJ, 436, 467

Gerritsen, J. P. E., Icke, V. (1997), *Star formation in N-body simulations. I. The impact of the stellar ultraviolet radiation on star formation*, A&A, 325, 972

Ghigna, S., Moore, B., Governato, F., Lake, G., Quinn, T., Stadel, J. (1998), *Dark matter haloes within clusters*, MNRAS, 300, 146

Gingold, R. A., Monaghan, J. J. (1977), *Smoothed particle hydrodynamics – Theory and application to non-spherical stars*, MNRAS, 181, 375

Giovanelli, R., Haynes, M. P., da Costa, L. N., Freudling, W., Salzer, J. J., Wegner, G. (1997), *The Tully-Fisher relation and H_0*, ApJ, 477, 1

Gott, J. R., Dickinson, M., Melott, A. L. (1986), *The sponge-like topology of large-scale structure in the universe*, ApJ, 306, 341

Gott, J. R., Miller, J., Thuan, T. X., Schneider, S. E., Weinberg, D. H., Gammie, C., Polk, K., Vogeley, M., Jeffrey, S., Bhavsar, S. P., Melott, A. L., Giovanelli, R., Haynes, M. P., Tully, R. B., Hamilton, A. J. S. (1989), *The topology of large-scale structure. III - Analysis of observations*, ApJ, 340, 625

Gottlöber, S., Klypin, A. A., Kravtsov, A. V. (1998), *Evolution of isolated halos and halos inside groups and clusters in a LCDM model*, in *Evolution of large-scale structure: From recombination to Garching*, preprint, astro-ph/9810191

Götz, M., Huchra, J. P., Brandenberger, R. H. (1998), *Group identification in N-body simulations: SKID and DENMAX versus friends-of-friends*, preprint, astro-ph/9811393

Governato, F., Baugh, C. M., Frenk, C. S., Cole, S., Lacey, C. G., Quinn, T., Stadel, J. (1998), *The seeds of rich galaxy clusters in the Universe*, Nature, 392, 359

Governato, F., Reduzzi, L., Rampazzo, R. (1993), *Can isophotal shape discriminate between possible origins of elliptical galaxies?*, MNRAS, 261, 379

Groom, W. (1997), *The formation and evolution of galaxies in a cold dark matter universe*, Ph.D. thesis, University of Cambridge

Guth, A. H. (1981), *Inflationary universe: A possible solution to the horizon and flatness problems*, Phys. Rev. D, 23, 347

Hamilton, A. J. S., Gott, J. R., Weinberg, D. (1986), *The topology of the large-scale structure of the universe*, ApJ, 309, 1

Heller, C. H., Shlosman, I. (1994), *Fueling nuclear activity in disk galaxies: Starbursts and monsters*, ApJ, 424, 84

Hernquist, L. (1992), *Struture of merger remnants. I. Bulgeless progenitors*, ApJ, 400, 460

Hernquist, L. (1993a), *N-body realizations of compound galaxies*, ApJS, 86, 389

Hernquist, L. (1993b), *Structure of merger remnants. II. Progenitors with rotating bulges*, ApJ, 409, 548

Hernquist, L., Bouchet, F. R., Suto, Y. (1991), *Application of the Ewald method to cosmological N-body simulations*, ApJS, 75, 231

Hernquist, L., Katz, N. (1989), *TREESPH: A unification of SPH with the hierarchical TREE method*, ApJ, 70, 419

Hernquist, L., Spergel, D. N., Heyl, J. S. (1993), *Structure of merger remnants. III. Phase-space constraints*, ApJ, 416, 415

Heyl, J. S., Cole, S., Frenk, C. S., Navarro, J. F. (1995), *Galaxy formation in a variety of hierarchical models*, MNRAS, 274, 755

Heyl, J. S., Hernquist, L., Spergel, D. N. (1994), *Structure of merger remnants. IV. Isophotal shapes*, ApJ, 427, 165

Hibbard, J. E., Mihos, J. C. (1995), *Dynamical modeling of NGC 7252 and the return of tidal material*, AJ, 110, 140

Hibbard, J. E., Yun, M. S. (1999), *Luminosity profiles of merger remnants*, ApJ, 522, L93

Hiotelis, N., Voglis, N. (1991), *Smooth particle hydrodynamics with locally readjustable resolution in the collapse of a gaseous protogalaxy*, A&A, 243, 333

Hultman, J., Källander, D. (1997), *An SPH code for galaxy formation problems*, A&A, 324, 534

Hunter, C. (1972), *Self-gravitating gaseous disks*, Ann. Rev. Fluid Mech., 4, 219

Jenkins, A. R., Frenk, C. S., Pearce, F. R., Thomas, P. A., Colberg, J. M., White, S. D. M., Couchman, H. M. P., Peacock, J. A., Efstathiou, G. P., Nelson, A. H. (1998), *Evolution of structure in cold dark matter universes*, ApJ, 499, 20

Jernigan, J. G., Porter, D. H. (1989), *A tree code with logarithmic reduction of force terms, hierarchical regularization of all variables, and explicit accuracy controls*, ApJS, 71, 871

Kalberla, P. M. W., Kerp, J. (1998), *Hydrostatic equlibrium conditions in the galactic halo*, A&A, 339, 745

Kang, H., Ostriker, J. P., Cen, R., Ryu, D., Hernquist, L., Evrard, A. E., Bryan, G. L., Norman, M. L. (1994), *A comparison of cosmological hydrodynamic codes*, ApJ, 430, 83

Katz, N. (1992), *Dissipational galaxy formation. II. Effects of star formation*, ApJ, 391, 502

Katz, N., Weinberg, D. H., Hernquist, L. (1996), *Cosmological simulations with TreeSPH*, ApJS, 105, 19

Kauffmann, G. (1995a), *Hierarchical clustering and the Butcher-Oemler effect*, MNRAS, 274, 153

Kauffmann, G. (1995b), *The observed properties of high-redshift cluster galaxies*, MNRAS, 274, 161

Kauffmann, G. (1996a), *The age of elliptical galaxies and bulges in a merger model*, MNRAS, 281, 487

Kauffmann, G. (1996b), *Disc galaxies at $z = 0$ and at high redshift: an explanation of the observed evolution of damped Ly α absorption systems*, MNRAS, 281, 475

Kauffmann, G., Charlot, S. (1998), *The K-band luminosity function at $z = 1$: a powerful constraint on galaxy formation theory*, MNRAS, 297, L23

Kauffmann, G., Colberg, J. M., Diaferio, A., White, S. D. M. (1999a), *Clustering of galaxies in a hierarchical universe - I. Methods and results at $z = 0$*, MNRAS, 303, 188

Kauffmann, G., Colberg, J. M., Diaferio, A., White, S. D. M. (1999b), *Clustering of galaxies in a hierarchical universe - II. Evolution to high redshift*, preprint, astro-ph/9809168

Kauffmann, G., Guiderdoni, B., White, S. D. M. (1994), *Faint galaxy counts in a hierarchical universe*, MNRAS, 267, 981

Kauffmann, G., Nusser, A., Steinmetz, M. (1997), *Galaxy formation and large-scale bias*, MNRAS, 286, 795

Kauffmann, G., White, S. D. M. (1993), *The merging history of dark matter haloes in a hierarchical universe*, MNRAS, 261, 921

Kauffmann, G., White, S. D. M., Guiderdoni, B. (1993), *The formation and evolution of galaxies within merging dark matter haloes*, MNRAS, 264, 201

Kendell, M. G., Stuart, A. (1973), *The advanced theory of statistics*, volume 2, Hafner, New York, third edition

Kennicutt, R. C. (1983), *The rate of star formation in normal disk galaxies*, ApJ, 272, 54

Kennicutt, R. C. (1989), *The star formation law in galactic disks*, ApJ, 344, 685

Kennicutt, R. C. (1998), *The global Schmidt law in star-forming galaxies*, ApJ, 498, 541

Kent, S. M., Gunn, J. E. (1982), *The dynamics of rich clusters of galaxies. I - The Coma cluster*, AJ, 87, 945

Kent, S. M., Gunn, J. E. (1983), *The dynamics of rich clusters of galaxies. II - The Perseus cluster*, AJ, 88, 697

Kersher, M., Schmalzing, J., Buchert, T., Wagner, H. (1998), *Fluctuations in the IRAS 1.2 Jy catalogue*, A&A, 333, 1

Klein, R. I., Fisher, R. T., McKee, C. F., Truelove, J. K. (1998), *Gravitational collapse and fragmentation in molecular clouds with adaptive mesh refinement hydrodynamics*, preprint, astro-ph/9806330

Klypin, A. A. (1988), *Percolation as a descriptive statistics – prospects and methods*, Acta Cosmologica, 15, 101

Klypin, A. A., Gottlöber, S., Kravtsov, A. V., Khokhlov, A. M. (1999), *Galaxies in N-body simulations: Overcoming the overmerging problem*, ApJ, 516, 530

Kormendy, J. (1989), *Did elliptical galaxies form by mergers or by dissipative collapse?*, ApJ, 342, L63

Kormendy, J., Bender, R. (1996), *A proposed revision of the Hubble sequence for elliptical galaxies*, ApJ, 464, L119

Lacey, C., Cole, S. (1993), *Merger rates in hierarchical models of galaxy formation*, MNRAS, 262, 627

Lacey, C., Guiderdoni, B., Rocca-Vomerange, B., Silk, J. (1993), *Tidally triggered galaxy formation. II - Galaxy number counts*, ApJ, 402, 15

Lacey, C., Silk, J. (1991), *Tidally triggered galaxy formation. I - Evolution of the galaxy luminosity function*, ApJ, 381, 14

Laughlin, G., Korchagin, V., Adams, F. C. (1998), *The dynamics of heavy gaseous disks*, ApJ, 504, 945

Lemson, G., Kauffmann, G. (1999), *Environmental influences on dark matter haloes and consequences for the galaxies within them*, MNRAS, 302, 111

Lima-Neto, G. B., Combes, F. (1995), *Isophote shapes of merger remnants*, A&A, 294, 657

Lockman, F. J., Gehman, C. S. (1991), *Vertical distribution and support of galactic HI*, ApJ, 382, 182

Lombardi, J. C., Sills, A., Rasio, F. A., Shapiro, S. L. (1998), *Tests of spurious transport in smoothed particle hydrodynamics*, preprint, astro-ph/9807290

Lucy, L. B. (1977), *A numerical approach to the testing of the fission hypothesis*, AJ, 82, 1013

Luo, S., Vishniac, E. (1995), *Three-dimensional shape statistics: Methodology*, ApJS, 96, 429

Mac Low, M.-M., Ferrara, A. (1999), *Starburst-driven mass loss from dwarf galaxies: Efficiency and metal ejection*, ApJ, 513, 142

Mac Low, M.-M., Klessen, R. S., Burkert, A. (1998), *Kinetic energy decay rates of supersonic and super-Alfvenic turbulence in star-forming clouds*, Phys. Rev. Letters, 80, 2754

MacFarland, T., Couchman, H. M. P., Pearce, F. R., Pichlmeier, J. (1998), *A new parallel P^3M-code for very large-scale cosmological simulations*, New Astronomy, 3, 687

Madau, P., Pozzetti, L., Dickinson, M. (1998), *The star formation history of field galaxies*, ApJ, 498, 106

Magorrian, J., Binney, J. (1994), *Predicting line-of-sight velocity distributions of elliptical galaxies*, MNRAS, 271, 949

Makino, J., Taiji, M., Ebisuzaki, T., Sugimoto, D. (1997), *GRAPE-4: A massively parallel special-purpose computer for collisional N-body simulations*, ApJ, 480, 432

Matsubara, T. (1996), *Statistics of isodensity contours in redshift space*, ApJ, 457, 13

Matsubara, T., Suto, Y. (1996), *Nonlinear evolution of genus in a primordial random Gaussian density field*, ApJ, 460, 51

Matsubara, T., Yokoyama, J. (1996), *Genus statistics of the large-scale structure with non-Gaussian density fields*, ApJ, 463, 409

McKee, C. F., Ostriker, J. P. (1977), *A theory of the interstellar medium - Three components regulated by supernova explosions in an inhomogeneous substrate*, ApJ, 218, 148

Mecke, K. R., Buchert, T., Wagner, H. (1994), *Robust morphological measures for large-scale structure in the Universe*, A&A, 288, 697

Melott, A. L., Dominik, K. G. (1993), *Tests of smoothing methods for topological study of galaxy redshift surveys*, ApJS, 86, 1

Melott, A. L., Weinberg, D. H., Gott, J. R. (1988), *The topology of large-scale structure. II - Nonlinear evolution of Gaussian models*, ApJ, 328, 50

Mihos, J. C., Bothun, G. D. (1997), *NGC 2442: Tidal encounters and the evolution of spiral galaxies*, ApJ, 481, 741

Mihos, J. C., Dubinski, J., Hernquist, L. (1998), *Tidal tales two: The effect of dark matter halos on tidal tail morphology and kinematics*, ApJ, 494, 183

Mihos, J. C., Hernquist, L. (1994a), *Dense stellar cores in merger remnants*, ApJ, 437, L47

Mihos, J. C., Hernquist, L. (1994b), *Star forming galaxy models: Blending star formation into TreeSPH*, ApJ, 437, 611

Mihos, J. C., Hernquist, L. (1994c), *Ultraluminous starbursts in major mergers*, ApJ, 431, L9

Mihos, J. C., Hernquist, L. (1996), *Gasdynamics and starbursts in major mergers*, ApJ, 464, 641

Mo, H. J., Mao, S., White, S. D. M. (1998), *The formation of galactic disks*, MNRAS, 295, 319

Monaghan, J. J. (1992), *Smoothed particle hydrodynamics*, Ann. Rev. Astron. Astrophys., 30, 543

Monaghan, J. J., Gingold, R. A. (1983), *Shock simulation by the particle method SPH*, J. Comp. Phys., 52, 374

Monaghan, J. J., Lattanzio, J. C. (1985), *A refined particle method for astrophysical problems*, A&A, 149, 135

Moore, B., Katz, N., Lake, G., Dressler, A., Oemler, A. (1996), *Galaxy harassment and the evolution of clusters of galaxies*, Nature, 379, 613

Moore, B., Lake, G., Katz, N. (1998), *Morphological transformation from galaxy harassment*, ApJ, 495, 139

Navarro, J. F. (1998), *The cosmological significance of disk galaxy rotation curves*, preprint, astro-ph/9807084

Navarro, J. F., Frenk, C. S., White, S. D. M. (1995), *The assembly of galaxies in a hierarchically clustering universe*, MNRAS, 275, 56

Navarro, J. F., Frenk, C. S., White, S. D. M. (1996), *The structure of cold dark matter halos*, ApJ, 462, 563

Navarro, J. F., Frenk, C. S., White, S. D. M. (1997), *A universal density profile from hierarchical clustering*, ApJ, 490, 493

Navarro, J. F., Steinmetz, M. (1997), *The effects of a photoionizing ultraviolet background on the formation of disk galaxies*, ApJ, 478, 13

Navarro, J. F., White, S. D. M. (1993), *Simulations of dissipative galaxy formation in hierachically clustering universes - I. Tests of the code*, MNRAS, 265, 271

Navarro, J. F., White, S. D. M. (1994), *Simulations of dissipative galaxy formation in hierarchically clustering universes - II. Dynamics of the baryonic component in galactic haloes*, MNRAS, 267, 401

Negroponte, J., White, S. D. M. (1983), *Simulations of mergers between disc-halo galaxies*, MNRAS, 205, 1009

Nelson, R. P., Papaloizou, J. C. B. (1994), *Variable smoothing lengths and energy conservation in smoothed particle hydrodynamics*, MNRAS, 270, 1

Norman, C. A., Ferrara, A. (1996), *The turbulent interstellar medium: Generalizing to a scale-dependent phase continuum*, ApJ, 467, 280

Norman, M. L., Bryan, G. L. (1998), *Cosmological adaptive mesh refinement*, preprint, astro-ph/9807121

Ostriker, J. P., Peebles, P. J. E. (1973), *A numerical study of the stability of flattened galaxies: Or, can cold galaxies survive?*, ApJ, 186, 467

Pacheco, P. S. (1997), *Parallel Programming with MPI*, Morgan Kaufmann Publishers, San Francisco

Padmanabhan, T. (1993), *Structure formation in the universe*, Cambridge University Press

Park, C., Gott, J. R., da Costa, L. N. (1992a), *Large-scale structure in the Southern Sky Redshift Survey*, ApJ, 392, L51

Park, C., Gott, J. R., Melott, A. L., Karachentsev, I. D. (1992b), *The topology of large-scale structure. VI – Slices of the universe*, ApJ, 387, 1

Peacock, J. A. (1999), *Cosmological physics*, Cambridge University Press

Pearce, F. R., Couchman, H. M. P. (1997), *Hydra: A parallel adaptive grid code*, New Astronomy, 2, 411

Pearce, F. R., Jenkins, A., Colberg, J. M., White, S. D. M., Thomas, P. A., Couchman, H. M. P., Peacock, J. A., Efstathiou, G. (1999), *A simulation of galaxy formation and clustering*, preprint, astro-ph/9905160

Pearson, R. C., Coles, P. (1995), *Quantifying the geometry of large-scale structure*, MNRAS, 272, 231

Peebles, P. J. E. (1993), *Principles of physical cosmology*, Princeton University Press

Press, W. H., Schechter, P. (1974), *Formation of galaxies and clusters of galaxies by self-similar gravitational condensation*, ApJ, 187, 425

Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P. (1995), *Numerical recipies in C*, Cambridge University Press, Cambridge

Protogeros, Z. A. M., Weinberg, D. H. (1997), *The topology of large-scale structure in the 1.2 Jy IRAS redshift survey*, ApJ, 489, 457

Quillen, A. C., Sarajedini, V. L. (1998), *Galaxies with spiral structure up to $z \sim 0.87$: Limits on M/L and the stellar velocity dispersion*, AJ, 115, 1412

Quinn, T., Katz, N., Stadel, J., Lake, G. (1997), *Time stepping N-body simulations*, preprint, astro-ph/9710043

Rees, M. J., Ostriker, J. P. (1977), *Cooling, dynamics and fragmentation of massive gas clouds - Clues to the masses and radii of galaxies and clusters*, MNRAS, 179, 541

Rhoads, J. E., Gott, J. R., Postman, M. (1994), *The genus curve of the Abell clusters*, ApJ, 421, 1

Rosen, A., Bregman, J. N. (1995), *Global properties of the interstellar medium in disk galaxies*, ApJ, 440, 634

Roukema, B. F., Peterson, B. A., Quinn, P. J., Rocca-Volmerange, B. (1997), *Merging history trees of dark matter halos: a tool for exploring galaxy formation models*, MNRAS, 292, 835

Ryden, B. S. (1988), *The area of isodensity contours as a measure of large-scale structure*, ApJ, 333, L41

Salmon, J. K., Warren, M. S. (1994), *Skeletons from the treecode closet*, J. Comp. Phys., 111, 136

Scalo, J. M. (1986), *The stellar initial mass function*, Fundamentals of Cosmic Physics, 11, 1

Shu, F. H., Adams, F. C., Lizano, S. (1987), *Star formation in molecular clouds: Observation and theory*, Ann. Rev. Astron. Astrophys., 25, 23

Silk, J. (1997), *Feedback, disk self-regulation, and galaxy formation*, ApJ, 481, 703

Simien, F., de Vaucouleurs, G. (1986), *Systematics of bulge-to-disk ratios*, ApJ, 302, 564

Snir, M., Otto, S. W., Huss-Lederman, S., Walker, D. W., Dongarra, J. (1995), *MPI: The complete reference*, MIT Press, Cambridge

Somerville, R. S., Kolatt, T. S. (1999), *How to plant a merger tree*, MNRAS, 305, 1

Somerville, R. S., Primack, J. R. (1998), *Semi-analytic modelling of galaxy formation: The Local Universe*, preprint, astro-ph/9802268

Springel, V., White, S. D. M. (1998), *Estimates for the luminosity function of galaxies and its evolution*, MNRAS, 298, 143

Springel, V., White, S. D. M. (1999), *Tidal tails in CDM cosmologies*, MNRAS, 307, 162

Springel, V., White, S. D. M., Colberg, J. M., Couchman, H. M. P., Efstathiou, G. P., Frenk, C. S., ans F. R. Pearce, A. R. J., Nelson, A. H., Peacock, J. A., Thomas, P. A. (1998), *Genus statistics of the Virgo N-body simulations and the 1.2-Jy redshift survey*, MNRAS, 298, 1169

Spurzem, R. (1997), *Astrophysical N-body simulations: Algorithms and challenges*, preprint, astro-ph/9711238

Steinmetz, M. (1996), *GRAPESPH: cosmological smoothed particle hydrodynamics simulations with the special-purpose hardware GRAPE*, MNRAS, 278, 1005

Steinmetz, M., Müller, E. (1993), *On the capabilities and limits of smoothed particle hydrodynamics*, A&A, 268, 391

Steinmetz, M., Müller, E. (1995), *The formation of disc galaxies in a cosmological context: structure and kinematics*, MNRAS, 276, 549

Strauss, M. A., Davis, M., Yahil, A., Huchra, J. P. (1990), *A redshift survey of IRAS galaxies. I – Sample selection*, ApJ, 361, 49

Sutherland, R. S., Dopita, M. A. (1993), *Low-density astrophsyical plasmas*, ApJS, 88, 253

Syer, D., Mao, S., Mo, H. (1997), *Global stability and the mass-to-light ratio of galactic disks*, preprint, astro-ph/9711160

Thacker, R. J., Tittley, E. R., Pearce, F. R., Couchman, H. M. P., Thomas, P. A. (1998), *Smoothed particle hydrodynamics in cosmology: a comparative study of implementations*, preprint, astro-ph/9809221

Theuns, T., Rathsack, M. E. (1993), *Calculating short range forces on a massively parallel computer: SPH on the Connection Machine*, Comp. Phys. Comm., 76, 141

Thomas, P. A., Colberg, J. M., Couchman, H. M. P., Efstathiou, G. P., Frenk, C. S., Jenkins, A. R., Nelson, A. H., Hutchings, R. M., Peacock, J. A., Pearce, F. R., White, S. D. M. (1997), *The structure of galaxy clusters in various cosmologies*, MNRAS, 296, 1061

Thomas, P. A., Couchman, H. M. P. (1992), *Simulating the formation of a cluster of galaxies*, MNRAS, 257, 11

Tissera, P. B., Lambas, D. G., Abadi, M. G. (1997), *Analysis of galaxy formation with hydrodynamics*, MNRAS, 286, 384

Toomre, A., Toomre, J. (1972), *Galactic bridges and tails*, ApJ, 178, 623

Tormen, G. (1997), *The rise and fall of satellites in galaxy clusters*, MNRAS, 290, 411

Tormen, G., Bouchet, F. R., White, S. D. M. (1997), *The structure and dynamical evolution of dark matter halos*, MNRAS, 286, 865

Tormen, G., Diaferio, A., Syer, D. (1998), *Survival of substructure within dark matter halos*, MNRAS, 299, 728

van den Bosch, F. C., Lewis, G. F., Lake, G., Stadel, J. (1999), *Substructure in dark halos: Orbital eccentricities and dynamical friction*, ApJ, 515, 50

van der Kruit, P. C. (1995), *Dark matter in spiral galaxies*, in *Stellar Populations*, edited by P. C. van der Kruit, G. Gilmore, 205, in IAU Symp. 164, Kluwer

van Kampen, E., Jiminez, R., Peacock, J. A. (1999), *Overmerging and M/L ratios in phenomenological galaxy formation models*, preprint, astro-ph/9904274

Vogeley, M. S., Park, C., Geller, M. J., Huchra, J. P., Gott, J. R. (1994), *Topological analysis of the CfA redshift survey*, ApJ, 420, 525

Wada, K., Norman, C. A. (1999), *The global structure of a self-gravitating multi-phase interstellar medium in a galactic disk*, ApJ, 516, L13

Wang, B., Silk, J. (1994), *Gravitational instability and disk star formation*, ApJ, 427, 759

Warren, M. S., Quinn, P. J., Salmon, J. K., Zurek, W. H. (1992), *Dark halos formed via dissipationless collapse. I - Shapes and alignment of angular momentum*, ApJ, 399, 405

Weil, M. L., Eke, V. R., Efstathiou, G. (1998), *The formation of disc galaxies*, MNRAS, 300, 773

Weinberg, D. H. (1988), *Contour – A topological analysis program*, PASP, 100, 1373

Weinberg, D. H., Gott, J. R., Melott, A. L. (1987), *The topology of large-scale structure. I – Topology and the random phase hypothesis*, ApJ, 321, 2

Weinberg, D. H., Hernquist, L., Katz, N. (1997), *Photoionization, numerical resolution, and galaxy formation*, ApJ, 477, 8

Weinberg, S. (1972), *Gravitation and cosmology*, Wiley, New York

White, S. D. M. (1978), *Simulations of merging galaxies*, MNRAS, 184, 185

White, S. D. M. (1979), *Further simulations of merging galaxies*, MNRAS, 189, 831

White, S. D. M., Davis, M., Efstathiou, G., Frenk, C. S. (1987), *Galaxy distribution in a cold dark matter universe*, Nature, 330, 451

White, S. D. M., Frenk, C. S. (1991), *Galaxy formation through hierarchical clustering*, ApJ, 379, 52

White, S. D. M., Navarro, J., Evrard, A. E., Frenk, C. S. (1993), *The baryon content of galaxy clusters: a challenge to cosmological orthodoxy*, Nature, 366, 429

White, S. D. M., Rees, M. J. (1978), *Core condensation in heavy halos – A two-stage theory for galaxy formation and clustering*, MNRAS, 183, 341

Whitmore, B. C., Gilmore, D. M., Jones, C. (1993), *What determines the morphological fractions in clusters of galaxies?*, ApJ, 407, 489

Wirth, N. (1986), *Algorithms and data structures*, Prentice-Hall, London

Wyse, R. F. G., Silk, J. (1989), *Star formation rates and abundance gradients in disk galaxies*, ApJ, 339, 700

Yepes, G., Kates, R., Khokhlov, A., Klypin, A. (1997), *Hydrodynamical simulations of galaxy formation: effects of supernova feedback*, MNRAS, 284, 235

Danksagung

Ein großes Dankeschön gilt meinem Betreuer, Prof. Simon White, für seine hilfreiche Unterstützung während meiner ganzen Zeit am MPA. Simon White hat stets versucht, mir den Blick für das wesentliche zu öffnen. Seine Anregungen und seine Kritik waren immer überaus präzise und haben entscheidend zum Gelingen dieser Arbeit beigetragen. Simon White hat mir auch eine Unzahl toller Kontakte eröffnet und mich bei der Suche nach einer wissenschaftlichen Anstellung hervorragend unterstützt und beraten.

Ganz besonders möchte ich meinen Bürokollegen Jörg, Martin und Georg danken, die mir die jahrelange Zeit im stets abgedunkelten '221' erst erträglich gemacht haben. Kaum ein Tag verging ohne politische Diskussionen oder vernichtende Filmkritiken und nirgendwo sonst gab es solch ein Fülle neuester, sensationeller Erkenntnisse aus der Erforschung des Frau-Mann Verhältnisses. Jungs, ich werde unsere Streifzüge durchs Internet vermissen und die Lachsalven über alle möglichen absurden Facetten dieser Welt.

Und da gibt's noch viele andere im Institut, die gute Freunde geworden sind und meine Zeit in München zu einem unvergeßlichen Erlebnis gemacht haben. Jan, Gerti, Conrad, Guido "Professor!" Kruse. Ich werde all die Nächte im Atomic Cafe, im Kunstpark Ost, das viele Bier, die Taxifahrten quer durch die Stadt, nie vergessen. Unzählige bizarre Dinge haben sich ereignet, und einige davon sind längst zu legendären Geschichten geworden.

Meine WG in der Schleißheimer Straße war Erlebnis und Schicksal zugleich. Ort der Ruhe, des Chaos, immer mal wieder 'Party-Base' und natürlich Drehort der Daily-Soap 'Gute Zeiten, Schlechte Zeiten'. Steffi, Karin und Alexandra haben stets Nachsicht mit meinen erratischen Putz-Gewohnheiten und meinen überall rumstehenden, halbvollen Kaffee-Tassen gehabt und auch diverse spontane Feste am frühen morgen stets verziehen. Danke für all die Abwechslung, das Lachen, und die Gespräche. Ohne meine Mitbewohnerinnen hätte ich womöglich vergessen, daß es noch ein Leben außerhalb des Instituts gibt.

Liebe Sonja, ohne Dich wäre mir in der letzten Phase dieser Arbeit wohl irgendwann der Mut ausgegangen. Dein riesiges Verständnis und Deine Geduld mit meinen Marotten sind echt bewundernswert. Das war unheimlich lieb von Dir, Sonja.

Ganz besonders danken möchte ich auch meinen Eltern und meinen Brüdern. Sie haben mich immer rückhaltlos unterstützt, mir Sicherheit gegeben, und mir ins Gewissen geredet, wenn mich die Arbeit zu sehr absorbierte. Ohne die leckeren Essenspakete meiner Mutter wäre ich wohl völlig abgemagert, und ohne die finanziellen Kenntnisse

meines Vaters hätte ich noch nicht mal meine Steuererklärung ausfüllen können... :-)

Vielen meiner altgedienten Freunde möchte ich auch herzlich danken. Nur allzuoft habe ich sie sehr vernachlässigt und viele Versprechungen aus schlimmem Zeitmangel heraus nicht eingelöst. Trotzdem hat mich keiner von ihnen fallengelassen.

Ich möchte auch allen Mitarbeitern des MPA ganz herzlich danken, die immer für optimale Arbeitsmöglichkeiten gesorgt haben. Mein spezieller Dank gilt 'unseren' Sekretärinnen Cornelia Rickl und Maria Depner, sowie unseren System-Administratoren, die ein hervorragend funktionierendes Computersystem bereitgestellt haben. Auch dem Rechenzentrum Garching gilt mein Dank, speziell Gabriele Bachmeister und Ingeborg Weidl, die sich um den Superrechner T3E ausgezeichnet gekümmert haben.

Lebenslauf

Persönliche Daten

	Volker Springel geboren am 18. November 1970 in Backnang, ledig

Schulausbildung

Aug. 1977 – Juni 1981	Grundschule in der Taus, Backnang
Aug. 1981 – Juni 1990	Gymnasium in der Taus, Backnang, Abitur

Zivildienst

Aug. 1990 – Okt. 1991	Arbeiterwohlfahrt, Backnang

Studium

Okt. 1991 – Juli 1994	Studium der Physik, Universität Tübingen
Aug. 1994 – Mai 1995	graduate studies in physics, University of California at Berkeley
Juni 1995 – Dez. 1995	Studium der Physik, Universität Tübingen
Jan. 1996 – Okt. 1996	Diplomarbeit am Max-Planck-Institut für Astrophysik, Garching. Thema: "Topology and Luminosity Function of the PSCz Redshift Survey"
3. Oktober 1996	Diplom in Physik, Eberhard-Karls-Universität Tübingen
Okt. 1996 - Herbst 1999	Promotionsstipendium am Max-Planck-Institut für Astrophysik, Garching

Auszeichnungen

Juli 1989	XX. Internationale Physikolympiade, Bronze-Medaille
Mai 1990	2. Bundessieger beim Wettbewerb *Jugend forscht*
Juli 1990	XXI. Internationale Physikolympiade, Silber-Medaille
Mai 1989 – Sep. 1996	Stipendium *Studienstiftung des deutschen Volkes*