Predictability, Infrared Galaxies and Weak Lensing Contributions to the Semi-Analytic Approach to Galaxy Formation

Jaime Ernesto Forero-Romero

N^o d'ordre: 427
N^o attribué par la bibliothèque: 07ENSL0 427

THÈSE

en vue d'obtenir le grade de

Docteur de l'Université de Lyon -
École Normale Supérieure de Lyon
Spécialité: Physique
Laboratoire: C. R. A. L.
École doctorale de Physique et Astrophysique

présentée et soutenue publiquement le 30/11/07

par Monsieur Jaime Ernesto FORERO ROMERO

Titre:

PREDICTABILITÉ, GALAXIES INFRAROUGES ET LENTILLES GRAVITATIONNELLES: APPLICATIONS DE L'APPROCHE HYBRIDE

Directeur de Thèse:
Bruno GUIDERDONI

Aprés avis de:
Monsieur Stefan GOTTLÖBER, Membre/Rapporteur
Monsieur Mika JUVELA, Membre/Rapporteur

Devant la Commission d'examen formée de:

Monsieur, Stéphane COLOMBI, Membre
Monsieur, Stefan GOTTLÖBER, Membre/Rapporteur
Monsieur, Bruno GUIDERDONI, Membre
Monsieur, Mika JUVELA, Membre/Rapporteur
Madame, Guilaine LAGACHE, Membre

Contributions to the Semi-Analytic Approach to Galaxy Formation

Predictability, Infrared Galaxies and Weak Lensing

Jaime Ernesto FORERO ROMERO

para Gilma y para Olga

Contents

List of Illustrations

Acknowledgments

It's been a long way until this PhD Thesis has been finished and written. I'm counting, of course, since the beginning of my undergraduate studies. I'm grateful to a lot of people. The easiest way to acknowledge them all is in reverse, systematical, chronological order.

`(2005-2007)@Lyon` Thanks above all to Bruno for all the freedom and support. To Hélène for making my life in Saint Genis Laval a lot easier. To Julien and Jérémy for all the patience and scientific support. To Dylan and Thierry for making the office a lively place. To all the cosmology team. To CNP for the great movie selection, and to the bouchon *Le Galyon* for the great food at very affordable prices.

`(2003-2005)@Paris` To Anna, Simone and Ernesto for his patience and good company. Without them the first months in Paris would have been impossible to live. To Ernesto, again, for introducing me to Olga. To Effrosyni, Annie, and Sara for making the IAP a better place to live and work. To Patrick Boissé for introducing me to Bruno. To Bruno and Ludovic for great support since then. To Andrea for making such a good EARA 2003 meeting, it's his fault that I am involved now in galaxy formation. To the crêperie *Au p'tit grec* for the best crêpes in the world.

`(2001-2003)@Bariloche`. To Kimura, Roberto, Víctor and Ariel for growing up (!?) with me. To Mara, for her support in critical times. To César, hoping we can be friends again. To Ezequiel for all the blank sheets of paper he lent me when I didn't have some. To Víctor (again), Gladys and Paco, without them I could not be accepted to the ENS in Paris, literally. To Luis Masperi for his work at CLAF and his follow up until the latest days. To Alberto for the meat and the pasta.

`(1999-2001)@Bogotá` To Luisa, Viviana, Aura, Omar and Doris for their friendship. To Benjamin for the opportunity to be part of the OAN in spite of me being so young. To Balaguera, Gio and Pilar for countless *peches*. To

Srta. Murcia for always being present at the right moments. To Wilson for keeping up alive an important part of my life during those times. To David for many initial lessons in friendship, bad taste and good humor. To Juan Roberto and John for setting up high standards in all the things they did, in academics as well in personal life.

`(2002-)@online` To maestrozen, bluelephant, clarousse, catalombia, avn, cavorite, salidaenfalso, crazyrobot, bluedevil, angelripper and many others for being a good source of discussions and support in one way or another.

Always and everywhere: thanks to my family back in Colombia (mami, papi, mali, Juan y Tilin) for being a continuous support during all these times.

To Olga for being the main driver of my life, this thesis truly could not have been possible without her.

Yo creo en muchas cosas que no he visto, y ustedes también, lo sé.
No se puede negar la existencia de algo palpado por mas etereo que sea.
No hace falta exhibir una prueba de decencia de aquello que es tan verdadero.
El unico gesto es creer o no. Algunas veces hasta creer llorando.
Se trata de un tema incompleto porque le falta respuesta;
respuesta que alguno de ustedes, quizas, le pueda dar.
Es un tema en technicolor para hacer algo util del amor.
Para todos nosotros, amén.

Willie Colón, *Oh que será*

1
INTRODUCTION

Take a good telescope, go far away from the city, and point it to the sky. After some time of this routine you will be looking at a kind of diffuse clouds which seem to be far way. With a little help from your friends you will see that they are receding from us. Will you find anything else as interesting as this with your telescope? Leaving our galaxy aside, maybe not. Maybe that is why galaxies, as we call those clouds. are still –to our understanding– the most basic description of structure in the Universe.

During the last century of pointing out our telescopes to the sky to observe galaxies, we have also made them bigger, sensitive to radiation we cannot see with our eyes and even sent them to space. At the same time we have tried, of course, to make sense of all these observations. To do that we have applied our best theory of gravitation (thoroughly tested in our solar system only) on the entire Universe. We have also translated the laws of physics into iterative processes. Created pieces of data to represent physical entities, using them to create Universes at will

As a result of all that (observations, theory and numerical simulations) we are confident enough to make a guess about the ingredients of our Universe. About five percent is composed by the matter we know from everyday experience. Another 25 percent should be some kind of matter that only feels gravity, and the rest is some kind of homogeneous background energy that is pushing things apart. Thanks to these advances the struggle to understand how the full variety of structure in the Universe came to being still continues, with a central role played by phenomenological (also called semi-analytic) models of galaxy formation.

In this chapter we will describe very shortly all the observational and theoretical evidence that have built the need for the semi-analytic approach to galaxy formation. From that point I will setup the tree different axes of this thesis: predictability, infrared galaxies and weak lensing.

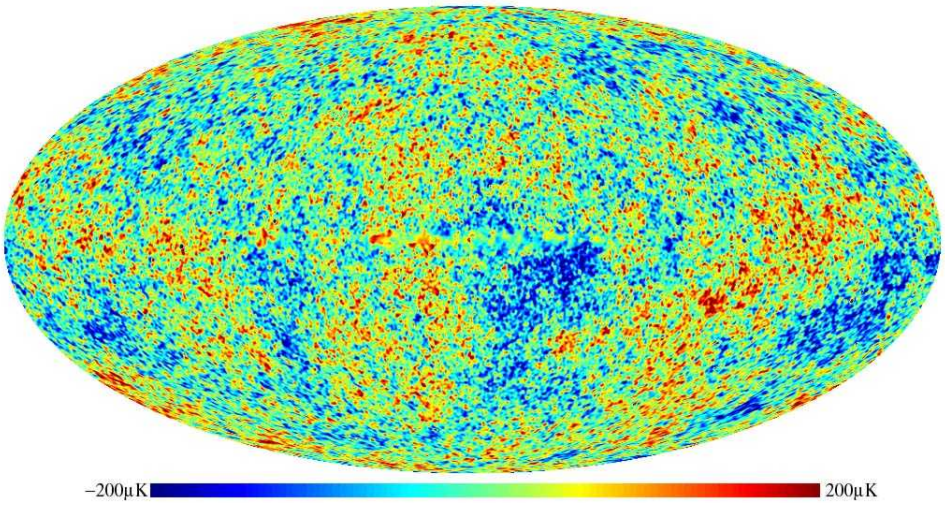


Fig. 1.1. Full sky image of the Cosmic Microwave Background (CMB) as seen by the Wilkinson Microwave Anisotropy Probe (WMAP). The image is the result of heavy post-treatment that removed the mean value and dipole of the CMB itself and all the foregrounds, most notably the Milky Way's radio emission. The color code have been chosen to highlight the inhomogeneities of a few μK (see the color bar at the bottom).

1.1 The Observational Input

The last decade have seen the emergence of a clear cosmological context into which extragalactic astronomy is developed. This was allowed both by the observational breakthroughs but also to the upswing of numerical techniques that allowed the astrophysicists to perform numerical experiments and test their theories.

Observationally there are three major breakthroughs that have shaped our modern understanding of structure formation in the universe: the observations of the Cosmic Microwave Background, the large surveys of galaxy populations and the detection of the weak gravitational lensing.

The **Cosmic Microwave Background** (CMB) is a background that is homogeneous (up to one part in one million) in all directions and its energy as a function of wavelengths fits nicely into a description by a black body spectrum. The discovery of the CMB in 1963 provided strong evidence towards the picture of a Universe that was much hotter and denser in the past than it is now. In other words it was fundamental support to the Hot Big Bang Cosmology. Cosmologist realized that over this smooth background, the initial fluctuations in the Universe, responsible for the structure that we see today, must have been imprinted. It was only in 1992 that the results

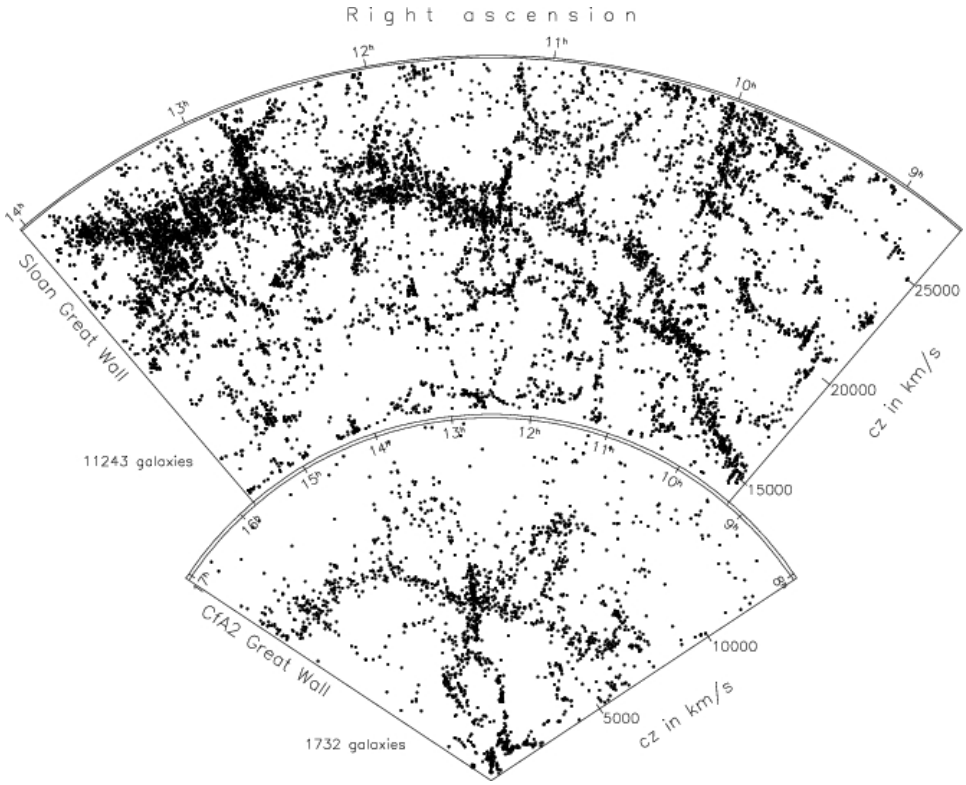


Fig. 1.2. Galaxies in the local Universe for two of the most influential galaxy surveys Center for Astrophysics (CfA, Harvard) Survey and the Sloan Digital Sky Survey (SDSS). The total number of galaxies included in each sample is also indicated. Figure taken from [GJS+05]

from COBE discovered these tiny fluctuations of order 10^{-5} of the CMB temperature [SBK+92].

The analysis of these anisotropies have yielded the most solid constraints we have over cosmological parameters. Nevertheless, there is much more information the remains to be extracted from the CMB, especially in its polarization, which could confirm the nature of the initial density fluctuations.

Galaxies surveys reached during the last decade a peak in performance. They covered wide portions of the sky up to distances of a few thousand billions of light years, while others went very deep, showing the properties of galaxies at very large distances showing galaxies when the universe had only tenth of its age. For wideness, we think especially of the Sloan Digital Sky Survey (SDSS) [YAA+00] and the Two Degrees Field (2dF) [FRP+99], for deepness we think of the Hubble Space Telescope (HST) [ATS+96].

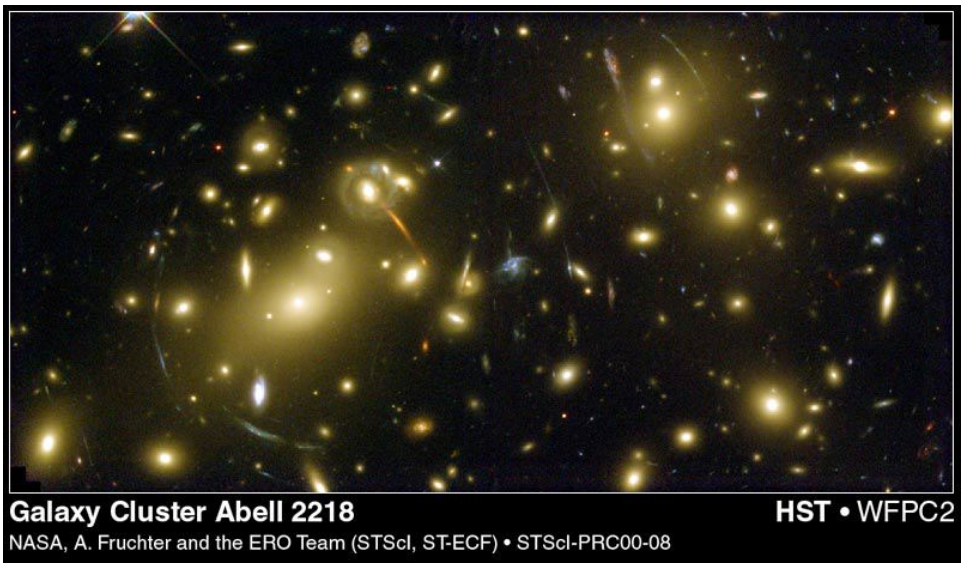


Fig. 1.3. Image taken with the Hubble Space Telescope. It shows a cluster of galaxies acting as a lens for smaller galaxies in the background, which are the light source of the arcs seen around the brightest galaxies in the cluster.

From wide surveys we can have an idea of the large-scale structure in the universe up to the scales were in fact the distribution starts to be homogeneous and isotropic (Fig.1.2); from the deep surveys we have the proof that galaxies have suffered strong evolution, merging, changing their morphology and color, interacting with their environment. We have enough evidence to think from a hierarchical view, where interactions between galaxies play an important role on its evolution [KWG93].

All these observations represent a zoo asking to be understood, but first of all, classified. This has been the work of astronomers during the last decades. Dividing galaxies by mass, luminosity, color. Watching how they evolve through the history of the universe. Observed from the optical to infrared wavelengths, also in radio and now in energetic γ-rays or cosmic rays.

Matter deflects light. In 1919 when deflection of the star light by the Sun was measured, Einstein's General Theory of Relativity was confirmed. Gravitational lensing, as this effect is called, applied to cosmological scales is today a highly valuable tool to measure the properties of matter distribution in the Universe [VMR$^+$01].

Conceptually, gravitational lensing is an effect composed by a source emitting light which will be later deflected by matter. Its detection not only gives

information about the masses of the intervening "lenses" but also on the geometrical configuration of the system source-lens-observer – provided that we have some information on the light source.

The ability to infer properties of the cosmological matter field (lenses) and the geometry of the universe (distances) has put weak gravitational lensing at the forefront of observational tools able to give us a better understanding of our Universe. At the same time, as observations grow in complexity and precision, and we have additional information about the general composition and geometry of the Universe, the interpretation of lensing measurements is limited by the information available from the light sources, i.e. the galaxies. Much of the premises that allow the study of gravitational lensing are not longer valid when looking at finer levels of measurement, demanding a finer understanding of galaxy evolution to make further advances [HVB$^+$06].

1.2 The Theoretical Effort

The generic picture we can have after all the sets of observations is that 13 billion years ago the Universe looked very simple. Homogeneous and isotropic to one part in 10^5. Today the universe is homogeneous and isotropic on large scales (hundreds of millions of light years) but in smaller scales there is a diversity of complex structures with statistical properties we can measure [Pee93].

The goal is to understand what (when and how) happened in the meantime. Using the known physics to date to do this job, we will require gravity to deal with the large-scale processes and electrodynamics and quantum physics to understand the small scale process.

As the early universe was **homogeneous and isotropic** (just like our present Universe on large scales), our first task is to understand the behavior of such a space. The most immediate theoretical choice we have is the theory of General Relativity, which is a theory describing the interaction of a space time with its matter content. In this theory the basic -mathematical- ingredients are the metric $g_{\alpha\beta}$ (describing space-time) and the energy momentum tensor $T_{\alpha\beta}$(describing matter content).

In the appendix we summarize the principal results for the theoretical description of the large scale evolution of an homogeneous and isotropic universe, emphasizing as well the different approaches to describe an inhomogeneous one. The main conclusion we can derive is that the analytical approach is only useful to describe the average evolution of the universe. As soon we intend to describe the growth of a density perturbation under the

effect of gravity, we be must satisfied to study very simple cases with high symmetry such as the collapse of an spherical over-density.

When we come back from the vastness of cosmological context to galactic scales, to discuss the **formation of the galactic structure** we see the outlook is not much more promising. This is equivalent, to a great extent, to perform a study of star formation, as most of the information we receive from galaxies comes in the form of radiation emitted by the stars.

Downgrading our interest to a more simple problem, we can say that the central problem behind the formation of galaxies is to understand how a tenuous gas can collapse and form a star. We can think that at some point a gravitational instability sets in making a clump of gas start to collapse under its own weight. As it collapses it will lower its gravitational energy, and to keep things in balance the thermal energy of the gas will rise, rising the pressure that at some point will balance the pull of the cloud's own gravity.

To continue the collapse there must be a way to reduce the pressure draining thermal energy away [RO77, Sil77, Bin77, WR78]. In other words, there should be cooling mechanisms that allow the gas to radiate and loose thermal energy. These can be done through different radiation mechanisms as in the case of an excited atom that going back to its ground state emits a photon, at higher temperatures one can think of an ionized gas where the accelerated electrons radiate.

Describing the subsequent steps of collapse can be attempted through the study of hydrodynamics. In general one should also include magnetic fields, as they are observed in the interstellar medium. Complex as it is, analytical descriptions of the evolution of the gas under these circumstances are not available. The numerical approach is the only way out.

1.3 The Computational Trap

The methods of extragalactic astronomy are closer to archaeology than anything else. We only have access to traces of past events, and we can not reproduce experimentally our observations over and over again with different conditions just to test some hypothesis about the formation or evolution of galaxies.

Under these conditions we can only receive help from the third way of exploring nature that started to be established in the scientific and technical community during the last fifty years: numerical simulations. Besides, the complexity of most of these situations, very often resist analytical treatment, leaving numerical simulations as the only available tool.

In our case we are interested in following a great variety of processes. Gravity to follow the large-scale structure; hydrodynamics to follow the fate of gas that is going to collapse and form stars, magnetohydrodynamics (as we observe magnetic fields in the universe) and radiative transfer (as we seek to understand how radiation make its way from its originating process to us). Just to mention the most important.

Of all the fundamental process we mentioned, we have just obtained a satisfactory numerical implementation in the case of the gravity [FWB+99]. For hydrodynamics we have the control in some special cases, and form magnetohydrodynamics in very restrictive conditions. All the rest is territory largely unexplored [ICA+06]. But even including all the physics we are confident enough to model together, from large cosmological down to a galactic scale, it is still very challenging in the usual numerical approach that discretize space and time (or mass), to try to solve a relevant set of equations (gravity, hydrodynamics) to capture the physics [HLF+07]. From the computational point of view it involves achieving an effective resolution spanning 5 to 6 orders of magnitude in mass, length and time. Bearing always in mind that a meaningful comparison with the observations must include theoretical predictions for at least hundreds of thousands of galaxies representing our local universe.

This fully numerical approach, even if reachable for simplified cases, is still reserved for the few with enough computational power available. That said, in order to advance our comprehension of the process of galaxy formation and evolution more democratic and flexible tools must be employed, allowing a broader range of people from the community to test efficiently hypothesis about the process of galaxy evolution. Hypothesis which might not be properly resolved in more detailed numerical schemes.

One approach that has been developed uses numerical simulations for the formation of structure of the dark matter (which is the problem we can model the best) and uses the heuristics for star formation and other hydrodynamical processes to include them into the model as analytic prescriptions (to model the physics we know the worst). This is known as the **semi-analytic approach** which is the one we have used for our work [KWG93, SP99].

1.4 The Semi-Analytic Hope

To reconcile the need for large simulated volumes with detailed populations describe in it, the semi-analytic model (hereafter SAM) approach proposes to describe first the non-linear clustering of dark matter on large scale at coarse resolution, and describe later the small scale baryonic physics through

analytical prescriptions. The connection between the two scales is provided through the dark matter halo, which is the most basic unit of non-linear dark matter structure.

These models are the only window we have on the process of galaxy formation in a explicit cosmological context. In order to be fully confident of its predictive power, the work on these models have been focused on the reproduction of the observed luminosity functions and galaxy counts over a wide range of wavelength and redshift. The dialogue between these two trends (**exploration** of the model results and **construction** of new parts in the model) has been very fruitful, and most of the semi-analytic models have reached a point of internal self-consistency where they are able to reproduce a large set of observations, opening the path for a third element in the discussion; the **capitalization** where the semi-analytic model opens new trends for the theoretical study of extragalactic astronomy.

In this thesis these three points — exploration / construction / capitalization — were addressed with one of the three or four fully fledged semi-analytic models of galaxy formation models existing in the world [GLS$^+$00, HDN$^+$03, BBM$^+$06].

Exploration

The Semi-Analytic approach is just another numerical technique that must be controlled to know which part of the results is coming from artifacts in the numerics, and which part is coming from the physics we intend to model. Furthermore, a SAM seen as a numerical technique is *sui-generis*. The plethora of possible implementations of the same physical phenomenon and the infinite ramifications of possible interactions between different parts of the model convert into a challenge the simplest intention of defining, for instance, an equivalent for the concept of convergence in usual numerical simulations. It makes almost impossible the comparison of numerical performance between two different semi-analytic codes.

During my thesis I have proposed a protocol to provide a framework to address all these issues. The core idea is to introduce small **perturbations in the scalar parameters controlling the baryonic physics** inside the SAM and see how this affects the final results. The final impression is that for large objects, which form in the most hierarchical way after many mergers, the SAMs show a low power of predictability for its properties, which is very likely to be a common feature for a hierarchical aggregation paradigm.

Another theme I have been interested in, comes from the fact that it is

widely recognized that the backbone of a SAM is the merger tree, which is the structure describing the growth history of a dark matter halo. In apparent contradiction to its relevance, the methods to describe and compare merger trees obtained by different approaches are rather poor, as they oversimplify their structure in the sake of simple tests. For that reason I have also proposed a way to describe merger trees by its **contour process**, a simpler one-dimensional structure that keeps intact all the geometrical information. I applied this description to the merger trees of a large numerical simulation and to merger trees derived from Monte Carlo simulations. This allowed me to describe the coarse geometry of the merger trees and find critical physical scales where the geometric properties of the merger tree change, a fact that can have an influence over the description of the underlying galaxy population.

Construction

Our semi-analytic model was incomplete. It gave a good description of the local universe, but it was not the case for the high redshift universe. It offered a good description of Lyman Break Galaxies at redshift $z \sim 3$ but not for the galaxies observed at submillimeter wavelengths. At those redshifts the most strict probe of galaxy formation and evolution is given by the observations of infrared galaxies, as it have been understood from the last decade of observations.

The problem we faced at high redshift is generic to all semi-analytic models. It can be stated as the impossibility to obtain more bright galaxies in the infrared during the first half of the age of the universe.The best solution, after exploring different paths, seems to be a change, at high redshift only, of the initial mass function for stars. This agrees with the result of other groups [NLB$^+$05].

Later, we implemented this solution for a new cosmological context inspired by the latest results reported from the WMAP team, whose more important characteristic is having a lower normalization (in other words, massive structures are less abundant and structure formation is delayed). We find out that in that case we miss again the bright infrared galaxies we had gained by changing the initial mass function.

The overall feeling after all is that, of course, the high redshift universe is very far from being understood, even in the simplified and general picture of semi-analytic models of galaxy formation.

Capitalization

The local Universe still has many aspects to explore and information to extract. Within the new wave of extragalactic measurements, it has been brought to public attention that the Universe is not only expanding, but also accelerating. Further understanding of this fact needs the measurement — and simulation — of galaxies inside a large volume of the Universe. This being feasible by using semi-analytic models.

Along the same line, new measurements of weak gravitational lensing with space based telescopes (reducing systematic effects and scanning large portions of the sky) will find big limitations in the interpretation of its data, most of them coming from our ignorance about the coevolution of galaxies and dark matter. Once again, in order to prepare for this kind of missions, and to later extract all the possible astrophysical information, one needs to simulate galaxies within a large volumes of the universe. Furthermore, it is desirable to create these galaxies inside a realistic cosmological context.

I have given a step in that direction. I built a code capable to extract the weak lensing signal from a N-body simulation and to later produce lensed images of the underlying galaxy population. It was the first time it was done, and as a test of the possibilities offered by the code I explored a potential bias that could affect high precision determination of cosmological parameters from the upcoming weak lensing measurements.

Structure of this document

This thesis explores the three axes I have just described. This was done as an effort to improve the development of the `GalICS`code (GALaxies In Cosmological Simulations). The model is the result of a large collaboration spanning almost a decade [HDN$^+$03, BGD$^+$04, BSC$^+$06] and now various researchers across Europe (Lyon, Potsdam and Oxford).

The chapter 2 is dedicated to the **Exploration** of the model. I present as a technical introduction useful for the rest of the document the general characteristics of SAMs and the particular structure of `GalICS`.

The chapter 3 describes the **Construction** of the improved infrared modeling inside `GalICS`.

The chapter 4 is dedicated to the **Capitalization** of `GalICS`through its use to make realistic weak lensing simulations and, in particular, the simulation code I developed.

How to read this document

This document has been written as one of the requirements to obtain a PhD degree in astrophysics. It has to show the results of my research during the last years and must be read and approved by two referees. On the other hand I want to write a document that would be useful for students beginning their studies on galaxy formation, or simply to people with some general background in astrophysics who want to have a very general idea of what is going in the world of semi-analytic models of galaxy formation.

Communicate what I have done, in a way clear to people ranging from master students in astrophysics to professional astrophysicists, with out shocking the expert in the field with too many imprecision. With this goal in mind, the thesis is split in two parts: chapters and appendices.

The chapters will introduce the general ideas about galaxy formation, infrared galaxies and weak gravitational lensing. The appendices will contain the technical work around each one of these areas in the form of submitted (or published) papers. **The chapters are aimed the student and the appendices the specialist**. The only exception is the Chapter 3 dealing with infrared galaxies, which has the most important results summarized at the end.

Furthermore, each chapter presents in the first part the general concepts, reserving to the end a series of subsections (*The problem, Our Solution, Work in Progress*) describing in a concise manner the motivations for the work we performed, the results we obtained, and the work that remains to be done.

The first chapter about galaxy formation deserves a larger comment. In this chapter I had to introduce ideas about galaxy formation that will be important for the rest of the document, and at the same time explain the detailed points about the particular model of galaxy formation I have used. For that purpose the broad tone of the chapter is directed to the general ideas, and the specificities of the code I used are written inside rectangular boxes easy to spot and skip if wanted. The work around the model (predictability and merger trees) are summarized in two submitted papers in the Appendix.

In the second chapter I do not make that distinction between general ideas and particular implementation, I only give the details about the particular model with a few general considerations, summarizing the final results.

The last chapter on weak lensing deals almost only with general ideas and concepts. The published paper is in the Appendix.

2
GALAXY FORMATION

The study of galaxy formation in the Universe has lead us to formulate the situation in terms of two competing processes: large scale aggregation driven by gravity and small scale fragmentation driven by magneto-hydrodynamical instabilities.

On large scales we think that the dominant matter component in the universe —known as Dark Matter— interacts only through gravity, creating a scaffolding for small scale processes originating structure on galactic scales.

Over galactic scales the question of structure formation is equivalent to asking how tenuous gas can form stars. Even star formation is not understood, but some key processes should relate to (magneto)hydrodynamical instabilities that allow the gas to fragment and collapse.

This duality between scales and physical processes has been exploited at the maximum in the context of semi-analytic models of galaxy formation. These models obtain first the evolution of dark matter structure on cosmological scales to follow afterwards the evolution of baryonic structure on small scales.

The formation of dark matter structure is fairly well understood, and the numerical techniques dealing with gravity are the best managed in the astrophysical community. On the other hand the small scale processes are poorly understood from every point of view, and their implementation into semi-analytic models is done through analytical prescriptions inspired from observations.

In this chapter we review all the physical processes that are being included to date into semi-analytic models of galaxy formation. We present first the themes related to the formation of dark matter structure, and later all the different baryonic process involved in the formation and evolution of single galaxies.

Along all the chapter we will present the details of our semi-analytic

model. In the last part we present two different results that we have obtained in the process of scrutinizing the semi-analytic approach itself. The first is related to a new description of merger trees, and the second is related to the predictability of the semi-analytic models.

2.1 Dark Matter Structure

In this section we review the process of growth of dissipationless structure in the Universe. The physics behind this topic is basically the dynamics of a matter density field under the effect of its own gravity, as we will focus on the physics of the collisionless, non-baryonic dark matter.

We specify first the **cosmological model**, we go on then to describe the formation of the **population** of dark matter halos in this context, to finally make a pass on the detailed structure and formation history of a **single halo**. This top-down approach to link these two scales will allow us to tackle the first important problems in galaxy formation.

2.1.1 Cosmological Scales

Currently we have a concordance cosmological model, and its parameters estimated with unprecedented accuracy [SBD$^+$07]. In that model the dominant matter component is the cold dark matter † nevertheless the universe seems to be under the commanding presence of a cosmological constant [Car04] which today is believed to drive the accelerated expansion of the Universe.

In this theoretical Universe we assume that the initial fluctuations, that originated the seeds for the structure formation in the universe, are Gaussian. Once we have set up the initial conditions for the density fluctuations with the cosmological parameters at hand, we can try to make a guess about the abundance of dark matter structures. The first attempt to do so was done by Press and Schechter [PS74, Sch76], long before the cold dark matter model was widely accepted in the community. They assumed a Gaussian density field and smoothed it on different physical scales. By varying the physical scale of the smoothing window they could construct a mass for that scale, then the abundance of halos above a given mass is calculated from the fraction of volumes that have a given smoothed density.

The mass function (i.e. the number density of structures with a determined mass) predicted by this simple approximation agrees very well with

† A kind of non-collisional, non-baryonic kind of matter that only interacts with itself through gravity. We do not have, as of 2007, a stablished picture of its physical nature

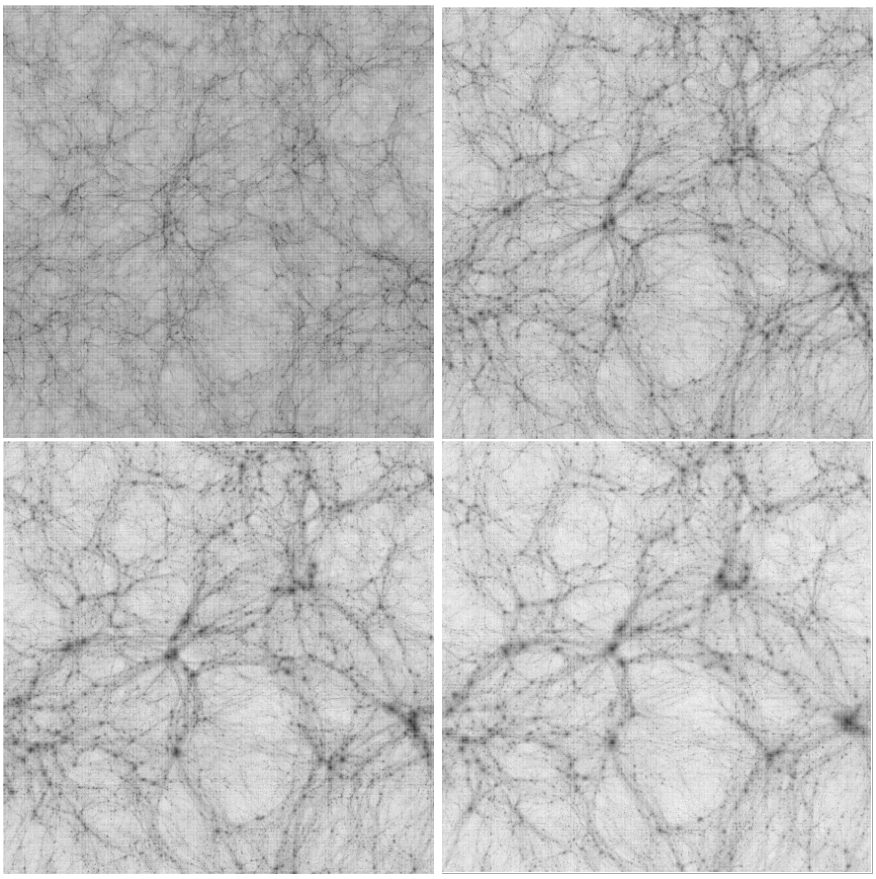


Fig. 2.1. Evolution of a density field for a dark matter only simulation. This represents an slice trough the volume of one tenth the box side ($100h^{-1}$Mpc). The redshifts correspond to redshifts $z = 10, 3, 1, 0$. High densities are represented by darker colors, the color scale is proportional to the logarithm of the column density. This simulation was done by Christophe Pichon using the publicly available code GADGET-2.

the results of numerical simulations of a dark matter density field. The accuracy with which we can produce mass functions of halos using the Press-Schechter approach have been improved in the nineties by further refinements, most notably by Sheth and Tormen [SMT01].

2.1.2 Galactic Scales

From the cosmological context, we can start to describe the structure of single dark matter halos, which are equilibrium entities created in the developed stages of nonlinear stages of dark matter collapse. In a first approximation

we can say that a dark matter halo is spherical with a density profile depending only on the radius. Using numerical simulations in the early 90s Navarro, Frenk and White found that dark matter halos follow a universal density profile (known as NFW profile) [NFW97]. For this density profile we do not have a clear explanation for its emergence. A much simpler profile, with a longer career in astrophysical contexts, providing a rather good approximation to the results in numerical simulations for the most of the volume of the halo, is the isothermal profile

$$\rho_{DM}(r) = \frac{M_{vir}}{4\pi R_{vir}} \frac{1}{r^2}, \tag{2.1}$$

where M is the mass of the halo, R its radius, and the subscript vir refers to virialized (equilibrium) quantities. We can associate a temperature to the halo, which is naturally called virial temperature

$$T_{vir} = \frac{\mu m_p}{2k_B} \frac{GM_{vir}}{R_{vir}}, \tag{2.2}$$

where $\mu = 0.57$ is the mean molecular weight for an ionized gas composed of 75% hydrogen and 25% Helium, m_p is the proton mass, k_B is the Boltzmann constant and G the gravitational constant.

Another important characteristic of a dark matter halo is its rotation that is quantified through the spin parameter

$$\lambda = \frac{|E|^{1/2}|\vec{J}|}{GM_{vir}^{5/2}}, \tag{2.3}$$

where E is the total energy of the halo and $\vec{J}$ is its angular momentum

Later, when higher resolution simulations where done, the community saw that dark matter halos were not featureless [MGQ$^+$98, MQG$^+$99, KKVP99, CAV00]. They host a good deal of substructure, which can be thought as smaller halos inside a larger halo. Typically, around 15 percent of the total mass of dark matter halo is found in the form of those substructures [GMG$^+$00]. How was created this substructure? Understanding that process imploes the study of growth of a dark matter halo as function of time. In the same way genetic information is passed from progenitor to descendant, some information of the halo final structure should be found in the history of its progenitors.

In order to do this genealogical analysis, the simplest approach is to analyze the distribution of dark matter halos at different times, and to try

to link the populations from one generation to the next, concluding by the construction of a genealogical tree for each halo in the last generation we are considering (Figure 2.2).

2.1.3 Halos and Merger Trees

The genealogical tree of a dark matter halo, in the context of structure formation, is known as a merger tree. This structure is fundamental to the construction of galaxies inside hierarchical scenarios. Accordingly, there is a great variety of ways to build such a tree.

For an analytical approach to merger trees, the natural step is to sample the Press-Schechter halo distribution at different times, in order to link the different populations. Usually the implementation of this idea starts from the final halo and draws a conditional probability for the mass of their parents, continuing recursively until every halo has a parent. Naturally, there is also a condition on the minimal mass a parent can have, implying that the tree cannot be infinitely large with smaller and smaller halos on it. This is known as a Monte-Carlo approach to merger trees [ND08, **?**, PCH07]. It has been criticized in the past because a merger tree built in that way, won't naturally reproduce the halo mass function at different times. It means that if we select all the halos at present, and take all its parents at different times, the halo mass functions we build from the parents in the tree won't match the original halo mass functions from which we drew them [BKH05]. More recent implementations seem to have fixed this problem using additional ad-hoc conditions [ND08, **?**, PCH07].

The fundamental assumption behind all the algorithms implementes in a Monte-Carlo approach is that the formation history of the halo depends only on its mass. This assumption has been challenged based on the results of large high resolution dark matter simulations. It has been shown, for instance, that the halo environment has an impact on its formation history [GSW05]. This can be seen as a major caveat to using merger trees from Monte-Carlo simulations, but it remains to be shown that indeed the merger tree properties change significantly with the halo environment.

Another approach to merger trees is numerical. They are based on the same principle as before, detecting populations at different cosmological times to link them trough parenthood relationships. The difference is that we define the structures at different timesteps from the snapshots of numerical simulations. Now, all the magic (and variety of applications) is in the algorithm used to identify structures. The most popular algorithms identify structures based on a percolation process, where particles in simulation are

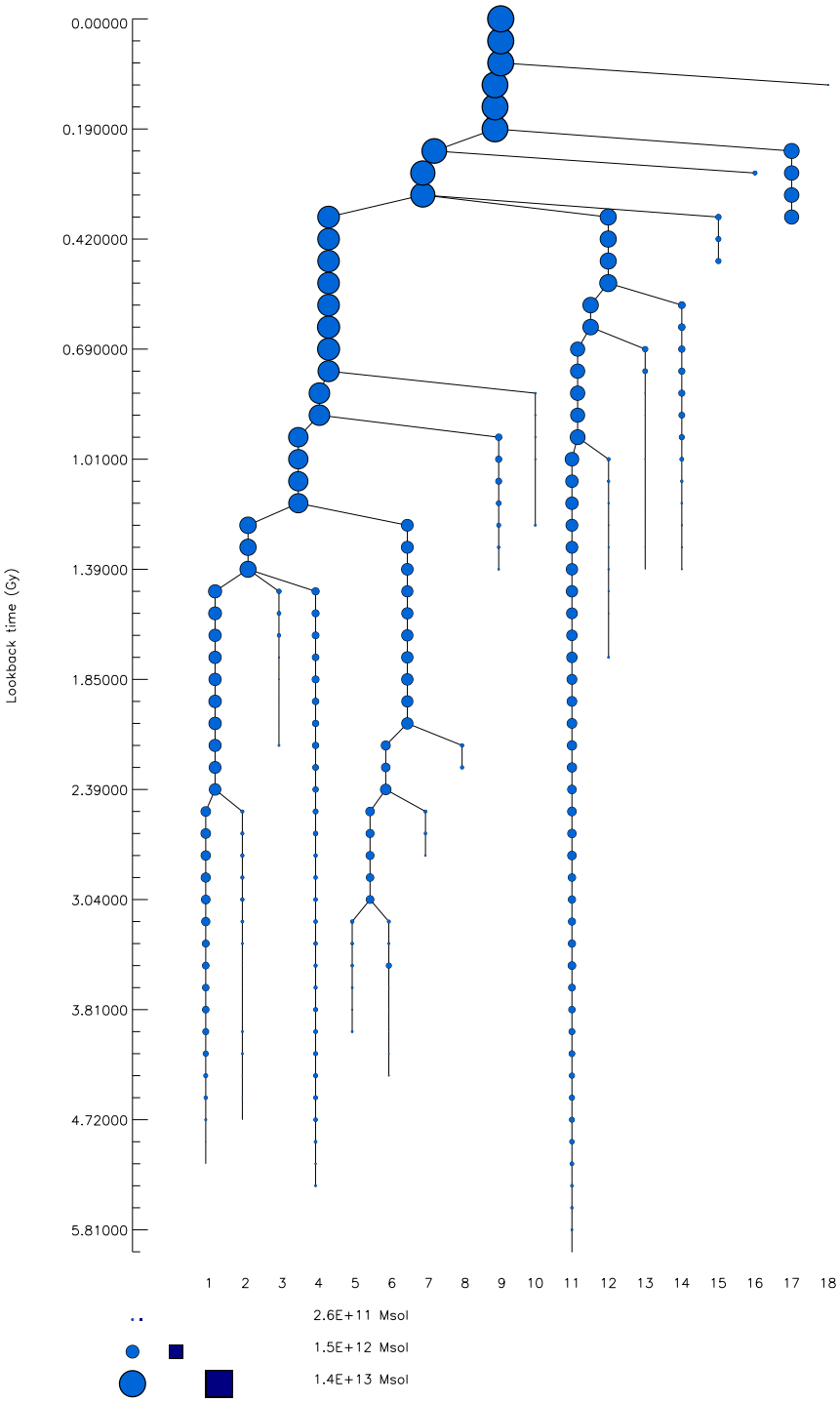


Fig. 2.2. Merger tree of a dark matter halo of mass $\sim 10^{13}\mathrm{M}_{\odot}$ constructed from a dark matter simulation. Each circle represents a dark matter halo with an area proportional to its mass. The horizontal levels mark each timestep in the simulation, and the lines between the circles show parenthood links. Time advances from the bottom to the top of the page. The mass of the smallest halos present in the tree is fixed by the resolution of the dark matter simulation. Figure kindly prepared by Dylan Tweed using data from the simulation presented in Figure 2.1.

linked to each other if they are found to be closer than a given distance. The linking length, as it is called this distance, is expressed usually as a fraction of the mean distance between particles. Its value is selected to have objects (sets of particles) of a particular global overdensity. The numerical techniques based on this percolation algorithm have been also extended to account for this substructure in halos. Substructure identification algorithms are based on the detection of the highest density regions in the simulation and proper profiling of the density morphology in the neighborhood of these density peaks. [DEFW85, KGKK99, SWTK01].

Once the dark matter structure has been identified inside the simulation, different halos (and subhalos if the algorithm permits) can be linked from one timestep to the next. This way of constructing merger trees is in principle more exact and correct than the Monte-Carlo approach, nevertheless the construction of the trees is obstructed by the fact that dark matter simulations have a finite resolution, meaning that for a given halo at redshift zero, we can only describe its progenitors above a given minimum mass [HDN$^+$03]. This fact is in contrast with the Monte-Carlo trees that have, in comparison, an infinite resolution.

We have now two different approaches. The numerical is usually thought as the correct one, all the results from Monte-Carlo trees are usually validated in comparison with numerical simulations. Even we can say that they are tuned to reproduce some of its characteristics. This line of thinking suggests the relevance of having good parameters and techniques to analyze merger trees (i.e. the ability to compare different trees) and saying if they are similar one to the others. In general, the analysis of merger trees has been reduced to the most massive branches, cutting from the root all the complexity encoded in the tree. Other methods compare mean properties over all the generations, wrapping up all the information under a single blanket [ND08] .

We propose a new way to describe merger trees and discuss some of its consequences based on the analysis of the Millennium Simulation [LS06] and the Monte Carlo code made public by [ND08]. See the Appendix with the submitted paper.

We follow the dark matter distribution using numerical simulations. We use the publicly available GADGET2 [Spr05] for the simulations of growth of dark matter structure. To identify the halos we use a friend of friends algorithm [DEFW85].

For each halo we store its mass, radius and spin. Afterwards we build the merger tree using the identified halos in the snapshots of the simulation. For every calculation that assumes a density profile of the dark matter halo, we use an isothermal profile truncated at 0.1 kpc.
We have also used merger trees constructed with the SUBFIND [SWTK01] algorithm on the Millennium Simulation [SWJ+05]. This was used only in the description of merger trees by contour walks (see the Appendix).

Now that we have the tools to describe the growth of dark matter structure, we can consider a toy model of galaxy formation. In this model the simplest assumption one can make, is that galaxy luminosity is proportional to the halo mass. If we do that the abundance of galaxies as a function of the luminosity should follow the same trend as the abundance of dark matter structures which is $d(\log N)/d(\log M) \propto M^{-1}$. But knowing from observations that the abundance of bright galaxies is best described by a sharp exponential cutoff, and that at faint luminosities the slope α in the relation $d(\log N)/d(\log L) = L^{-\alpha}$ is less than 1, we can say that in the toy model we would have too many galaxies at both the bright and faint end.

This gives us a hint that the processes building the luminous part of the galaxies are far more complex than those forming dark matter halos. This is the motivation of the current section dealing with the physics of gas and how it transforms into stars.

This sections explain the general understanding we have of these complex processes in three parts. The first one deals with the cooling of the gas (how much gas do I have available to form stars?). The second will deal with star formation process (how much gas transform into stars?), and the last one will treat the physics of feedback meaning that we complete a full circle to the first question, and will try to see how the formation of stars and structure will modify the quantity of gas available to form stars.

2.1.4 Cooling

In the context of a cold dark matter model, where dark matter halos are thought to be the cradles of galaxy formation it is of major importance to list the available mechanism of gas cooling inside dark matter halos. In other words, find the way about how a halo can trap and cool down mixture of tenuous hot gas to start forming stars.

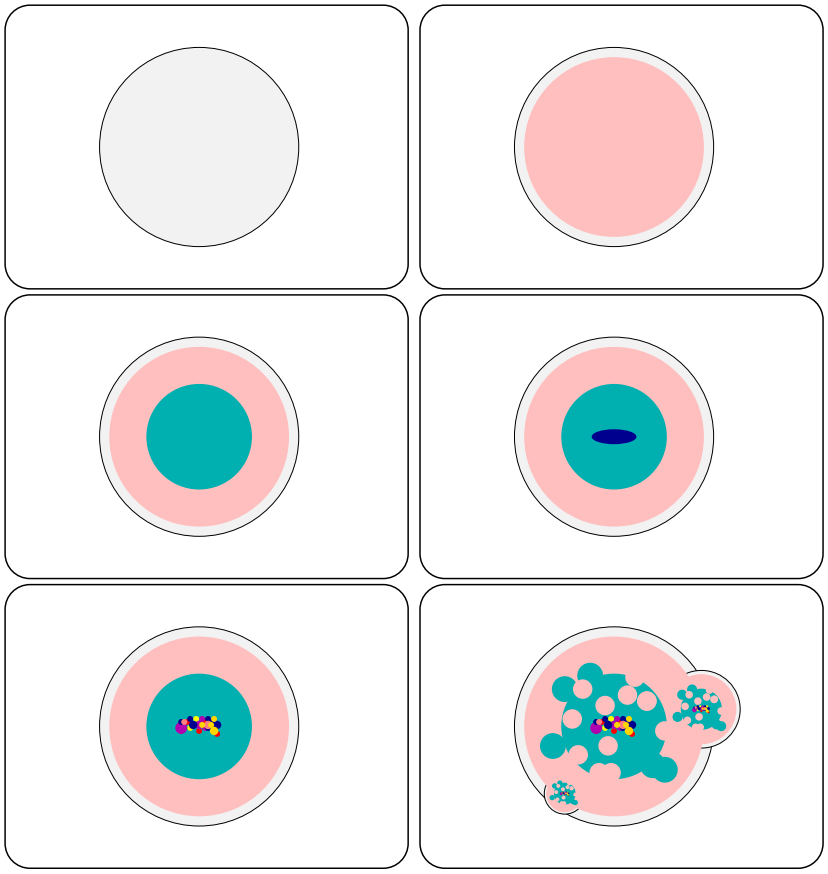


Fig. 2.3. Principal stages in the process of galaxy formation in the classical hierarchical point of view. From left to right, top to bottom: 1. single dark matter halo forms. 2. gas falls into its potential well getting shock heated. 3.Gas starts to cool down radiatively. 4. A rotationally supported disk is formed. 5. Stars start to form 6. SN explosions, AGN feedback, and merging add additional complexity to the process.

The basics of this picture where gas cools inside dark matter halos was set up about thirty years ago, and still is inside this framework that studies of the galaxy formation continue today. In a schematic way, these theories try to explain how much, and how fast, can gas cool down in terms of the parameters describing the dark matter halo.

One picture tells us that we can see things as a two stage process. First the halo is formed, then the gas falls into the potential well of the dark matter halo and is shock heated to the virial temperature of the halo [Sil77, WR78]. This temperature depends only on the mass of the halo as can be seen in

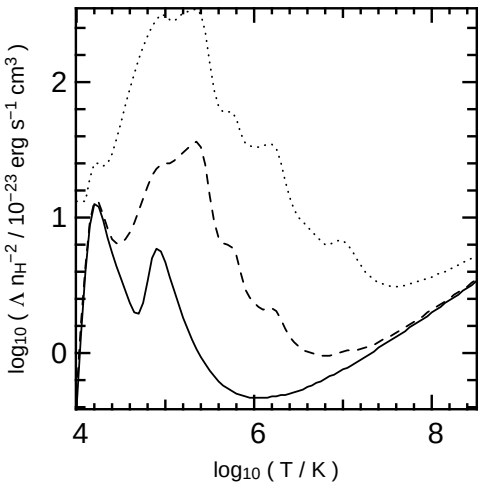


Fig. 2.4. Cooling curve from [SD93] as a function of the temperature for zero metallicity, 1/10 $Z_\odot$ and 1/3 $Z_\odot$ (continuous, dashed and dotted lines, respectively).

the Eq.2.2. Gas can now cool, on a physical timescale that depends on its temperature, density, and chemical composition. As the gas cools it will fall to the center of the halo to form a rotational supported disc.

This picture have been confronted from the very beginning [Bin77]. The opposing picture states that the gas cannot be shock heated at all. It has been suggested by theoretical and numerical work [BD03], that in fact the process is more complex than the two-stage picture. The halo could be fed with cold gas (relative to the virial temperature of the halo) resulting in a much more efficient supply a of material to form stars.

One detail remains fundamental: the ability of gas to radiate energy away. Gas can cool from a variety of physical process, their relative importance depends on the physical conditions of the gas such as temperature, density and chemical composition. The most relevant processes are the emission of photons following transitions between atomic levels; the gas cools down when the excited levels decay emitting a photon. This is the dominant mechanism between 10^4K to 10^6K. Bremsstrahlung radiation from electrons accelerated in the plasma is the dominant emission mechanism for temperatures higher than 10^7K. Below 10^4K the only relevant process would come through H_2 cooling, which is only relevant for the formation of the first structures when the universe was not ionized.

The amount of radiated energy can be quantified through the cooling function $\Lambda(T, \rho, Z)$. This curve is a function of the temperature T, density

ρ metallicity Z of the gas. The rate of radiative energy loss is $\rho^2\Lambda$ where ρ is the density of the gas. A detailed, and very popular, description of this cooling curve have been calculated by Sutherland and Dopita [SD93]. The Figure 2.4 shows this curve for three different metallicities.

From this cooling efficiency we can define a typical cooling time, as the ratio of the energy stored thermally in the gas to the rate of energy emission from radiative processes.

$$t_{cool} \equiv \frac{E_{th}}{\dot{E_{rad}}} = \frac{3}{2}\frac{\rho_g k_B T_g}{\mu m_p} \times \frac{1}{\rho_g^2 \Lambda(T_g, Z)}. \tag{2.4}$$

The temperature that is included in the previous equation, is the temperature of the gas. This is why is so important to be sure about which picture is correct about the infall of the gas inside a dark matter halo. If indeed the gas is shock heated to high temperatures, or if the gas is already cool this is very likely to make a distinction about the fate of baryonic structure in the halo.

Nevertheless, it can be argued the most important time scale is not the cooling time scale, but the free fall time scale, which is is the typical time that will take for the cold gas to reach the center and form the galaxy disc. Detailed discussion on this issue can be found in [Bau06], the conclusion is that the most striking difference on the two formation processes is not rate of cold material available to form star but the **geometry** of the feeding. The usual picture we have explained until now, allows the gas to cool only into the central galaxy. In a picture where the cold gas can invade the halo through filaments, it would seem reasonable to let the gas cool on some satellite structures, if any. Or even think about a perturbed initial disk.

The cooling curve we use comes from Sutherland and Dopita, it takes into account all the atomic process available to cool the gas, such as electronic transitions between bound states, bound-free and Bremsstrahlung. In particular, it imposes an ionization state dependent on the temperature, the metallicity, and density.
In our model we define the mass that is going to cool during the time Δt as

$$M_{cool} = \int_0^{r_{max}} 4\pi r^2 \rho_g(r) dr, \tag{2.5}$$

where r_{max} is the minimum between the cooling radius (defined by $t_{cool}(r_{cool}) = \Delta t$) and the free fall radius defined by $\Delta t \times V_c$ and V_c is the circular velocity of the halo

There are plenty of phenomena that must be taken into account to make the thermal balance of the gas. Virtually every physical situation that can modify, not only the temperature, but also the chemical composition must be taken into account.

One important mechanism of that kind, is the background of UV photons generated by the star formation process on cosmological scales, this background increases the degree of ionization of the hot gas, erasing the possibility for gas to cool through atomic transitions, removing the low temperature mountain in the zero metallicity cooling curve (solid line in Figure 2.4). This process is most efficient in low mass halos with masses below $\sim 10^{10}\mathrm{M}_{\odot}$, imposing lower baryonic fraction than the cosmic one [Som02, TSTV02, HYG07]. Other processes one can consider are heating from supernovae or AGNs. We will pass a review on these processes in the section dealing with feedback processes. But a necessary condition to get to that section is the discussion about the physics of star formation.

As an end to this theme of cooling, one can ask: what happens to the gas that cooled? In semi-analytic models of galaxy formation the typical answer is: it forms a rotationally supported disc. The answer from detailed numerical simulations is the same, with the caveat that the produced discs are smaller and with less angular momentum than the observed ones [NW94].

In SAMs this problem is avoided by modeling the size of the disc from the size of the virial radius of the halo R_{vir} and its spin parameter λ (Eq. 2.3). The prescription used in SAMs assumes that the specific angular momentum of the gas is conserved and that the final surface density profile is exponential $\Sigma(r) = \Sigma_0 \exp(-r/r_d)$ [FE80, MMW98]. Where the disc size r_d is

$$r_d = \frac{\lambda}{\sqrt{2}} R_{vir}. \tag{2.6}$$

In some models the size of the disc is adjusted at every timestep to follow the evolution of the spin parameter in the halo. In models that follow the dark matter evolution from numerical simulations, this approach produces strong changes in the size of the discs, which might be considered as unphysical.

In our model we setup the initial disc in a new dark matter halo using Eq.2.6, and then follow its growth. At every timestep we calculate r_d from the mass already in the disc and the newly accreted mass of cold gas.

Once we have fixed the disc size we define its circular velocity at the half-mass radius $r_{1/2} = 1.68 r_d$ assuming rotational equilibrium

$$v_d^2 = \frac{GM(r < r_{1/2})}{r_{1/2}}, \tag{2.7}$$

where $M(r < r_{1/2})$ is the total mass inside the half-mass radius including the other components of the galaxy and the dark matter. The disc can growth until it merges with another galaxy or until a disc instability installs. We will detail the merger implementation later. We describe now the instabilities implemented in our model following [van98]

We define α as the ratio between the disc circular velocity v_c at the half mass radius, calculated from the disc mass only, and v_d as defined above: $\alpha = v_d/(2v_c)$. The disc will be stable if $\alpha < \alpha_{critical}$, where $\alpha_{critical}$ is given by [van98] as 0.7.

If the disc is unstable we transfer a part of its gas and stars to the starburst component. The transfered mass is calculated to bring α to a marginally sub-critical value. This procedure gives a channel to morphological evolution, i.e. a mechanism to grow bulges.

2.1.5 Star Formation

We do not understand star formation. However, we have relatively detailed observational studies that can correlate different environmental properties in a galaxy with this star formation process. These empirical scaling relations are at the heart of the semi-analytic models, which can implement these relationships as black boxes with a few efficiency parameters that can be chosen to reproduce observations.

Star Formation Rate

The Kennicutt-Schmidt empirical law for star formation states that rate of star formation per unit the area of galaxy is proportional to a power of the surface density of the cold gas, with a surface density threshold below which star formation does not take place [Ken98].

In terms of a simple recipe for a semi-analytic model, we can recover the Kennicutt-Schmidt law by simply saying that the star formation rate is proportional to the ratio of cold gas in the galaxy, m_{gas} and a typical time scale t_{dyn}. This time scale generally is associated with the dynamical time of the galaxy or to some other representative time.

$$\Psi_\star = \alpha \frac{m_{gas}}{t_{dyn}} \tag{2.8}$$

The free parameter α is ~ 50 following Kennicutt and Guiderdoni [Ken89, GHBM98]. We set the dynamical time for each component. For discs it is $\pi r_{1/2}/v_d$ and for spheroids it is $r_{1/2}/\sigma_b$, where v_d is the rotational velocity in the disc and σ_b is the velocity dispersion in the bulge. We impose the observed column density threshold of 10^{20} atoms/cm^2 to trigger star formation.

Initial Mass Function

The star formation rate is just half the story. We still need to know what kind of stars will be formed. The type of stars we have in our formed galaxy will determine, based on the well tested stellar evolution theories, the amount of metals that will be produced, the quantity of supernovae that are going to explode, the proportion and stars that might become black holes and, above all, the spectral signature of the galaxy, which in the end is the only quantity we can measure with our telescopes.

The concept we are looking for is the one of the Initial Mass Function (IMF). Its observational determination is a complex problem that needs to deconvolve many parameters such as star formation histories, mass luminosity ratios of stars,dust extinction, just to mention as few of them [Kro02].

But also in this case, we can use a simple description that fixes the number of stars in a given bin of mass; $dN = \xi(m)dm$ is the number of stars in the mass interval m, $m + dm$ and ξ is the IMF.

In our model we use a Kennicutt initial mass function which is defined by $\xi(m) \propto m^{-1.4}$ for $0.1 \leq m \leq 1$ M$_\odot$ and $\psi(m) \propto m^{-2.5}$ for $1 \leq m \leq 100$ M$_\odot$. In order to reproduce the counts of submillimetre galaxies we have to introduce in the starburst a top-heavy IMF which only produces stars with masses ≥ 8 M$_\odot$.

Chemical Evolution

In comparison to stellar evolution or hierarchical formation of structure, galactic chemical evolution (GCE) models are poorly understood. Their predictions are not robust as is the case for the fate of stars of a given mass, or the timetable for dark matter halo formation. This is due in great part

to the fact that the key physical process for GCE is collective star formation which, as we stated at the beginning, is not well understood.

In the simplest GCE models, the ejecta of dying stars are instantaneously mixed in the interstellar medium, which means that we can parameterize the time evolution in our system by a chemical composition which is fixed at every time. In this simple model the gas is initially composed by a mix of primordial abundances of helium and hydrogen. As the gas is transformed into stars with a given star formation rate $\Psi_\star$, stars are produced with an initial mass function $\xi(m)$, with each star having its own lifetime τ_m which depends only on its mass, m. The stars of mass m created at a time t will die at the time $t + \tau_m$ to return a part of its mass (through stellar winds or supernovae explosions) making the interstellar medium more and more enriched in elements heavier that Hydrogen and Helium [Pra07].

The chemical evolution of the system can be expressed in a system of integrodifferential [Tin80]. The important point being that to solve them we will require the following information

- Individual stellar properties
  Stellar lifetimes τ_m, the masses of stellar residues, $w(m)$, and metal yields, $Y_Z(m)$.
- Collective stellar properties
  The initial mass function and the star formation rate, which are imposed by a observational constraints and are implemented as we explained previously.
- Gas flows
  Exchange into and out of the system, in our case the most important infall process is the continuous accretion of gas, the outflow process may come from the supernova-wind driven ejection of the gas.

In our model the quantities τ_m, $w(m)$ and $Y_Z(m)$ come from the Geneva stellar evolution tracks [LS01], and are calculated for 5 different metallicities ($Z = 0.001, 0.004, 0.008, 0.04$). The evolution is solved in the general condition of instantaneous mix, but not instantaneous ejection of metals as we follow the time evolution of the star population, first with a timestep of 1 Myr (for 10Myr), then with a timestep 10Myr (during 100Myr) and finally with a timestep of 100Myr up to 20Gyr.

2.1.6 Feedback

In our toy model of galaxy formation, where the galaxy luminosity was proportional to the mass of the dark matter halo, we found that we had to do something at the bright and the faint end to save the model. All the physical justifications that have been thought during the last decade to iron out these discrepancies receive now the name of feedback. For the faint end we invoke feedback from supernovae. To decrease the number of galaxies at the bright end, one speaks in terms of feedback from Active Galactic Nuclei (AGN).

It is highly controversial to model these effects. Not only because of the physical uncertainties but mainly because they have been invoked *afterwards* to save a model. Besides, even today, with the ever-growing available computational power, simulation these processes is far from trivial without invoking again dubious couplings between different physical scales. For instance, it is not clear how the physics on parsec scale around a black hole (the engine believed to be powering AGNs) can affect the thermodynamical state of the gas on kiloparsec scales. Which is the process that makes the energy transfer possible? The same goes for the supernova feedback in galaxy discs, the hydrodynamical methods to simulate these explosions count on releasing energy directly into the gas without any certainty about the physical processes that will make the energy from the explosion available to the interstellar medium [STWS06].

We explain now how this new layer of complexity is included in our simulations, giving again a clear example of the semi-analytic spirit in galaxy formation.

2.1.6.1 Faint End Feedback

In the case of supernovae feedback, the historical first implementation was done by Dekel and Silk [DS86]. It was based on a detailed balance between the energy in ejecta by a supernovae and potential energy of the gas in the galaxy. This feedback is not very efficient on massive galaxies, as the potential well of the galaxy gets deeper and deeper as we increase its mass, making it more difficult for the ejecta to escape.

In contrast, there is a feedback model proposed by Silk [Sil01] that includes two novelties. First, a factor quantifying the mass of cold gas taken away by the hot gas, second, and more importantly, the porosity of the interstellar medium. This feedback implementation is far more aggressive on newly formed discs regardless of its mass.

In our model we have used the Silk prescription, where the rate of gas mass loss is written assuming a stationary model

$$\frac{dm_{out}}{dt} = \Psi_\star \times \eta_{SN} \Delta m_{SN} \times (1 + L) \times (1 - e^{-P}) \tag{2.9}$$

where $\Psi_\star$ is the SFR, η_{SN} is the number of supernovae per unit mass of formed stars (which is a fixed number for a given IMF) and Δm_{SN} is mean mass loss of one supernova (~ 10 $M_\odot$).
We define

$$1 + L \equiv \epsilon \frac{m_{gas}}{m_{gal}} \tag{2.10}$$

where m_{gas} and m_{gal} are the gas and total mass of the galaxy component, and ϵ is an efficiency parameter. We further define the porosity of the galaxy component as

$$P = \alpha \left(\frac{m_{gal}}{m_{gas}} \right)^{1/2} \left(\frac{17.8}{\sigma} \right)^{2.72} \tag{2.11}$$

where σ is a typical dispersion velocity in the ISM (in km/s), which we fix to 10km/s for discs and to the velocity dispersion for the spheroidal components. The parameter α is the star formation efficiency.
In this model, usually the ejected amount of gas is of the same order of the mass of formed stars in young discs.

2.1.6.2 Bright End Feedback

The biggest problem haunting hierarchical models of galaxy formation from the early days, is the production of large galaxies with too much gas and too much ongoing star formation at redshift zero [KWG93]. There is a need to suppress cooling and star formation in the most massive halos that host the massive galaxies.

The implementation of feedback on massive halos is centered in heating the intra-halo gas, and has basically two forms. In the first, the extra heating is **external** to the halo, in the second the heating is **internal** to the halo. An example of the first one is the convection model, where energy conduction from the outer parts of the halo towards inner parts can prevent cooling in halos above a threshold circular velocity. The most popular and typical example of the second kind of heating (internal heating) is composed by the the zoo of AGN models [GDS$^+$04, MF05, CBDG05, CSW$^+$06, BBM$^+$06],

which propose a direct injection of energy from the accretion process around the central black hole. It is necessary to clarify that there are two marked different trends for AGN feedback. One is the thermal mechanism where the energy input from the AGN is used to keep the hot gas hot. The other trend is a mechanical effect that stirs the gas, even ejecting some of it out of the halo. The observational evidence in favor of this AGN feedback are not clearly established and accepted yet.

> In our case, we have not used a detailed implementation that follows the growth of the central black hole to estimate the feedback. Instead, we say that for halos with galaxies that have a total bulge mass higher that a given threshold mass (on the order of $10^{11}\mathrm{M}_{\odot}$) the cooling is stopped.
> This is based on the local correlation between the bulge and black hole mass established by Magorrian [MTR^{+}98] and on the assumption that the injection of energy from the AGN will keep the intra-halo gas hot.

2.1.7 Merging

The previous sections (cooling, star formation and feedback) are fundamental to get the picture for the evolution of an isolated galaxy. But another key point in the context of hierarchical formation of structure is, of course, to understand what happens with two merging galaxies.

In the picture we have been following so far, we can have many halos merging along its history of formation, which allows the existence of many galaxies inside a dark matter halo at a given time. When a halo enters a larger dark matter structure, becoming a satellite, one can consider at least two process that lead to a galaxy merger: dynamical friction and satellite-satellite collisions.

The most important is dynamical friction [Cha43] causing satellite galaxies to sink gradually to the center of their host halo. The decrease of the radii of the orbits can be described by the following differential equation [BT87]

$$r\frac{dr}{dt} = -0.429\frac{Gm_{sat}}{V_c}\ln\Lambda \tag{2.12}$$

where r is the orbital radius, m_{sat} is the mass of the satellite, V_c is the circular velocity of the dark matter halo and Λ is the Coulomb logarithm

approximated by $1 + (m_{halo}/m_{sat})$. If the time to get to the center is less than the lifetime of the halo, then the satellite will merge with the central galaxy.

Besides being a mechanism to increase the mass of the central galaxy, there are many other consequences of the merger process. The numerical work done about galaxy mergers ([BH92, MH96, MH94] for early developments) is small considering the vast range of parameters and possible initial conditions to consider, but the general conclusions are that galaxy mergers lead to intense episodes of star formation and to a morphological evolution that builds a channel to construct spheroidal components from discs. Until very recently there has been a systematic effort to model galaxy mergers in realistic cosmological setups [DCC$^+$07], as well very pronounced developments in the described physics [CJPS06, dCMS07, KKS$^+$07].

Semi-analytic models usually parameterize the merger process using the mass ratio of the two galaxies involved. For instance when the ratio is ~ 1 the discs are assumed to be destroyed and form an spheroidal remnant. The recipes have a wide degree of sophistication.

In our case we implement both dynamical friction and satellite-satellite collisions. To define the post merger morphology we use the parameter X which varies smoothly from 0 to 1 and depends on the ratio of the heavier to the lighter progenitor, R, with the constraints $X(R = 1) = 0$, $X(R = 5/2) = 1/2$ and $X(R = \infty) = 1$.
If we consider a vector $\mathbf{V} = (m_d, m_b, m_s)$ with the masses of the disc, bulge and starburst respectively, and $\mathbf{V}_{1,gas}$, $\mathbf{V}_{2,gas}$ represent the gas masses of the two merging galaxies, the gas mass for the new galaxy is defined as

$$\mathbf{V}_{new,gas} = \mathbf{A}_{gas}\mathbf{V}_{1,gas} + \mathbf{A}_{gas}\mathbf{V}_{2,gas} \tag{2.13}$$

the same goes for the masses of stars

$$\mathbf{V}_{new,star} = \mathbf{A}_{star}\mathbf{V}_{1,star} + \mathbf{A}_{star}\mathbf{V}_{2,star} \tag{2.14}$$

with $\mathbf{A}_{gas}$ and $\mathbf{A}_{star}$ defined as follows

$$\mathbf{A}_{gas} = \begin{pmatrix} X & 0 & 0 \\ 0 & 0 & 0 \\ 1-X & 1 & 1 \end{pmatrix} \quad \mathbf{A}_{star} = \begin{pmatrix} X & 0 & 0 \\ 0 & 1 & 0 \\ 1-X & 0 & 1 \end{pmatrix}. \tag{2.15}$$

In order to understand the meaning of these equations let's take as an example the merging of two galaxies with very different masses $X(R = \infty) = 1$, meaning that the disc transfer all of its mass (gas and stars) to the new disc (the same goes for the bulge and the starbusrt). With the only exception for the gas in the bulge which is transfered to the starburst.

2.2 Spectra

Now we are closer to the end of the exposition of our rough understanding of the process of galaxy formation and evolution. The link that is missing is perhaps the most important one because its the key to the comparison of our model with the observations. We refer to the spectral energy distribution (SED) of the galaxies.

After some work we can create a model that will predict the amount of the gas and stars inside the galaxy, and with more detail we can fix chemical composition and the ages of the population of stars (even at some point we can include an AGN) and from that we can also make a rough estimation on the dust extinction and emission [DGS99, DG00].

With all those elements it is possible, in principle, to build the spectra of the galaxy. After years of research the most straightforward part to calculate is the contribution from the stars [BC03]. But we can also be rather confident with the description of the radiation processing by dust, and include its emission in the galaxy spectrum. In this section we describe only the stellar part, leaving the details about dust treatment to the next chapter that will treat the infrared emission of galaxies.

2.2.1 Stars

At every moment we can write this sort of a contribution to the spectra in the following way

$$F^{\star}_{\lambda}(t) = \int_0^t \int_{m_{down}}^{m_{up}} \frac{dM_{star}(t-\tau)}{dt} \xi(m) f_{\lambda}(m, \tau_m)\ dm\ dt, \tag{2.16}$$

where $f_{\lambda}(m, \tau_m)$ is the flux of one star of initial mass m and age τ_m at wavelength λ, $\xi(m)$ is the initial mass function, m_{up} and m_{down} are the maximal and minimal mass possible. The immediate way to solve this double integral is the discretization in mass and time [BC03]. But the best approach from the numerical pointed view uses the discretization only in time. This

schema is called isochrone, and describes the evolution of a starbursts of a population of stars at a given metallicity on discrete periods of time, that chosen in such a way to follow closely the most important moments in the evolution of the spectra of the star.

2.2.2 Gas

An important element to complete the comparison between synthetic spectra and observations is the inclusion of the spectra from regions ionized by star formation, and in general any emission line process that gives information about the chemical conditions in the interstellar medium of our galaxies. This is not usually included in semi-analytic models.

> At the moment Nicolas Bonhomme is working on the inclusion of nebular emission from ionized regions and gas, following the approach and results summarized in [MR02]. In this approach the outputs from the stellar modelisation part (spectra and metallicity) are coupled with the photoionization code CLOUDY [Fer96] which predicts the spectra of astrophysical plasma over a wide range of densities under extreme conditions, such as starbusts or quasar environments. This is under progress, but the results of the comparison with the results of SDSS are encouraging.

2.3 Observations

It has been often criticized to semi-analytical models of galaxy formation that they have too many free parameters. But at the same time we should say that they have to fit many observational constraints that nowadays outnumber the number of free parameters. Until now, no model of galaxy formation can reproduce them all.

In this section we enumerate all the well established and more relevant observations used to test semi-analytic models of galaxy formation. We start by the local optical observations (more numerous, with higher precision) to finally go higher up in redshift and wavelength, keeping to a separate chapter the observations and our original results dealing with infrared galaxies.

2.3.1 Low redshift

In this part we review all the structural information than can be extracted from local observations.

- **Luminosity function**
  Is the most basic description of a galaxy population. Counting the number density of galaxies inside a luminosity bin. In the last years with surveys such as SDSS and 2dF, we have a rather good description up to very faint luminosities in the local universe. Matching the bright and the faint end of the luminosity functions is what have motivated all the work in the last decade in galaxy formation models, pecially on the SN and AGN feedback recipes like the ones we have reviewed in the last section.
- **Tully-Fisher relation**
  Is an observed correlation between the rotation speed and luminosity of spiral discs [TF77]. To date, there are no semi-analytic models able to match the zero point of this relation while at the same time reproducing the luminosity function. The slope in the logarithmic plot speed vs. luminosity can be reproduced.
- **Fundamental plane**
  In a similar way as quoted before, there are tight correlations between the radius, the velocity dispersion and luminosity of elliptical galaxies which can be described by the fundamental plane [DD87, DLB$^+$87]. Semi-analytical models have provided a reasonable match to the local fundamental plane
- **Morphology**
  The morphology segregation (disc galaxies are blue, and the elliptical galaxies are red) have been quantified [BMdC94], this naturally provides another test (not mentioned in detail by many) of a galaxy formation model.
- **Size of the galaxies**
  There are observations quantifying the sizes of discs in the local universe [dL00]. Generally speaking, semi-analytical models of galaxy formation use oversimplified descriptions for the construction of these galaxies, and hence one does not find many comparisons of the predicted sizes with the observations. Despite that following correctly the growth of discs sizes provides the key point to understanding the build of the fundamental plane and the Tully-Fisher relation.

2.3.2 High redshift

One of the motivations to consider hierarchical models of galaxy formation is the hyper-activity of galaxies at high redshift. At the present there is a huge amount of multiband spectrometry across a wide range of redshifts. It is of special interest the case of sub-millimeter galaxies, that under the light

of the current cosmology, can only be interpreted as furious star forming galaxies around redshift two.

It also has been shown by observations that the dominant component of emission at high redshift galaxies is dust. This implies that the modelisation of chemical enrichment and dust properties is a key to the understanding the high redshift universe. This is completely the opposite to the case of the local universe where just a fair understanding of the process of star formation, or at least its timetable and bookkeeping, is enough to interpret the observations. This is one of the most stringent test of high redshift galaxy formation will be covered under the chapter of infrared galaxies.

On the other hand one can also classify as high redshift observations the following

- **Chemical evolution**
  The pollution of the interstellar medium by supernova explosions can be followed and tracked inside the models. Observations in clusters from the X-ray emitted by the hot gas give an idea about the metal enrichment. That provides another test to pass, that is not a very robust one given all the freedom in galactic chemical evolution models.
- **Cosmic star formation history**
  The Madau plot [MFD$^+$96] have become a fashionable thing during the last years. It is a very useful way to keep a bookkeeping of the star formation history of our universe and the constraints on different models of galaxy formation and evolution. It is a must for every person with its galaxy formation model.

2.3.3 Bimodality

Observations highlight a robust bimodality in galaxies, with a clear division around the characteristic mass of $3 \times 10^{11} M_\odot$ which in the standard picture of galaxy formation, corresponds to halo masses $1 \times 10^{12} M_\odot$. In short, less massive galaxies tend to be blue and star forming, and massive galaxies tends to be red with a population of old stars. It is important to note this also a color bimodality (with different threshold mass) has been found also in the colors of galaxies located in very underdense regions [Tik07].

This bimodality can be seen in a wide set of observations, which overlap with some of the points we mentioned in the section recapitulating the low redshift observations. We follow the presentation of the Dekel and Birnboim [DB06], when selecting the most relevant set of observations pointing towards the bimodality.

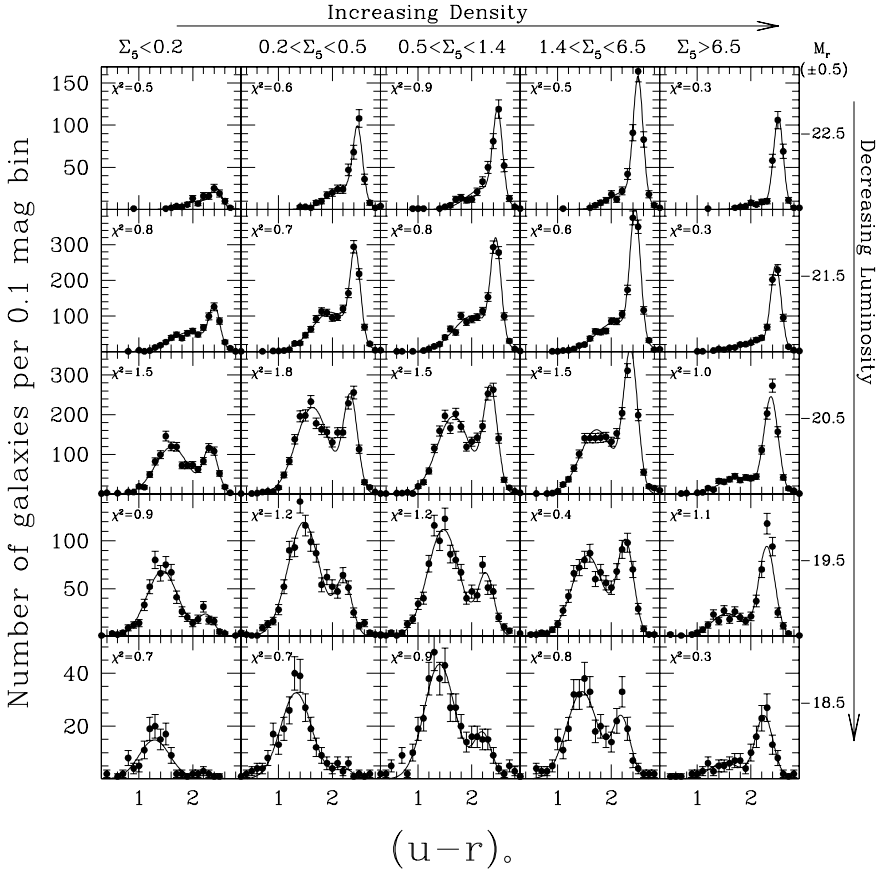


Fig. 2.5. Figure taken from [BGB+04] quantifying the galaxy color bimodality. It shows the color distribution for the indicated 1 mag interval (right axis) and the range of local projected density, in units of Mpc^{-2} (top axis). The solid line is a fit of a double Gaussian. On the top right part of the plot (luminous, crowded environment) shows only a peak on the red side. The bottom left part of the plot (faint, sparse environment) shows the peak on the blue side.

- **Luminosity function**
  Blue galaxies dominate the stellar mass function below the characteristic mass, while red galaxies takeover over this mass [BGB+04, BMKW03].

- **Color magnitudes**
  In color magnitude diagrams a color bimodality can be spotted quite robustly, where galaxies are divided into blue sequence and the red sequence separated by gap [BGB$^+$04, BEH$^+$05].

- **Star formation rate**
  Less massive galaxies are dominated by young populations while the more massive galaxies are dominated by old stars, in agreement with the color bimodality [KHW$^+$03, MCC$^+$03].

- **Bulge to disc ratio**
  The red sequence seems to be dominated by spheroids, while the blue sequence by discs [KHW$^+$03, BEH$^+$05].

- **Environmental dependence**
  Distributions in color and star formation rate depend strongly on the galaxy density around 1 Mpc. Young, disky populations are dominant in low density environments, while the old red and spheroidal sequence is dominant in high density environments [HBE$^+$03, KWH$^+$04, BEH$^+$05, BEM$^+$04]. This correlation is stronger than the well-known morphology density relation [Dre80].

- **Halo occupation**
  In typical models of halo occupation distribution, or more detail semi-analytical models, halos less massive than $\sim 10^{12}\mathrm{M}_\odot$ typically host one dominant galaxy, while more massive halos tend to house groups and clusters of luminous galaxies [YMW03, AZZ$^+$05, KBW$^+$04].

- **Dust Lanes**
  In low mass edge-on disc galaxies dust lanes are largely absent, while remaining a notable feature of higher mass disc galaxies. The dust lane phenomena disappears abruptly when the rotational speed of the galaxy is lower than 120km/s, this speed roughly correspond to the transitional bimodality scale [DYB04].

2.4 Our model

In this section we review the important points in our semi-analytic model of galaxy formation, as well the software we use to build mock catalogues from the outputs of the model, which consists of the same cosmological box seen at different stages of its evolution.

2.4.1 Galics

This semi-analytic model of galaxy formation have been under development during the last ten years, it has been successful in reproducing some observations. Its strong points in comparison to other semi-analytic models are in its modeling of the dust emission and inclusion of spatial information for the distribution of galaxies, which allows us to simulate observations, and not only report luminosity functions from a box of galaxies. For a detailed pipeline overview of the `GalICS`codes, see the Appendix.

We summarize now the most important points and ingredients in our model of galaxy formation.

- **Dark matter description**
  The process of fusion of dark matter halos is followed through N-body simulations. Detection of the halos is done through a friend of friends algorithm. The detected halos or later used to build the merger tree.
- **Star formation efficiency**
  The parameter α in Eq. 2.8 describes the efficiency of star formation. This is one of the most influential parameters in a model, as it not only changes the amount of stars but also indirectly regulates the amount of the gas through the feedback processes.
- **Feedback efficiency**
  The parameter ϵ in Eq. 2.9 describes the dragging efficiency of cold gas by the hot gas at the moment of galactic winds. Rising ϵ implies an ejection of the gas from the galaxies and the halos, leaving less material to form stars.
- **global AGN feedback mass**
  The inclusion of feedback processes that affect the bright end of the luminosity the function is done through this parameter. Given that we do not follow the growth of black holes, we cannot make an estimation of the energy input from an AGN. Instead, we use the correlation between bulge mass and black halo mass to say that if the total bulge mass inside a dark matter halo is greater than a given threshold, the AGN is going to stop the cooling in that halo. The values are selected to match the behavior of the bright end of the luminosity function.

Secondary elements in the model include the probability to have satellite-satellite collisions and the parameter defining the transition between what is considered as minor or major merger.

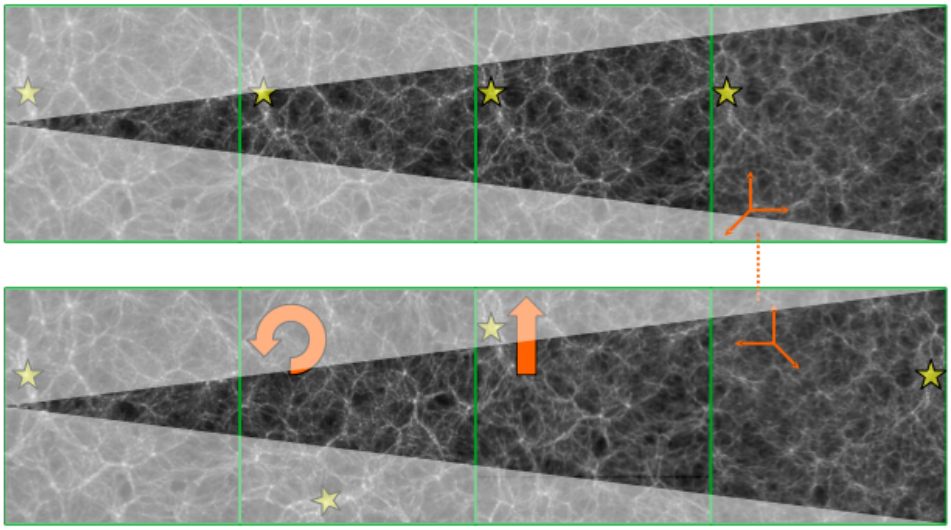


Fig. 2.6. Illustration of the random tiling technique used in momaf. Each box represents a different snapshot. The observation cone is extracted from the tiling. To avoid replication of structures (the star in the upper tile) one can randomize the disposition of each snapshot (as depicted in the lower tile) or extract the cone from an inclined line of sight. Figure prepared by Jeremy Blaizot.

2.4.2 Momaf

The only way to improve models of galaxy formation relies on the comparison with observations. But having in mind that the outputs from the model are simply snapshots of the Universe, the only quantities that one could compare are the luminosity functions at different epochs, which would leave out a huge amount of information out, specially spatial correlations.

The best approach is to attempt the construction of mock catalogues from the output of the simulations at different times. Several issues must be addressed to weight the effects of using a finite the volume. This has been already addressed by Jeremy Blaizot and its software package `MoMaF`(Mock Map Facility) that generates observing cones from the outputs of galaxies in cosmological simulations. A detailed discussion of the limitations of the method is found in a series of papers [BWG+05, BGD+04, BSC+06], which basically justify the limit on the number of statistically independent catalogues one can make as a function of the simulated Field Of View (FOV) size.

The production of the mock catalogues is based on the random tiling technique, where the snapshots are piled up to extract an cone of galaxies. Additional rotations, translations and inversion of the underlying snapshots

are used to decorrelate one volume from another. This is illustrated in the Figure 2.6

2.5 Our results

2.5.1 Geometry of Merger Trees

The problem

The Universe, as depicted by numerical simulations, shows us that the growth of structure is hierarchical. Smaller structures of dark matter form first, and merge to build bigger structures.

The description of this process has been usually done through the merger tree structures, which in turn serves as the fundamental input data for semi-analytic models of galaxy formation. This explains the great amount of research work done to date to generate these merger trees in fast and reliable ways.

As we have seen, there are two principal ways to build such trees. Analytical methods based on Monte-Carlo simulations, and fully numerical methods based on N-body simulations. In the absence of a theory to describe the nonlinear growth of structure, the results of numerical simulations are considered as the standard to compare the analytical results.

The comparison between analytical trees and numerical trees starts with, in our opinion, a harsh simplification. They just take the most massive branch and verify that the timeline of growth of that branch is comparable in the two approximations. Other comparisons simply measure the mass distribution (number of halos per unit mass) of all the progenitors .

We consider that this amputation of the richness in the structure of merger trees is actually unjustified. Is there a simple way to describe a merger tree without loosing all its complexity? Would it be useful?

Our proposition

As an answer to the first question we proposed a translation of the information in the merger trees into an easier to handle one-dimensional structure.

The translation is based on the duality that exists between a merger tree and the contour process around that tree. The illustration of that fact can be seen in the Fig.1 in the appendix.

We applied this duality to the merger trees constructed with one of the largest dark matter simulations run to date - the Millennium Run [SWJ$^+$05]. Then, we explored simple statistics such as the longitude of the walk and the action, which a useful way to measure the waviness of the walk.

At the same time we applied the same measurement to trees constructed with a so-called Markovian approach, using a publicly available code. The code was calibrated to yield similar statistics as the trees in the Millennium Run, providing us with an interesting set of trees which could serve as a meaningful point of comparison.

From this geometrical description we found a transition in the structural properties of the tree as a function of the final halo mass. This transitional scale (located for a halo mass $\sim 10^{12}h^{-1}$ $M_\odot$) is clearly spotted in the Millennium trees, but is barely hinted in the Markovian trees.

We explored the dependence on resolution of this transitional mass-scale using the Markovian trees. We produced merger trees at lower resolution, meaning that the minimal mass for a progenitor was 8 times bigger that the minimal mass in the first run, which mimicked the Millennium Run resolution. We find that the transitional scale is not a strong function of the mass resolution in the tree. We do believe that the existence of this scale can only be explained by the existence of a minimal mass for a progenitor.

Nevertheless, this minimal mass can have a physical explanation. Dark Matter halos must achieve a minimal mass in order to harbor a galaxy. Conveniently enough, this minimal mass estimated from numerical simulations and theoretical considerations, is on the same order as the minimal mass in the Millennium Run. So, in the end, this transitional scale might have a physical meaning.

Work in progress

Currently we are exploring finer statistics of the contour walks, which could describe at the same time the geometrical information of the tree and the masses for each node in it. This is possible because the construction of the contour walk implicitly uses the information of the masses in the tree nodes (see the Appendix).

2.5.2 Predictability of SAMs

The problem

The existing codes of semi-analytic galaxy formation have achieved a good level of internal consistency where they are able to reproduce a wide range of observational constraints, though not all of them. Nevertheless, the numerical robustness of these models has not been studied in the literature.

Some effort has been put into the exploration of semi-analytic codes and numerical simulations, to compare if they yield the same results under simple conditions. Even if this is useful for the study of galaxy formation, this

is useless from the point of view of studying the semi-analytical models themselves and understanding, for instance, how errors propagate.

The understanding of the overall performance, in the numerical sense, of SAMS would be greatly simplified if we had a way to measure this across different semi-analytic implementations of different groups. This is a conundrum because we do not have a way to measure the code's performance. For instance, if we wanted to compare hydro solvers, we could use analytic solutions for simple cases or convergence tests in order to qualify the performance of one code. We do not have any of these equivalents available for semi-analytic codes.

Furthermore, after the success of SAMs in the reproduction of the local luminosity functions over a wide range of wavelengths, it is becoming fashionable to select subsamples of galaxies in order to make predictions. We think that a previous step to being confident in such selection is to understand the biases and complications introduced by the semi-analytic technique.

Our solution

We propose a framework to make performance measurements for semi-analytic codes. The core of this protocol comes from a perturbative measurement of the code.

Generally speaking, there are only two common elements among all the semi-analytic codes available:

- The input merger tree, describing non-linear dark matter dynamics.
- The scalar inputs, controlling the baryonic processes of the model.

We introduce small variations in the scalar inputs of the model, and measured how this affected the output, i.e. individual galaxies. We focused our study on the effect of the variation in the parameters that regulate more directly the physics of star formation: the efficiency of star formation (α) and efficiency of supernovae feedback (ϵ).

We found that for high mass halos, with a rich and complex merger history, small fluctuations on the input parameters can have very noticeable effects in the final output galaxy. And even more, these fluctuations seem to be random. For a low mass halo, which grew with almost zero mergers, the properties of its central galaxy vary in a smooth fashion.

We quantified this behavior through an objective scalar function we call **predictability**, which is upper bounded by 1. It reaches values ~ 1 for the case of galaxies in low mass halos, and negative values for the case of galaxies in high mass haloes. Its meaning is related to the smoothenes of the

transition in the results of the code to small changes in the input parameters. Smooth transitions are represented for values close to one.

The predictability could be helpful to quantify semi-analytic models, regardless of their internal details. The definition of such objective scalar function allows the creation of **standardized tests** for semi-analytic models which at the end could report values for this function.

The big question around these results is why central galaxies in big halos have an highly fluctuating/unpredictable behavior. The answer must be composed by parts of three different, yet related, sources:

- The hierarchical building of galaxies, it provides a framework inside which is reasonable to think of small fluctuations that are accumulated for the most hierarchical objects, having as a final results two different results for similar initial conditions.
- The semi-analytic prescriptions, which in general can be noisy. For instance, the selection of the central galaxy in halo is determined by a ranking in mass. The most massive galaxy always sits at the center. This is ambiguous when various galaxies inside a dark matter halo have similar masses, in that case the selection of the central galaxy might be subject to noise
- Our semi-analytic prescriptions lead intrinsically to low predictability values for high halo masses.

Which part is more important to explain the result?

Work in progress

A detailed presentation of all these results is done in the Appendix. The work in progress moves around finding an answer to the question if there is something wrong within the semi-analytic framework. In order to answer it we must step out from our code and either make comparisons with other semi-analytic codes, or try to make the same kind of study with full hydrodynamic simulations. The first option requires great political power to move very different groups of research and try to make some standard tests using of the differing codes of galaxy formation. We have preferred to explore the second one, which only requires computational power.

3

THE INFRARED UNIVERSE

Following the discovery by the instrument FIRAS† (on board COBE ‡), in 1996 it was announced the detection of a cosmic infrared background with an unexpected high intensity, interpreted as the radiation output from all the galaxies in all the universe through all its history.

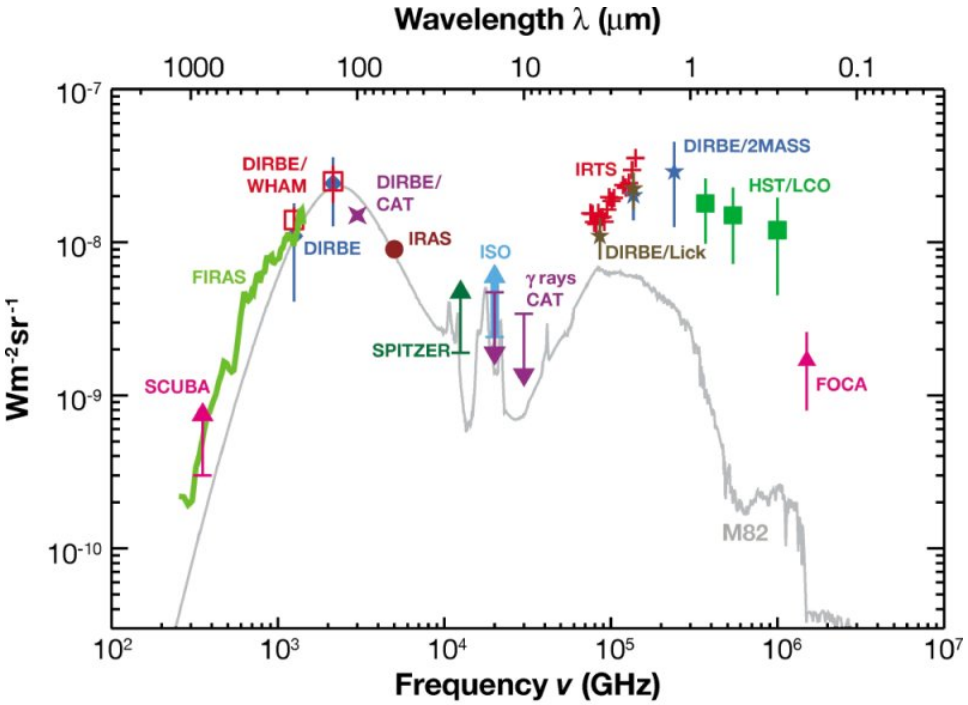
The cosmic background due to extragalactic sources (Figure 3.1) shows two maxima. One in the optical, one in the far infrared, both with roughly equal brightness. Thus, from an observational perspective, it can be stated that in order to understand the formation and evolution of galaxies it is often of equal importance to understand the optical and the infrared emission.

Further observational studies have shown that another generalization can be done, strong optical emission is mainly correlated with local galaxies, while the infrared emission is particular to high redshift galaxies.

Looking from the perspective of someone working inside the standard hierarchical model of galaxy formation, infrared galaxies provide one of the most basic tests that must be passed if one wants to argue that the standard model of structure formation gives a correct model for our universe at high redshift.

In this section we will review the physical processes that are relevant for modeling infrared galaxies, the infrared observations available that will provide us with data to match, to finally pass a quick view on the theoretical efforts that have been done to date to explain the observations. In the last part we present the detailed characteristics of our infrared model and our results.

† Far Infra Red Absolute Spectrophotometer
‡ Cosmic Background Explorer



Lagache, G et al. 2005
Annu. Rev. Astron. Astrophys. 43: 727–68

Fig. 3.1. Extragalactic background from the near UV up to millimeter wavelengths. The points are the reported constraining measurements. For comparison is shown the spectral energy density of the typical starburst galaxy M82, with bolometric luminosity $\sim 3 \times 10^{10} L_{\odot}$. Image taken from [LPD05]

3.1 Physical Processes

The dominant physical process in the infrared is dust emission. More precisely, it is the process of remission of energy absorbed by dust. This implies from the very beginning that we have to understand two parts. On one side, the sources that are radiating their energy on the dust, and second, the dust that is going to absorb that energy to remit it in the infrared.

3.1.1 Heating

We consider now the source of radiation that can heat the dust, safely restricting our study to two options: stars and Active Galactic Nuclei (AGN). Star formation as we saw in the previous chapter, is far from being understood. Nevertheless, using empirical correlations of star formation, we can

guess the stellar composition of a galaxy, which allows us to calculate an energy output. The case for AGN is a little bit more fuzzy, as it does not imply a specific physical process. AGN are usually classified by their great variety of observational properties. However, specific spectral properties of AGN (such as high X-ray emission compared to stars) and their point-like emission can allow an observational differentiation between these two kinds of heating and observations seems to favor the picture in which most of the heating is done by the star forming process and not by AGNs.

3.1.2 Extinction

The extinction properties of our Galaxy are rather well-known, the average optical depth perpendicular to the disk of our galaxy is typical of any spiral galaxy. However, for galaxies in general this is more complex, as one must take into account the particular geometry distribution of the dust in the galaxy with respect to the stars and the chemical composition of dust.

To simplify the geometry, galaxies are presented as an homogeneous mix of dust and stars, and to estimate the optical depth one uses observational fits from local galaxies as a function of metallicity and the mean column density of gas.

In `GalICS`, following [GR87] we assume that the mean perpendicular optical depth of a galaxy at wavelength λ is

$$\tau_\lambda^z = \left(\frac{A_\lambda}{A_V}\right)_{Z_\odot} \left(\frac{Z_g}{Z_\odot}\right)^s \left(\frac{\langle N_H \rangle}{2.1 \times 10^{21} \text{atom cm}^{-2}}\right), \tag{3.1}$$

where the mean hydrogen column density (accounting for the presence of helium) is given by:

$$\langle N_H \rangle = \frac{M_{cold}}{1.4 m_p \pi (a r_{1/2})^2} \text{atom cm}^{-2}, \tag{3.2}$$

and a is calculated such that the column density represents the average (mass-weighted) column density of the component, and is 1.68 for disc, 1.02 for bulges and starbursts. The dependence on the metallicity is calculated based on interpolation from the solar neighborhood and the Large and Small Magellanic Clouds, with $s = 1.35$ for $\lambda < 2000$. The extinction for solar metallicity $(A_\lambda/A_V)_{Z_\odot}$ is taken from [MMP83].

For the spherical component we used a formula where the scattering is taken into account via the dust albedo, ω_λ

$$A_\lambda = -2.5 \log_{10} \left(\frac{a_\lambda}{1 - \omega_\lambda + \omega_\lambda a_\lambda}\right), \tag{3.3}$$

where $a_\lambda = \sqrt{1-\omega_\lambda \tau_\lambda^2}$. We assume that dust and stars are distributed together uniformly and thus there is no need to average over inclination angle, as the bulge is spherically symmetric.

For the disc, we use a uniform slab model [GR87] for the extinction as a function of the inclination angle, θ:

$$A_\lambda(\tau,\theta) = -2.5\log_{10}\left[\frac{1-\exp(-a_\lambda \sec\theta)}{a_\lambda \sec\theta}\right]. \tag{3.4}$$

To calculate this quantity, we pick a random inclination of the galactic disc to the line of sight of the observer. For comparisons of the same galaxy at different epochs, and for examining inclination corrected statistics (e.g. the Tully-Fisher relation), we also compute and store the face-on magnitudes of the disc ($\sec\theta = 1$).

The final result is the extinguished stellar spectrum is then described by

$$F_\lambda = F_\lambda^\star \mathrm{dex}(-0.4A_\lambda). \tag{3.5}$$

3.1.3 Reemission

The information that defines the emission properties of dust are the size and chemical composition of the dust grains. The dust models that are now widely accepted include a mixture of amorphous silicate grains and carbonaceous grains. With a wide size distribution ranging from molecules of dozens of atoms to large grains of about 0.1 microns in diameter that can be coated with ice.

The approach we have followed in our model is based on phenomenological estimations of the Spectral Energy Distribution (SED) of infrared galaxies. Under this approach one needsfirst to collect data over a wide range of infrared luminosities (10^8 to 10^{12} $L_\odot$) and morphologies.

Usually these models use three dust components: large "classical" grains that are in thermal equilibrium, small grains heated semi-stochastically and Polycyclic Aromatic Hydrocarbons (PAHs) stochastically heated by individual ultraviolet or optical photons. The shape and normalization for each component is fixed by only one parameter: the intensity U of the interstellar radiation field, normalized such that $U = 1$ for the local interstellar radiation field. The range of variation of the models is $0.3 \leq U \leq 10^5$.

As the total emission from the galaxies comes from a superposition of different environments and components (each one composed by a sum of large grains + small grains + PAH) one creates the final infrared spectra

from the combination of each SED assuming a power-law distribution in a given galaxy dust mass over heating intensity:

$$dM_d(U) \propto U^{-\alpha} dU, \tag{3.6}$$

where the $M_d(U)$ is the dust mass heated by a radiation field at intensity U, and the exponent α is a parameter that represents the contribution of the different local SEDs. Varying the α parameters is equivalent to give different weights to active and quiescent regions in the galaxy. Then, all the needed normalizations in the model are constrained with the set of infrared observations we gathered at the beginning. The range values for the exponent α are constrained to be $1 \leq \alpha \leq 2.5$.

At the end of the day, for every value of α we have a total energy input U_T that will be associated with one infrared spectrum.

In our model we use the spectra from [DH02] (Fig. 3.2). These SED use the approach we have just described, with some modifications on the emissivity of large grains to get the models in better agreement with data beyond 100μm, using an empirical calibration to reproduce infrared *ISO* and sub-millimeter SCUBA data.

Summarizing, we calculate first the stellar component of the spectrum and the total energy output in the optical wavelengths, then from the estimation of the amount of dust (from the mass of gas) we calculate an optical depth which will tell us how much energy from the stellar component will be absorbed. With that information (the amount of obsorbed energy) we select an spectrum from the Dale& Helou model. We put together the extincted stellar spectra and the IR Dale& Helou IR spectra to build the SED of our galaxy.

3.2 Infrared Observations

Infrared observations provide tests at high redshift (z¿1)in the same way as local optical observations. We make a chronological presentations of the most relevant infrared observations during the last $\sim$ 25 years.

3.2.1 1983

The beginning of the extragalactic infrared astronomy can be safely set with IRAS (InfraRed Astronomical Satellite), the first infrared telescope, launched in 1983. It covered 96 percent of the sky in four filters centered at 12, 25, 60 and 100 μm down to 0.5 Jansky [NHv$^+$84]. During this mission

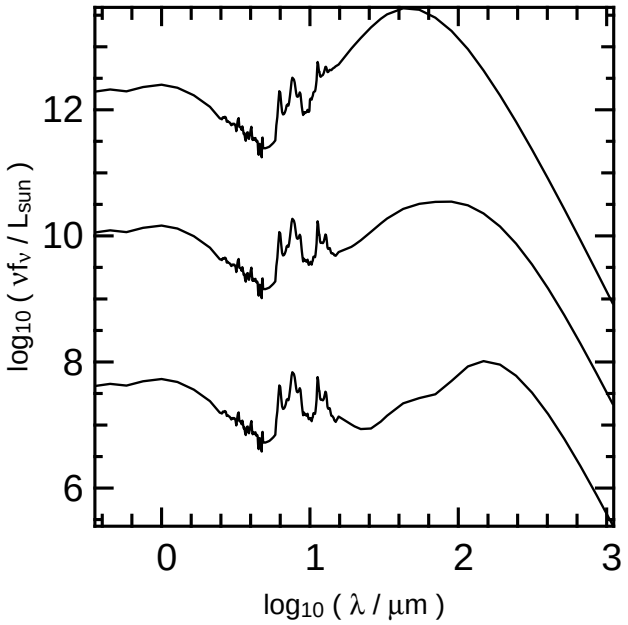


Fig. 3.2. Infrared spectra from [DH02], used in our model. Each spectra corresponds to a different bolometric infrared luminosity, from top to bottom, 5.5×10^{13} $L_\odot$, 6.3×10^{10} $L_\odot$ and 9.4×10^{7} $L_\odot$.

a new class of luminous infrared galaxies with infrared luminosities beyond $10^{11} L_\odot$ was discovered, they were named ULIRGs (Ultraluminous Infrared Galaxies) [SNH$^+$84].

When redshift surveys were coupled to IRAS it was found that as the bolometric luminosity increases, the relative density of luminous infrared sources rises as well, meaning that ULIRGs where rather common at high redshift. A global estimate of the local universe energy balance from IRAS observations shows that one-third of the star light is reprocessed by dust, demonstrating the importance of understanding infrared emission to determine cosmic evolution of galaxy populations.

3.2.2 1985

The continuation of the observations from IRAS was established by its successor, the Infrared Space Observatory (ISO) launched in 1985. This time it was equipped with a multi pixel camera (ISOCAM) and a multiband spectrophotometer (ISOPHOT) [ORA$^+$00].

The immediate improvement of ISO over IRAS was its sensitivity, which allowed to observe far deeper while having detailed imaging and spectroscopy from all the galaxies that had been detected by IRAS.

The interpretation of the observations from ISO had important consequences over the extinction corrected spiral morphology that can be quite different to the uncorrected morphology. ISO also permitted to pin down the mechanism of the infrared emission in elliptical galaxies. The imaging showed that it was mainly due to three things: photospheres of AGB stars, the interstellar medium and AGN. It was found that about two-thirds of sample galaxies had a mid infrared emission consistent with stellar origin, meaning that the infrared bands could be a good tracer of ongoing star formation.

ISOCAM covered pieces of the sky at a wavelength of 15 μm, extragalactic source counts coming from this instrument have limiting flux densities of around 2 mJy at 15 μm and around 1 μm at 6.7 μm. ISOPHOT data at 90 μm was also used to produce extragalactic catalogues.

Differential number counts from the survey have shown excess about of factors of three from the non evolution model predictions at the faintest end of the survey.This confirmed the strong evolution detected at 15 and 170 μm in other ISO surveys [ELDM05].

3.2.3 1997

A milestone in the observation of the dusty universe was set by the commissioning of the sub-millimeter Common-User bolometer array (SCUBA) camera in 1997. It took advantage of the atmospheric window at 850 μm and, of course, of the cutting-edge research on bolometric detectors. This instrument had the combination of field of view and sensitivity to allow the search for the sub-millimeter emission from unknown sources.

The first extragalactic surveys using SCUBA revealed a population of very active high redshift galaxies [CBIS03]. In the recent results from the SCUBA half degree survey, SHADES, it has been confirmed that the mean redshift of this population of galaxies is $z \sim 2-3$ [AHC$^+$07]. Other studies suggest that the implied star formation rates are extremely high about 100-1000 $M_\odot$/yr [AcA06].

The high star formation rates, resulting in high metallicity along with the large masses and high fractions of dust and gas along the strong clustering suggest a link to the formation of the most massive galaxies. Nevertheless, the used samples are quite small and should be used with caution.

3.2.4 2003

In the quest for higher sensitivity we go back again to a space based telescope. The Spitzer space telescope became operational at the end of 2003. With the high sensitivity of the multiband imaging photometer for Spitzer (MIPS) at 24 μm it became possible for the first time to conduct systematic studies of the infrared galaxies at higher redshifts than those achieved by IRAS and ISO. Also on board is the Infrared array camera - IRAC - which has four channels at 3.6, 4.5, 5.8, and 8 μm [LPS$^+$04]. The data from this mission have allowed a better description of which populations (as a function of redshift) construct the intensity of the infrared background.

3.2.5 2006

In a complementary approach to that of Spitzer, we have the AKARI infrared satellite. This space based mission carried out scans of all the sky with one to two orders of magnitude higher sensitivity than IRAS. The instrument aboard has four photometric bands of 65, 90, 140 and 160 μm, it is designed to detect dusty objects at wavelength near the peak of the dust emission. In general all the infrared results suffer from the same problem we have pointed out in the case of SCUBA galaxies, namely that the statistics are infered for a quite small number of objects. It has been suggested that some details of the discrepancies between different observations and models may come from field variance. Thus the importance of AKARI, it will provide important constraints because of these higher statistics and high sensitivity levels over the whole sky [MBB$^+$07]. During this year - 2007 - AKARI has reported its first results of far infrared observations of the Lockman hole [MSK$^+$07], giving source number counts at 65, 90, and 140 μm.

3.2.6 2008

Planck is a mission of the European Space Agency. It is designed to make measurements of the Cosmic Microwave Background with unprecedented sensitivity and resolution. It will be launched the next year (2008). Planck

will image the whole sky with a temperature sensitivity of $\Delta T/T \sim 10^{-6}$, angular resolution ~ 5 arcminutes and frequency coverage $30 - 857$GHz. The interesting part for the people working with infrared galaxies, is that a byproduct of the Planck mission will be a great amount of information available on the sub-mm properties of extragalactic sources, a task impossible to achieve from the ground.

However from the CMB point of view, these extragalactic sources provide a foreground contaminant that must be removed. Hence the interest of building a description of these sources, in order to control this potential source of error in Planck's cosmological results.

Together with Planck, the Herschel telescope will be launched. At that time it will become the largest infrared space telescope ever in use. It will have the following instruments available: HIFI (Heterodyne Instrument for the Far Infrared) a very high resolution spectrometer, PACS (Photodetector Array Camera and Spectrometer) an imaging photometer and medium resolution grating spectrometer and SPIRE (Spectral and Photometric Imaging Receiver) an imaging photometer and an imaging Fourier transform spectrometer. It will observe across the range of 60μm to 670μm, providing data unattainable with previous instruments.

3.3 Theoretical Predictions

There are different kinds of models of infrared galaxies. Depending on the price you are willing to pay and on your sense of adventure, you will surely find a model that will suit you well. In this section we review the two extremes in the spectra of infrared models. On one side we find the phenomenological models that play it safe and have low computational cost (albeit a limited physical description of the physics involved). On the extreme we show semi-analytic models that strive to produce infrared galaxies within an explicit cosmological context.

3.3.1 Phenomenological models

In these models one fundamental unit of description of the reality is the luminosity function. These models use purely empirical expressions for the luminosity function and its evolution. This is complemented by the template spectra of starbursts and normal galaxies [LDP03].

Usually the free parameters in these models do not have any physical explanation, they are just conveniently applied to fit the available information. The closest one can go to physical reality is in these models, is to state that

there are two kinds of star formation in the universe. One is related to a passive phase of galaxy evolution, and the other associated with a more active star forming activity, one for each template available. Fixing the proportion of those two models can provide a fit to a wide range of observations.

The caveat is the absence of a physical link to the astrophysical and cosmological context into which galaxies evolution is developed, without mentioning that the energy budget is not well constrained by comparisons with observations at other wavelengths.

3.3.2 Hierarchical models

In this family of models we start seeing large amounts of physical detail that is not present in the phenomenological approach. These models place the galaxy formation process in an explicit cosmological context. They use Monte-Carlo or numerical simulations to build the merger tree of dark matter halos, and run on top of it a semi-analytic implementation of all the physical process we reviewed in the previous chapter [GDS$^+$04].

Inside these kind of models we find two very different ways of calculating the infrared emission. On one hand we have models that use a phenomenological approach to derive the spectral signature of infrared emitting galaxies, which is our case. We only calculate the total amount of energy that will be absorbed by dust (which implied previous estimation of the absorption properties and amount of dust) and then, for a given total infrared luminosity we assign a spectrum in the infrared region. The caveat is that the spectra may be completely inconsistent with the formation history of the galaxy.

Another way to simulate the infrared emission is to be fully consistent with the derived properties of the galaxies, deriving a range of dust temperature and composition from the star formation history of the galaxy [GLS$^+$00].

In some sense, the best model inside this family (from the internal coherence point of view) will be the one using merger trees from numerical simulations and using self-consistent model for the infrared emission. Unfortunately, such a model does not exist yet, partly because of the lack of confidence on other people's results, but also because of the complete absence of open source initiatives for the codes of galaxy formation and the lack of standards around what could be considered a good semi-analytical model.

3.4 Our results

The problem

At the beginning of this chapter we stressed the importance of understanding infrared emission based on the argument that in the history of the universe there is as much energy emitted in the infrared than energy emitted in the optical. Then, we reviewed the relevant elements to obtain a theoretical description of galactic infrared emission. Finally, we reported the main results of almost twenty years of infrared observations.

In a nutshell, the observations state that locally the emission of galaxies is dominated mainly by optical wavelengths, and at high redshift it is dominated by the output in the infrared wavelengths. When you go further in redshift to longer wavelengths you have the strongest energetic constraints.

Furthermore, the strong evolution observed for infrared galaxies is one of the strongest arguments in favor of the hierarchical picture. In particular the observations at 850μm can only be interpreted as hyperactive star-forming galaxies at redshifts $z > 1$. The logic behind it is that during the process of merging, strong starbursts are triggered. The starburst embedded in the gas and dust of the galaxy will cause that almost all the radiation from star formation will be reprocessed by dust and re-emitted in the infrared.

Semi-analytic models of galaxy formation should in principle be able to reproduce easily these observations. However, this is not the case. In fact one of the most difficult observational tests for SAMs comes with the galaxies at the far infrared and sub-millimeter wavelengths. In particular, the counts at 850μm from SCUBA are very difficult to reproduce for semi-analytical models. The predicted galaxy counts are one order of magnitude below the observed counts. Our model is no exception to the difficulty.

Thus, our objective was to produce the right amount of SCUBA galaxies without breaking the rest of the model.

Our proposition

We find that in order to reproduce the counts at 850μm while keeping a good match to the observations at other wavelengths (from the infrared to the optical), we must modify the initial mass function for galaxies at high redshift.

From a Kennicutt Initial Mass Function we pass to a Top Heavy IMF which only produces stars with masses greater than 8 $M_\odot$. In this way we obtain more UV luminosity for the same amount of formed stars, providing more energy to heat the dust. The reason to constrain the change for

redshifts higher that $z > 1.0$ is that we must keep under control the IR luminosity at low redshift.

We will go step by step. First we define the observations that guided our search. Those were the integral counts for the SCUBA 850μm and SPITZER at 160μm, 70μm and 24μm. This gives us a general panorama from the sub-millimeter to the infrared. In order to keep in order the energetic and baryonic budget of our galaxies, we also track the results for the JOHNSON$_K$ band in the near infrared (giving us a measure of the bulk of formed stars) and in the optical through the SDSS$_r$ luminosity function.

In the Fig.3.3 we overplot the results for two models. The solid line represents what we call our standard model, and the dashed lines represent the results of the standard model including a top-heavy IMF in the galaxy components with a short dynamical time $t_{dyn} < 100$ Myr, all the rest being equal.

If we focus our attention in the upper left panel, showing the SCUBA data, we see that the standard model is almost one order of magnitude below the observations (represented by the symbols with error bars). There is also an overall agreement with the rest of the data, except maybe for a too bright luminosity function in the SDSS$_r$ filter (lower left panel).

If we switch on the top-heavy IMF (dashed-line), the SCUBA counts are finally reproduced. Moreover the JOHNSON$_K$ and SDSS$_r$ data are better fitted. The problem is that the model explodes for the rest of infrared wavelengths between 24μm and 170μm.

We know from observations, and also from the results of our model, that the SCUBA galaxies are located around redshifts of $z \sim 2$. We also know that the galaxies seen with SPITZER at 160μm, 70μ and 24μm are located for redshifts $z < 1$. So, what if we just turn on the top-heavy IMF at redshifts $z > 1$?

In Fig.3.4 we show the result of this experiment. This time the solid line represents the standard model with the top-heavy IMF everywhere, and the dashed represents the model with the top-heavy for redshifts $z > 1$ with an additional modification on the threshold mass for stopping the cooling in massive halos. The last modification is motivated in such a way to have a good fit of the SDSS$_r$ and JOHNSON$_K$ data.

We end up with a satisfactory model which presents good results from the sub-millimeter down to optical wavelengths. But all this was calculated from a cosmology compatible with WMAP first year data (WMAP1 hereafter), and today we have access to the new set of cosmological parameters from the third year data (WMAP3 hereafter). The biggest difference comes from the parameter σ_8, which measures the density fluctuations at present in a

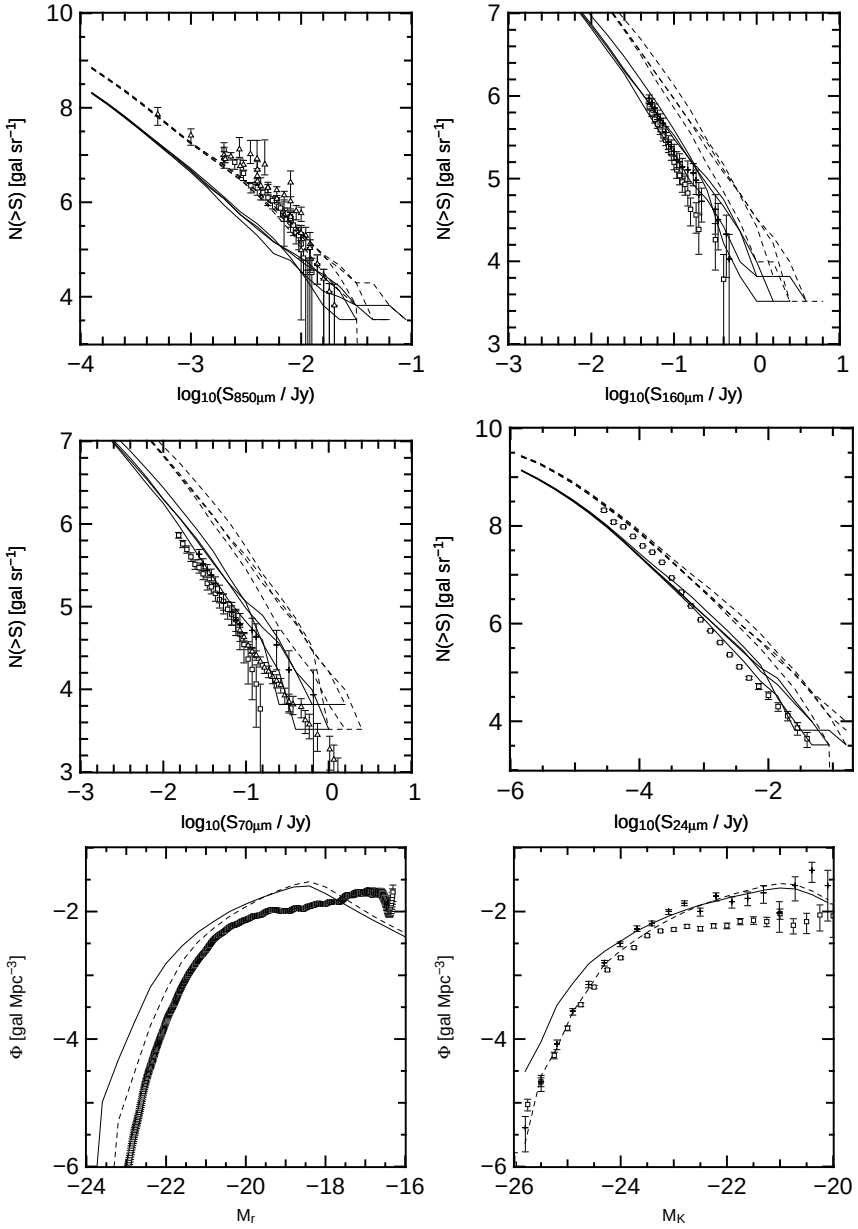


Fig. 3.3. Plots showing the results for our standard model (thin solid lines) and the model with a top-heavy IMF (thin dashed line). The other symbols represent the observed data. From left to right and top to bottom, the datasets are: integrated counts for SCUBA 850μm, SPITZER 160μm, 70μm and 24μm. The lines for the models in this case come from various runs of `MoMaF`which produced mock observations of size $1^\circ \times 1^\circ$ and depth up to $z \sim 4$. The last two panels correspond to the luminosity functions in the SDSS$_r$ and JOHNSON$_K$ filters. The model luminosity function is constructed from the $100h^{-1}$ Mpc side cube of galaxies at redshift $z = 0$.

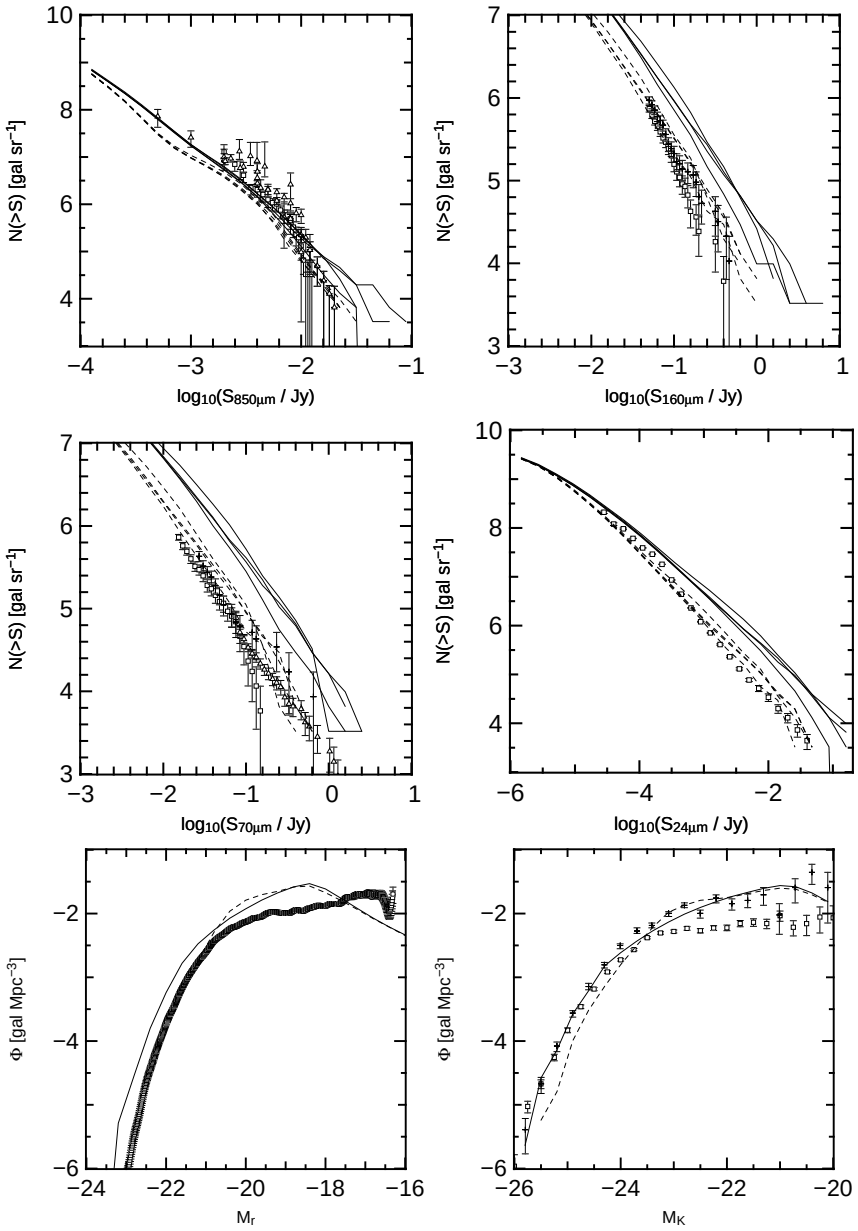


Fig. 3.4. Same as Figure 3.3 for the model with a top-heavy IMF at all redshifts (thin solid line) and the model with a top-heavy IMF at $z > 1$ (thin dashed line). Read the main text for a comment on these results.

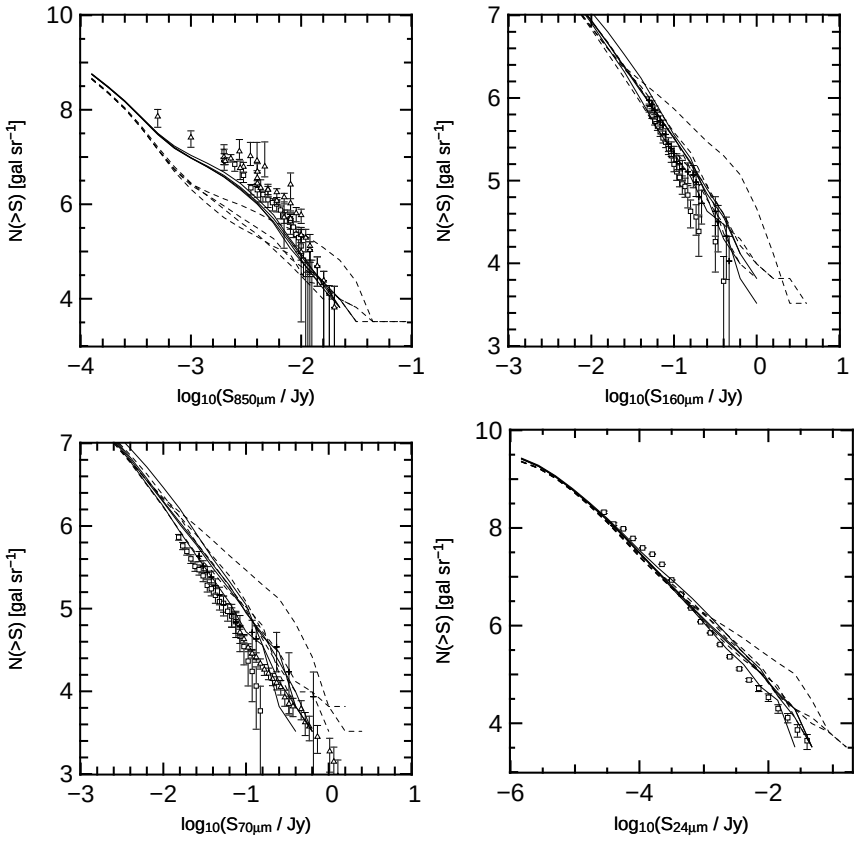


Fig. 3.5. Integrated counts for two models. Thin solid line: model with a top-heavy initial mass function at $z > 1$ and WMAP1 cosmology. Thin dashed line: same model in a WMAP3 cosmology. For a comment on the results read the main text.

sphere of 8 Mpc in radius. From the WMAP3 data σ_8 is lower than the value deduced from WMAP1 data, going from 0.9 to 0.72. The overall effect of this change in σ_8 is that the formation of structure is delayed, an effect that is stronger as we go to higher redshifts.

In Fig.3.5 we report the results for the infrared counts. The solid line is our best model in a WMAP1 cosmology, the dashed is exactly the same model in a WMAP3 cosmology. In the WMAP3 cosmology, we lost again the SCUBA counts, while the other wavelengths remain the same. As we expected, only the high redshift populations are strongly affected by the change in the cosmological parameters between WMAP1 and WMAP3.

It is possible that we could add up new elements to the model, like including some special kind of ultra-cold dust which could help to boost the infrared luminosity. At any rate, the final message one should keep in mind is that our understanding of the high redshift universe is very far from being clear and complete. Of course any further understanding won't come from the models itself but from the data of missions like Planck or Herschel.

Work in progress

We are proceeding with the capitalization of the good match we have to the data (using a WMAP1 cosmology) to make predictions for the Planck and Herschel missions. We are also working on playing a part in the discussion of a WMAP1 vs. WMAP3 normalization, based on our results for SCUBA galaxies, a perspective that has not been addressed yet in the literature.

4
WEAK GRAVITATIONAL LENSING

The gravitational deflection of light was first discussed more than 300 years ago by Newton. Today this effect is the most powerful cosmological tool available to observers seeking the comprehension of the structure of our universe. To arrive to this point we had to wait for Einstein to build a more correct theory of gravitation than that of Newton's, without forgetting the impressive developments in astronomical instrumentation.

Perhaps the most popular example of deflection of light in an astronomical context are those that recall the effect of a magnifying glass, where very distorted images of galaxies take the shape of arcs following the curvature of the lens. This is a typical example of what we know as strong lensing. However, the powerful tool we refer to in the beginning is related to the case where the light rays are slightly deflected, generating distortions that are generally very difficult to detect in individual objects, and can only be verified on a statistical sense over an ensemble of images. This receives the name of weak gravitational lensing.

The first calculations that aimed to exploit the effect for cosmological purposes, were done in the early 90s by theoreticians. Further comprehension of the power of this tool had to wait for numerical simulations to extend the analytical work. Finally, it was only at the beginning of the 21st century that weak gravitational deflection of light on cosmological scales could be detected. Few years later it was providing tight constraints of cosmological parameters, and today it is aiming to produce 3-D maps of the matter distribution in the universe.

To attain this ambitious goal observers have to overcome a great deal of problems coming from systematics in the observations that have to be corrected and uncertainties in the models of galaxy formation. All this calls for a general framework that can include simulations of galaxy formation

and weak lensing observations to help in the task of enhancing all the power that weak lensing has to offer.

In this chapter we present the first step in that direction. We offer first a review of the basic concepts in gravitational lensing, emphasizing the importance of weak lensing in a cosmological context. Then we explain the principal observational concepts around weak lensing and its current limitations. The next to last section explains in detail the most common technique to perform weak lensing simulations.

Finally we explain the importance of understanding the coevolution of galaxies and dark matter structure for the next generation of weak lensing surveys, in order to present the computational tool we have constructed to tackle this problem.

4.1 Theoretical Bases

To present the basic concepts of gravitational lensing we will follow a route of increasing complexity. We start with the simplest example that can illustrate the most relevant concepts, the case of one point source and one lens. Then we extend our discussion to describe the case of an extended source. Finally, we jump to the case of an extended source in an explicit cosmological context.

4.1.1 One Lens, Point Source

Physically speaking, the results we present here are based on two assumptions. First, that the properties of the lenses are obtained in the limit of a slowly varying gravitational field. Second, the lensing object is considered to be thin, meaning that the deflecting mass is bounded to region of size much smaller than a distance between source, deflector and the observer. All these approximations hold very well in the astrophysical case of galaxies.

The presence of the intervening lens between the light source and the observer has two main effects. It can change the brightness of the source and its apparent position on the sky. We do not include changes in shape as we are restricting the discussion to the case of a point source.

We want to find a relationship between the intrinsic angular position of an object in the sky in the absence of the lens to its observable image on the sky in presence of the lens. This relation is known as the lens equation, and can be set-up in a more precise form by taking a look at the geometry of the problem. To follow more clearly the argumentation, we present the

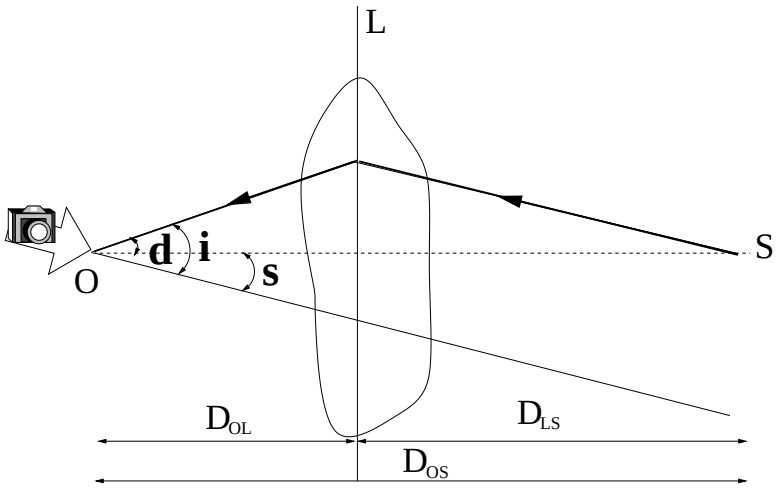


Fig. 4.1. Spatial configuration for the system describing the lensing of a point source S by one lens L and the observer at O.

system in a diagram (Fig. 4.1), which reminds us of the settings in a optical laboratory.

On one side we have the observer, in the middle we place the lens (the perturbing mass) and on the other extreme the source. In the conditions we specified at the beginning (slowly varying gravitational field, thin lens) we can sketch the deflection of the light coming from the source, as happening at the lens plane which is perpendicular to the line joining the observer and the source.

From this simple geometrical setup we can obtain the observed position on the sky as function of the deflecting angle and the original position of the source of the sky

$$D_{OS}\mathbf{s} + D_{LS}\mathbf{d} = D_{OS}\mathbf{i}, \tag{4.1}$$

where **s**, **d** and **i** (**s**ource, **d**eflection and **i**mage) are 2-d vectors representing angular variables over the lens plane, and $D_{...}$ mark the distances between the observers (O), the lens (L) and the source (S).

We still have to express the deflecting angle as a function of the physical properties of the lens. We do that integrating the perpendicular deflection as a function of the gravitational potential

$$\frac{\mathbf{d}}{D_{OL}} = 2\int_{-\infty}^{\infty} dz \left[\frac{^{(2)}\nabla\phi}{c^2}\right], \tag{4.2}$$

where ${}^{(2)}\nabla$ is the gradient in the plane perpendicular to the trajectory. Taking now the two dimensional divergence over the same plane and using the Poisson equation, we can write

$$ {}^{(2)}\nabla \cdot \mathbf{d} = \frac{2D_{OL}}{c^2} \int_{-\infty}^{\infty} dz 4\phi G\rho = \frac{8\pi G D_{OL}}{c^2}\Sigma, \tag{4.3} $$

where Σ is the surface mass density corresponding to ρ. This equation has the solution

$$ \mathbf{d}(\mathbf{i}) = \frac{4GD_{OL}}{c^2} \int d^2x \Sigma(\mathbf{x}) \frac{(\mathbf{i}-\mathbf{x})}{|\mathbf{i}-\mathbf{x}|^2}. \tag{4.4} $$

Using this equation together with the definition of the lens equation (Eq.4.1) we can express the true source position as a function of quantities at the image plane,

$$ \mathbf{s} = \mathbf{i} - \frac{4G}{c^2}\frac{D_{LS}D_{OL}}{D_{OS}} f(\Sigma(\mathbf{i})), \tag{4.5} $$

where f is the function defined by the integral in Eq.4.4. This relationship by virtue of the Poisson equation can be recast in terms of an scaled, projected gravitational potential Ψ, so the lens equation becomes

$$ \mathbf{s} = \mathbf{i} - \nabla\Psi. \tag{4.6} $$

4.1.2 One Lens, Extended Source

The next level of complexity deals with the description of an extended source. Fortunately, we can use the same machinery we have developed. We simply note that all we have done is build a mapping from the source plane to the observer plane. This allows us to describe the lens properties through the Jacobian matrix of the lens mapping, which can also be seen as the ratio of the images at the image plane and the source plane

$$ \mathsf{A}_{ab} = \frac{\partial s_a}{\partial i_b}, \tag{4.7} $$

where i_a are the two components of the position vector at the image plane, and s_b are the components of the position vector at the source plane. From the last equation in the previous section we can write

$$\mathsf{A} = \begin{bmatrix} 1 - \frac{\partial^2 \Psi}{\partial i_1^2} & \frac{\partial^2 \Psi}{\partial i_1 \partial i_2} \\ \frac{\partial^2 \Psi}{\partial i_2 \partial i_1} & 1 - \frac{\partial^2 \Psi}{\partial i_2^2} \end{bmatrix}. \tag{4.8}$$

This kind of 2×2 matrix can be expressed **generally** as the decomposition in the following terms

$$\mathsf{A} = \begin{bmatrix} 1 - \kappa - \gamma_1 & \gamma_2 \\ \gamma_2 & 1 - \kappa + \gamma_1 \end{bmatrix} \tag{4.9}$$

$$= \mathsf{I} + \kappa \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \gamma \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix}, \tag{4.10}$$

where $\gamma = \sqrt{\gamma_1^2 + \gamma_2^2}$ and I is the identity matrix. The term κ is know as the convergence and describes the isotropic magnification of the source, γ_1 and γ_2 are the shear components and describe the image shear along perpendicular axis. The Figure 4.2 illustrates these concepts.

From the last two equations one can now write the convergence and the shear from linear combinations of second order derivatives of the the 2-d potential Ψ:

$$\kappa = \frac{1}{2} \left[\frac{\partial^2 \Psi}{\partial i_1^2} + \frac{\partial^2 \Psi}{\partial i_2^2} \right], \tag{4.11}$$

$$\gamma_1 = \frac{1}{2} \left[\frac{\partial^2 \Psi}{\partial i_1^2} - \frac{\partial^2 \Psi}{\partial i_2^2} \right], \tag{4.12}$$

$$\gamma_2 = \frac{\partial^2 \Psi}{\partial i_1 \partial i_2} = \frac{\partial^2 \Psi}{\partial i_2 \partial i_1}. \tag{4.13}$$

The inverse A^{-1} will give us the change in image size at the observed plane in terms of the image at the source plane, this inverse matrix can be written as:

$$\mathsf{A}^{-1} = \left(\frac{1}{1 - \kappa} \right) \left(\frac{1}{1 - |g^2|} \right) \begin{pmatrix} 1 + g_1 & -g_2 \\ -g_2 & 1 - g_1 \end{pmatrix}, \tag{4.14}$$

where g is the reduced shear of components $g_i = \gamma_i / (1 - \kappa)$.

At the end the important point to remember is that the deformation of the source image can be expressed in terms of second derivatives of the scaled two-dimensional lens potential Ψ.

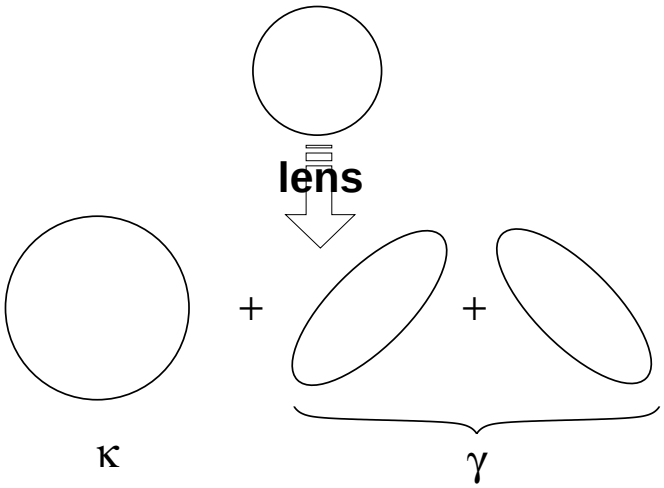


Fig. 4.2. Illustration of the meaning of the different terms in the amplification matrix on a extended image of circular shape. The convergence κ represents an isotropic magnification and $\gamma_{1,2}$ the shear.

4.1.3 Cosmological context

If we now pass to a explicit cosmological context, there is no other way to start but to state the metric into which we are working and follow the path of a photon as it propagates over it.

We will use the approximation of perturbing the Robertson Walker metric after choosing a definite gauge, meaning that there is only one degree of freedom to express the perturbation of the matter [MVVH06] (see the Appendix for details). In that case the metric is given by

$$ds^2 = a(t)^2\{-(1+2\phi)dt^2 + (1-2\phi)\gamma_{ij}dx^i dx^j\}, \tag{4.15}$$

with the spatial part of the metric that can be expressed as:

$$\gamma_{ij}dx^i dx^j = d\chi^2 + r(\chi)^2(d\theta^2 + \sin^2\theta d\phi^2),$$

where $a(t)$ is the scale factor, ϕ is the gravitational potential and $r(\chi)$ is the physical comoving angular diameter distance, which in a flat space-time geometry is equal to χ. The change in the photon's direction as it travels is governed by spatial part of the geodesic equation, which, applied to the metric equation gives

$$d\vec{\alpha} = -2\vec{\nabla}\phi d\chi, \tag{4.16}$$

this is basically the same result we found in the section 4.1.1, $\vec{\alpha}$ is change in the direction of the photon propagation on the symbol $\vec{\nabla}$ denotes the transverse derivative from the line of propagation. The deflection by $\vec{\alpha}$ at χ leads to a transverse excursion $d\vec{x} = r(\chi - \chi')d\vec{\alpha}(\chi')$, finally integrating along the photon's trajectory all the individual deflections we have

$$\vec{x}_\perp(\chi) = -2\int_0^\chi r(\chi - \chi')\vec{\nabla}_\perp\phi(\chi', \vec{x}_\perp(\chi'))d\chi' + r(\chi)\vec{\theta}(\chi = 0), \tag{4.17}$$

where the gravitational potential has to be calculated along the actual perturbed path, this allows us to write the angular direction on the sky a equation dividing by $r(\chi)$

$$\vec{\theta}(\chi) = -\frac{2}{r(\chi)}\int_0^\chi r(\chi - \chi')\vec{\nabla}_\perp\phi(\chi', r(\chi', \vec{\theta}(\chi')))d\chi' + \vec{\theta}(\chi = 0). \tag{4.18}$$

Now we can use again the differential expression of the ratio between image areas, given by the Jacobian:

$$\frac{\partial\theta_i(\chi)}{\partial\theta_j(0)} = \delta_{ij} - 2\int_0^\chi g(\chi, \chi')\nabla_i\nabla_j\phi(\chi', r(\chi', \vec{\theta}(\chi')))d\chi', \tag{4.19}$$

where

$$g(\chi, \chi') = \frac{r(\chi)r(\chi - \chi')}{r(\chi)}$$

is playing the role of the distances combinations we saw in the first section of this chapter. We can define now the analogous weighted 2-dimensional potential

$$\Psi_{ij} = 2\int_0^\chi g(\chi, \chi')\nabla_i\nabla_j\phi(\chi', r(\chi', \vec{\theta}(\chi')))d\chi', \tag{4.20}$$

which allows us to express convergence and shear in the same way as in the previous section.

But also in the case of κ, we can arrive to a much more explicit expression using also the relation between the gravitational potential and the density perturbation is given by Poisson equation:

$$\nabla^2\phi = \frac{3}{2}\left(\frac{H_0}{c}\right)^2\Omega_m\frac{\delta}{a}, \tag{4.21}$$

where H_0 is the Hubble parameter today and Ω_m is the matter density.

Using this expression we can express the convergence in terms of the density perturbation equation

$$\kappa = \frac{3}{2}\left(\frac{H_0}{c}\right)^2 \Omega_m \int_0^{\chi} g\frac{\delta}{a} d\chi'. \tag{4.22}$$

Moreover, one can express the power spectrum of the convergence field κ as a function of the power spectrum of δ:

$$P_\kappa = \frac{9H_0^4\Omega_m^2}{4c^4}\int_0^{\chi_H} d\chi \frac{g^2\chi}{a^2} P(l/r(\chi), \chi) \tag{4.23}$$

The equation 4.23 shows that the second order statistics of the convergence κ are sensitive to the matter density Ω_m, the normalization of the power spectrum of the density fluctuations, the Hubble parameter H_0 and the source distribution. In the case of a matter only universe, in the linear approximation for a power spectrum of density fluctuations $\propto k^n$ one can say that

$$\langle\kappa(\theta)^2\rangle^{1/2} \sim 10^{-4}\sigma_8^2\Omega_m^{1.6}\left(\frac{\theta}{1\text{deg}}\right)^{-(n+2)} z_s^{1.4} \tag{4.24}$$

where θ is the size of the window where κ is averaged and z_s is the mean redshift of the sources. This is an explicit example of the dependence of the convergence on the cosmological parameters.

Interlude

Before continuing we summarize the relevant points in this theoretical introduction

- All the lensing physics can be described by the **2-dimensional potential** of the lens.
- the deformation of an extended source can be expressed by a matrix of **second derivatives** of the 2-dimensional potential we just mentioned.
- the shear can be related to the **lens surface matter density**.
- In the cosmological context convergence gives us information of the cosmological density field integrated **along the photons path**.

4.2 Observations

4.2.1 Image from the shear

From the observational point of view the most common approximation is to describe galaxies as ellipses, using the quadrupole of the light distribution. If $f(\vec{\theta})$ is the 2-dimensional light distribution of the galaxy image over a plane of coordinates $\vec{\theta} = \{\theta_1, \theta_2\}$, the quadrupole is defined as

$$\mathsf{Q}_{ij} = \frac{\int d\theta_i d\theta_j f(\vec{\theta})(\theta_i - \bar{\theta}_i)(\theta_j - \bar{\theta}_j)}{\int d\theta_i d\theta_j}, \tag{4.25}$$

where $\bar{\theta}$ is the centroid if the light distribution. For an ellipse of major axis a, minor axis b and with the major axis making an angle β with respect to an horizontal reference axis, the shape matrix can be written as:

$$\mathsf{Q} = \frac{1}{\pi ab} \begin{pmatrix} a^2\cos^2\beta + b^2\sin^2\beta & (a^2-b^2)\sin\beta\cos\beta \\ (a^2-b^2)\sin\beta\cos\beta & a^2\sin^2\beta + b^2\cos^2\beta \end{pmatrix}. \tag{4.26}$$

The case of the circle is recovered by setting $a = b$ and $\beta = 0$. Supposing that the Jacobian —amplification—matrix describing the lensing does not vary much along the corresponding surface of the galaxy (which is a good approximation in the case of weak lensing caused by large scale structure) the lensed shape matrix Q' can be written as:

$$\mathsf{Q}' = \mathsf{A}^{-1}\mathsf{Q}\mathsf{A}^{-1} \tag{4.27}$$

this fixes the modification of the shape and magnitude of the galaxy, that can be expressed in terms of the following quantities

- Convergence, κ
- Shear $|\gamma| = \frac{a-b}{a+b}$
- Magnification $\mu = 1/\det \mathsf{A}$

that would mean, for instance, that a galaxy of apparent magnitude m would have its magnitude modified as $m' = m - 2.5\log\mu$.

4.2.2 Shear from the image

As we show in section 4.1.2, the convergence κ and the shear γ are combinations of the second derivatives of a two-dimensional potential with respect to the lens plane coordinates:

$$\kappa = \frac{1}{2}\left[\frac{\partial^2\Psi}{\partial i_1^2} + \frac{\partial^2\Psi}{\partial i_2^2}\right], \tag{4.28}$$

$$\gamma_1 = \frac{1}{2}\left[\frac{\partial^2\Psi}{\partial i_1^2} - \frac{\partial^2\Psi}{\partial i_2^2}\right], \tag{4.29}$$

$$\gamma_2 = \frac{\partial^2\Psi}{\partial i_1 \partial i_2} = \frac{\partial^2\Psi}{\partial i_2 \partial i_1}. \tag{4.30}$$

Kaiser & Squires in 1993 [KS93] used this relationship to obtain the convergence from the shear. We note first that the previous equations become even simpler in Fourier space (derivatives are now multiplicative factors)

$$\tilde{\kappa}(l) = \frac{1}{2}(l_1^2 + l_2^2)\tilde{\Psi}(\vec{l}), \tag{4.31}$$

$$\tilde{\gamma_1}(l) = \frac{1}{2}(l_1^2 - l_2^2)\tilde{\Psi}(\vec{l}), \tag{4.32}$$

$$\tilde{\gamma_2}(l) = -il_1 l_2 \tilde{\Psi}(\vec{l}), \tag{4.33}$$

where l_1 and l_2 are the components in Fourier space. This means that we can measure the convergence from the shear. We can now eliminate the potential $\tilde{\Psi}$ and express $\tilde{\kappa}$ as a weighted average that avoids divergences in the $l_1 - l_2$ axes:

$$\tilde{\kappa} = \frac{l_1^2 - l_2^2}{\mathbf{l}^2}\tilde{\gamma_1} + \frac{2l_1 l_2}{\mathbf{l}^2}\tilde{\gamma_2}. \tag{4.34}$$

The inverse Fourier transform of this yields immediately an estimator for the convergence. It must be noted that in general the observed quantity from the observed galaxies ellipticities is the reduced shear $g = \gamma/(1-\kappa)$ and *not* the shear γ, with $g \to \gamma$ only for the weak lensing regime.

The galaxies' ellipticities are used as unbiased estimators for the shear γ, which in turns allows a reconstruction of the convergence κ field from the approach in Eq.4.34. This is equivalent to a measurement of the correlations between galaxy shapes, providing the key to the actual measurements of this effect from observations [MVVH06].

4.2.3 Systematics

All the weak lensing measurements performed so far are limited by small fields of view. Future surveys are focusing on having a bigger field of view and reducing the instrument noise by going to space.

These future surveys still have a great number of challenges to overcome faced to the systematic effects that must be controlled. We review the main sources of systematics following the presentation of [Ref03]. In general, all these points will be addressed by the Shear Testing Program (STEP), [HVB+06] which aims to bring together all the expertise from all the researches in the field, to tackle these problems.

- **Point Spread Function (PSF) anisotropy**
  This is perhaps the worst systematic effect. The underlying problem is that the PSF ellipticities observed by all the groups are rather large, in the order of 10 percent. The objective of high precision measurements of weak lensing requires precision of 0.1 percent in shear, which calls for a reduction of two orders of magnitude in the estimation of the PSF effects on the ellipticity. It is a true technical limit and must be overcomed at every step of the measurement, from the design of the telescope to the analysis of the images.
- **Shear measurement methods**
  The present shear measurement methods are not well adapted to high precision observations. This is the most important theoretical limiting factor. Although new methods have been proposed, their accuracy has not been checked yet.
- **Redshift distribution**
  The information of cosmic shear measurements is only two dimensional. In order to make a 3-D reconstruction, or just simply deconvolve the information in order to make predictions of cosmological order, it is necessary to have constraints on the redshift of the sources. The only way out is to have dedicated spectroscopy surveys and photometric redshift studies, an expensive solution that anyway requires some knowledge on the possible values of the magnitudes and colors of high redshift ($z > 0.5$) galaxies.
- **Intrinsic correlations**
  The statistical basis of all shear measurement methods is that the galaxies' ellipticities are not correlated in the absence all the weak lensing effect. To which extent is this true? It is possible to think of physical mechanisms that produce a correlation between the ellipticities of the galaxies and the underlying dark matter distribution. The simplest one is the alignment of the disc with respect to the angular momentum of the dark matter halo.

A lot of work has been done to quantify this effect, and generally one can say that it is not important for the surveys going up to redshift one, but it can be important for shallow surveys [HRH00, Jin02]. It has also been found that these correlations are persistent over a quite large redshift range, something that could be controlled with good redshift information.

4.3 Numerical Simulations

The physical scales probed by weak lensing are found in the nonlinear regime where an analytical description is not available. Furthermore, as we have just seen, there are limiting factors coming from the galaxy properties and not much from our ignorance about the dark matter field.

The only way to describe the matter clustering at those scales is through the use of N-body simulation techniques, and to describe galaxy formation the best option is a semi-analytic model of galaxy formation. The extraction of the weak lensing signal from tese simulations is usually done using the technique called ray tracing, which basically follows the trajectory of a photon inside the simulation box.

In the next sections we will see in detail the multiple plane formalism and its use in the ray-tracing technique.

4.3.1 Multiple Plane Theory

So far we have been dealing with the continuous description of matter on cosmological scales. But in order to make numerical simulations we have to pass, at some point, into a discretized description. One way to do so, is to replace the density field from the observer to the source by a collection of particles. From that, we can put a grid over that volume and obtain the gravitational potential, to finally follow the path of photons through this clumpy distribution. It is doable, but it is expensive.

A technique that yields the same results, and is much cheaper, introduces multiple planes between the observer and the source, where the spacing between these planes is larger than the typical distance between two density perturbation. The intermediate mass is projected onto the planes, and now the photons propagate freely between the planes to get deflected at the planes only.

For convenience, the ray of light starts from the observer and is deflected up to the source plane. We can now define at every plane two dimensional quantities from the 3-D distribution, getting at the end a discrete version of the equations we have seen in the previous section. For instance for the

angular position as function of redshift, we have now the angular position at the n-th plane,

$$\vec{\theta}_n = \vec{\theta}_1 - \sum_{i=1}^{n-1} \frac{r(\chi_n - \chi_i)}{r(\chi_n)} \vec{\nabla}\Psi_i \tag{4.35}$$

$$\mathsf{A}_n = \mathsf{I} - \sum_{i=1}^{n-1} \frac{r(\chi_n - \chi_i) r(\chi_i)}{r(\chi_n)} \mathsf{U}_i \mathsf{A}_i \tag{4.36}$$

where I is the identity matrix and U_i and the Jacobian matrix at the i-th plane, is defined by:

$$\mathsf{U}_i = \begin{pmatrix} \partial_{11}\Psi_i & \partial_{12}\Psi_i \\ \partial_{21}\Psi_i & \partial_{22}\Psi_i \end{pmatrix} \tag{4.37}$$

where ∂_{ij} symbols stand for partial differentiation and Ψ_i is the two dimensional potential at the i-th plane, obtained from the Poisson equation.

Once we have integrated from the observer to the source plane we can now put an object in the last plane and ask how it will be modified by a weak lensing. It is usually done by finding the four closest rays to the center of the object and interpolating all the important quantities from the amplification matrix to apply Eq. 4.27.

4.3.2 Ray tracing algorithm

The ray tracing algorithm can be seen as the implementation of the multiple plane theory we have just described. The implementation can be divided into three steps, also illustrated in Fig.4.3 with additional details about how they were implemented in the parallel ray-tracing code written during my thesis.

(i) projection of the matter distribution
Once we have the dark matter particles of the simulation positioned inside a light cone built with `MoMaF`, we define an orthogonal coordinate system (x_1, x_2, y) at the origin of the cone, where the y-axis is directed from the observer along the symmetry axis of the cone. We then project (along the y-axis) all the particles on to N equidistant planes parallel to the one located at the origin of the cone

(ii) calculation of the potential
After we have the overdensity interpolated over the grid at every plane we solve the 2-d Poisson equation for the 2-d density over the

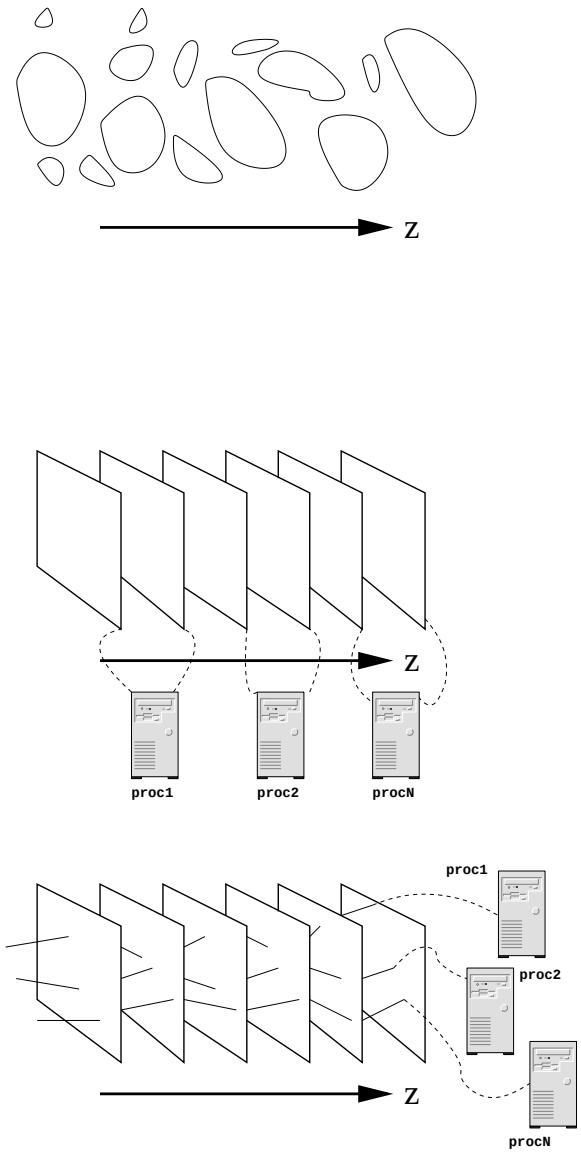


Fig. 4.3. Schematic approach to the ray-tracing algorithm over multiple planes. At a first stage (top) we have a 3-d dark matter distribution over a light cone, which can be created with `MoMaF`. At a second stage (middle), the 3-d distribution is projected over different planes perpendicular to the line of sight and the Poisson equation is solved to obtain the 2-d potential. In the `LeMoMaF` code, different processors take in charge different planes. In a third stage (bottom) rays are traced from the observer towards the source plane, calculating at each plane the deflection angle. In `LeMoMaF` each processor manages a bundle of rays for the ray-tracing part.

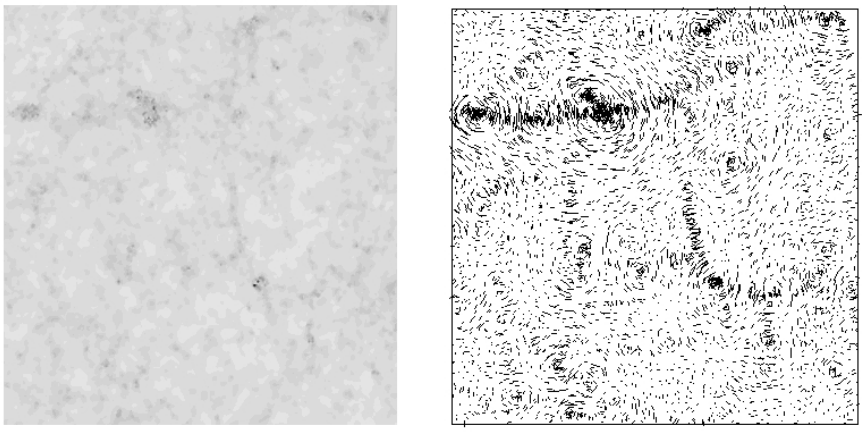


Fig. 4.4. $1^\circ \times 1^\circ$ convergence field (left) and shear field (right) for a source plane at $z \sim 1$. Obtained using our ray-tracing code LEMOMAF.

plane using a Fourier transform technique. This allows us to construct the matrix made up the partial derivatives of the gravitational potential, which give us deflection angles at every point in the grid.

(iii) ray propagation
We initialize the ray positions over a uniform grid on the closest plane the observer. For each ray we interpolate values of the angle of deflection and the derivatives of the gravitational potential. Applying the equation of propagation seen in the last section we figure out the position of the ray at the next plane, to again interpolate the same values to calculate the amplification matrix and get the position to the next plane. We continue in this way until we reach the last plane.

The results of the ray tracing technique is limited basically by three factors: the resolution of the N-body simulation, the size of the grids in the planes where the density is interpolated and the size of the simulation box that limits the size of the patches of the sky that can be simulated.

4.3.3 The galaxies

The ray tracing simulation technique is only concerned with extracting the lensing signal from the dark matter distribution. But in reality the lensing signal is only measured from the galaxies. So, strictly speaking one should

have to simulate not only the weak lensing signal, but also its effect on the ellipticity of simulated galaxies. An approach that is more necessary now that we have recognized their importance in the interpretation of high precision weak lensing measurements.

A first approximation to that kind of simulation, is to ask what happens if we only measure the convergence at the highest density points where one expects galaxies to be found. Could this correlation between the sources of light and the sources of magnification, could have any impact on the final statistics?

Moreover, the intrinsic possible correlations between the ellipticities and the dark matter distribution may have an impact on the shear reconstruction, as we first pointed out in the systematics section.

All this calls for an extension of the ray tracing simulation that includes the information of galaxies. This have been the motivation for the construction of a simulation tool better described at the appendix in the published paper (See the Appendix). In the next section we summarize the more important results we have obtained so far.

4.4 Our Results

The Problem

Weak lensing is rather challenging to detect. Requiring measurement of weak distortions on sets of observed images, which have been convolved with the distortion of the atmosphere telescope and camera. Even under very good conditions of observation and very good control of systematic effects of the measurement equipment the intrinsic galaxy shapes are of the same order as the shear effect we want to measure.

The need to push forward the weak lensing surveys have put the strain on the weakest link in the chain: the uncertainties in galaxy formation. The galaxies are the objects which are actually measured, and the information that can be extracted depends on the null hypothesis we make about the observable properties of galaxies, and controlling the validity of these hypothesis (i.e. degree of intrinsic alignments), and understanding the possible existence of other correlations of astrophysical origin is central to the success of future surveys.

Our Solution and Work in Progress

Our objective is to include the information of galaxy formation inside the simulation of weak lensing, and more specifically, create realistic lensed images of galaxies in an explicit cosmological context.

To do so we have used on one side the galaxy formation code we have described in the previous chapters,`GalICS`, and the tool for construction of mock catalogues (of galaxies and dark matter particles) `MoMaF`. On the other side we have constructed a new code that implements the ray tracing algorithm to extract the lensing information from cones of dark matter particles.

At this point, we can produce a density distribution, its underlying galaxy populations fully described by their physical properties, and the lensing quantities from the matter distribution. The first question we address with this tool is the following: what is the difference between sampling the lensing quantities with a realistic galaxy distribution compared to a random galaxy distribution?

Technically the consequences of this correlation receive the name of Source Lens Clustering effect. It is expected from theoretical grounds that this effect will modify the probability density function of the convergence towards a more Gaussian shape, leaving the two point statistics unchanged, but reducing the skewness.

This was first addressed in theoretical studies [Ber98], and later on numerical implementations that followed an approach very similar to ours [HCT$^+$02], with one big difference: the underlying galaxy distribution was not completely self consistent with the history of the formation of the dark matter structure. Thus, it relied heavily on functional fittings for the redshift distribution of the population of galaxies.

Summarizing, in our implementation we verified that the trend expected in the three point correlation functions is found. This means that the source lens clustering bias does produce a more Gaussian probability density function for the convergence. What was completely unexpected is that we found a minor effect on the variance, which was not suspected at all from theoretical considerations and was not reported before in first numerical attempt.

The non predictions from theoretical grounds can be understood from the highly simplified assumptions in those calculations. The non report of the effect on the variance followed from its almost irrelevant amplitude for the weak lensing surveys at that time, but that will definitely be an effect to consider in the upcoming high precision surveys.

In a second step we explored the influence of weak lensing on galaxy counts. We were guided by the analytical expression that estimates this effect. Physically speaking galaxy counts are modified in two ways by weak lensing first this effect modify the magnitudes, and second the change of positions translates into changes of the number density of objects in the sky. The total effect on galaxy counts $N(m)$ (the number of galaxies with

magnitude between m and $m + dm$) can be then expressed analytically

$$N(m)_{lensed} = \mu^{2.5s-1} N(m) \tag{4.38}$$

where s is defined as $d \log N(m)/dm$ and μ is the magnification defined in the usual fashion from the amplification matrix $\mu = 1/\det \mathsf{A}$. When we confront our simulations to this analytical expression, we could only see the effect of change in the magnitudes. The change of the number density of objects over the sky, produce by this change of positions, could not be observed in our simulations. Even when we did include the change of positions of the galaxies as we had the information to follow that properly. Our numerical results were instead reproduced by the following equation

$$N(m)_{lensed} = \mu^{2.5s} N(m) \tag{4.39}$$

The explanation for missing this effect in our simulations is related to the poor angular resolution we have. Indeed, the deflection angles are on the order of a few arcseconds, while in our simulations we only had an angular resolution of the order of one arcminute, meaning that adding up all the deflections wasn't done with enough resolution to follow deflection in a coherent way to see the effect in the galaxy counts. Using a finer grid is not a solution, because at that moment we will start seeing the particles of the dark matter simulation, and we will be dominated by the noise. The only way out is to increase both the resolution of the grid and the dark matter simulation.

Another part of the solution involves getting rid of the grid to make the computations of the potential. In the literature has been proposed to use particle-particle interactions for close interactions, and a grid for long-range interactions.

Summarizing, we will need higher resolution simulations and a new method complementary to interpolation over a grid. We are in fact starting to use catalogues of dark matter halos to describe the dark matter distribution. This is motivated by the fact that very large simulations are very likely to make the halo identification the fly and output just the halo catalogues, and not the list of particles. Besides dealing with analytical expressions of dark matter halos is much simpler from the computational point of view.

Checking that the halo model still gives meaningful results with a reasonable precision is still a work in progress.

Appendix 1
Cosmology

As the early universe was homogeneous and isotropic, our first task is to understand the behavior of such space. The only theoretical choice we have is the theory of General Relativity, which is a theory describing the interaction of a space time with its matter content. In this theory the basic mathematical ingredients are the metric $g_{\alpha\beta}$ (describing space-time) and the energy momentum tensor $T_{\alpha\beta}$ (describing the matter content).

The next level in complexity deals with the discussion of inhomogeneities. In that case we have two approaches. The conventional way takes the homogeneous and isotropic metric as a starting point and then adds the perturbations later. The second approach starts from the exact equations, and treats them up to linear order and finds then an appropriate expression for the metric.

Behind these two different approaches there is an important point. In General Relativity the average evolution of an inhomogeneous and anisotropic universe does not have to be the same as the evolution of a perturbed homogeneous and isotropic universe. In other words, solving the equations and averaging/linearizing the solutions is not the same as averaging/linearizing the equations to find the solutions.

In all the different cases we will deal with the same kind of matter: ideal fluids. For such a fluid, the energy momentum tensor can be written as

$$T_{\alpha\beta} = (p + \rho)u_\alpha u_\beta + p g_{\alpha\beta}, \tag{1.1}$$

where ρ is the density, p the pressure and u_α is the 4-velocity of the fluid element. In the coordinates we study in the next section we will have $u^\alpha = \delta^{\alpha 0}$ (Kronecker delta), making the energy-momentum tensor diagonal, with $T^0_0 = -\rho$ and $T^i_i = p$.

We will review the two most important approaches for describing the

dynamics of the Universe without deriving explicitly each equation, but indicating if a crucial approximation has been done.

Homogeneous Universe

The corresponding metric for that kind of universe is known as the Friedmann-Robertson-Walker (FRW) metric:

$$ds^2 = -dt^2 + \frac{a(t)^2}{\left(1 + \frac{K}{4}r^2\right)^2}, \tag{1.2}$$

where K is a constant describing the overall geometry of the Universe, $r^2 = \delta_{ij}dx_i dx_j$ and $a(t)$ is the expansion factor.

It is easier to picture the dynamics of the FRW model when the curvature term is set to zero, which seems to correspond to the case of our actual Universe. Using the Einstein's equations and the energy tensor (Eq. 1.1) and the metric we defined in Eq. 1.2 we can obtain the following equations:

$$\frac{\dot{a}^2}{a^2} = \frac{8\pi G\rho}{3} + \frac{\Lambda c^2}{3} \tag{1.3}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda c^2}{3}, \tag{1.4}$$

where the factor $\dot{a}/a$ is known as the Hubble parameter $H(t)$ and Λ is a constant.

Furthermore, we can consider that the Universe is composed of different kinds of matter, with different equations of state.

$$p = \frac{1}{3}\rho \quad \text{for relativistic matter,} \tag{1.5}$$

$$p = 0 \quad \text{for non relativistic matter,} \tag{1.6}$$

$$p = -\rho \quad \text{for the cosmological constant,} \tag{1.7}$$

and now we can write:

$$\frac{1}{H_0}\frac{\dot{a}}{a} = \frac{H(t)}{H_0} = \left[\frac{\Omega_m}{a^3} + \frac{\Omega_r}{a^4} + \Omega_\Lambda\right], \tag{1.8}$$

where H_0 is the value of the Hubble parameter at present, and we have

introduced the density parameters for the non relativistic matter, the relativistic matter (radiation) and the cosmological constant:

$$\Omega_m = \frac{8\pi G}{3H_0^2}\rho_m \quad , \Omega_r = \frac{8\pi G}{3H_0^2}\rho_r \quad , \Omega_\Lambda = \frac{\Lambda c^2}{3H_0^2} \quad . \tag{1.9}$$

These quantities are defined at present time, and its time dependence is given by the denominators in the equation 1.8.

The expansion factor $a(t)$ is useful as a time variable to parameterize the expansion of the universe. It is customary to use instead the redshift, z, defined as

$$1 + z = a(t_0)/a(t) \tag{1.10}$$

where t_0 is the present time. The redshift can be related to observable properties in the universe, in fact the origin of its name comes from the relation of a photon's wavelength emitted at a time t and measured at a later time t_0

$$\frac{\lambda(t_0)}{\lambda(t)} = 1 + z. \tag{1.11}$$

Inhomogeneous Universe

In this case we can follow two approaches: perturb the metric and solve the Einstein equations, or keep the metric general and see how far can we go.

First Way

The most general perturbation of the metric can be written as

$$ds^2 = -(1 - 2\phi)dt^2 + 2(B_{|i} - S_i)dx^i dt \tag{1.12}$$

$$+ \frac{a(t)^2}{1 + \frac{K}{4}r^2}((1 - 2\psi)\delta_{ij} + 2E_{|ij} + F_{i|j} + F_{j|i} + h_{ij})dx^i dx^j, \tag{1.13}$$

where ϕ, ψ, B and E are scalars, S_i, F_i are vectors, h_{ij} is a tensor and $|$ is the covariant derivative with respect to the spatial background metric, which in the case of $K = 0$ reduces to ordinary derivatives. A detailed analysis of this equation would show that, physically, tensors correspond to gravity waves, vectors describe rotation and scalars are related to the density perturbation.

Staying at the linear level we can neglect the tensor component of the perturbation, and because the equations for the perturbations decouple we can set the vectors to zero and study the scalar perturbation only:

$$ds^2 = -(1-2\phi)dt^2 + 2B_{|i}dx^i dt + \frac{a(t)^2}{1+\frac{K}{4}r^2}((1-2\psi)\delta_{ij} + 2E_{|ij})dx^i dx^j, \quad (1.14)$$

which still leaves four degrees of freedom (ϕ, ψ, B and E), not all of them corresponding to physical degrees of freedom. To isolate them one can perform a (linear) transformation of coordinates and ask that the form of the metric remains intact. Fixing the values of some of these scalars would be then equivalent to choose a coordinate system, or gauge. The fastest route to avoid these ambiguities is setting $E = B = 0$, which in the technical jargon is called the longitudinal gauge (or Newtonian gauge):

$$ds^2 = -(1-2\phi)dt^2 + a(t)^2(1-2\psi)\delta_{ij}dx^i dx^j. \quad (1.15)$$

If one goes through the Einstein equations with the previous form of the metric, and the diagonal energy tensor of the ideal fluid, it follows that from consistency conditions for an ideal fluid $\phi = \psi$, leaving us with just one physical degree of freedom:

$$ds^2 = -(1-2\phi)dt^2 + a(t)^2(1-2\phi)\delta_{ij}dx^i dx^j. \quad (1.16)$$

This degree of freedom, ϕ, is related to the density perturbation through the Poisson equation. This simple form of the metric (after a choice of gauge and neglection of the tensor modes) is the one used almost everywhere and everytime there is a study requiring a description of an mildly perturbed Universe. For instance, in the description of light propagation in a inhomogeneous universe, this is the metric at the base of all the developments in weak gravitational lensing on cosmological scales.

Second Way

The second way takes a general form of the metric, which under the supposition that the velocity field of the ideal fluid we are studying is non rotational ($\nabla_\alpha u_\beta - \nabla_\beta u_\alpha = 0$) can be written as

$$ds^2 = -dt^2 + h_{ij}dx^i dx^j, \quad (1.17)$$

a choice called synchronous comoving coordinates. Making projections of the scalar parts of the Einstein equations we obtain

$$\dot{\theta} = \frac{1}{3}\theta^2 = -4\pi G\rho - 2\sigma^2 \tag{1.18}$$

$$\frac{1}{3}\theta^2 = 8\pi G\rho - \frac{1}{2}{}^{(3)}R + \sigma^2 \tag{1.19}$$

$$\dot{\rho} + \theta\rho = 0, \tag{1.20}$$

where $\theta = \partial_t\sqrt{h}/\sqrt{(h)}$, with $h \equiv \det(h_{ij})$, is a measure of the expansion rate of the metric. σ is the average shear of the metric and ${}^{(3)}R$ is the spatial curvature. In order to make useful comparisons with the FRW equations we define a scale factor as an average over a volume of the hypersurface of constant t

$$a(t) \equiv \left(\frac{\int d^3x\sqrt{(h(t,\mathbf{x}))}}{\int d^3x\sqrt{h(t_0,\mathbf{x})}}\right)^{1/3}. \tag{1.21}$$

Taking the average of the scalar parts of the Einstein equations (Eqs. 1.18 - 1.20), with the average of a scalar calculated as

$$\langle f\rangle = \frac{\int d^3x\sqrt{h(t,\mathbf{x})}f(t,\mathbf{x})}{\int d^3\sqrt{h(t,\mathbf{x})}}, \tag{1.22}$$

we obtain the following equations

$$3\frac{\ddot{a}}{a} = -4\pi G\langle\rho\rangle + \mathcal{Q} \tag{1.23}$$

$$3\frac{\dot{a}^2}{a^2} = 8\pi G\langle\rho\rangle - \frac{1}{2}\langle{}^{(3)}R\rangle - \frac{1}{2}\mathcal{Q} \tag{1.24}$$

$$\partial_t\langle\rho\rangle + 3\frac{\dot{a}}{a}\langle\rho\rangle = 0, \tag{1.25}$$

where $\mathcal{Q}$ is called the reaction variable and is a new term comparing with the FRW equations, containing the effect of inhomogeneity and anisotropy

$$\mathcal{Q} \equiv \frac{2}{3}(\langle\theta^2\rangle - \langle\theta\rangle^2) - 2\langle\sigma^2\rangle. \tag{1.26}$$

These equations are called the Buchert equations, and are at the base of the study of dynamics of inhomogeneous universes. In contrast to the case of the FRW equations, this approach is still young and not even a simple description of the propagation of light in such a universe has been worked out in the astrophysical literature.

Appendix 2
Codes from the projects

The agenda after perfoming a N-body simulation is the following. First we perfom the identification of the halos using a FOF algorithm, this is done with the code `HaloMaker`. This code produces for each snapshot the list of particles belonging to dark matter structures.

With these outputs we construct the merger tree. The code performing this task is called `TreeMaker`. As an upgrade from the previous version of the code, it is now possible to leave the identified trees in different files, i.e. we can choose a subset of halos and write its full merger histories in separate files. Before, all the trees were written in the same file, which forced the galaxy formation code to run on one procesor only. Now, trivially, we can use different input files on independent processors. Typically we split the trees into ~ 1000 files.

After the construction of mergre trees, we build the galaxies with `GalaxyMaker`. The fact that the tree files are split, introduced additional complexity to the launch process which now is done through `python` scripts that manage the deployment of the directories and input data, as well the running on multiprocessor machines without a proper queuing system.

Once `GalaxyMaker` has run over the galaxies we still have to give positions and velocities to the galaxies. We do that through an additional code called `GSM_STD`. After this, we can finally construct mock catalogues with `MoMaF`.

The Weak Lensing code `LeMoMaF` stands for Lensed Mock Map Facility, and was created to take as inputs the outputs from `MoMaF`, but in no way is dependent on the detailed algorithm used to construct the cones, nor is a module inside the `MoMaF`.

`LeMoMaF` Is a code written in `ANSI C`, parallelized using `MPI`. It produces maps of convergence and shear. Also weak lensed catalogues to be interpreted by *SkyMaker* to produce images as observed by a telescope with all the relevant systematic effects.

Appendix 3
Submitted Paper: Merger Trees

Abstract

We introduce the contour process to describe the geometrical properties of merger trees. The contour process produces a one-dimensional object, the contour walk, which is a translation of the merger tree. We portray the contour walk through its length and action. The length is proportional to to the number of progenitors in the tree, and the action can be interpreted as a proxy of the mean length of a branch in a merger tree.

We obtain the contour walk for merger trees extracted from the public database of the Millennium Run and also for merger trees constructed with a public Monte-Carlo code which implements a Markovian algorithm. The trees correspond to halos of final masses between $10^{11}h^{-1}$ $M_\odot$ and $10^{14}h^{-1}$ $M_\odot$. We study how the length and action of the walks evolve with the mass of the final halo. In all the cases, except for the action measured from Markovian trees, we find a transitional scale around $3 \times 10^{12}h^{-1}$ $M_\odot$. As a general trend the length and action measured from the Markovian trees show a large scatter in comparison with the case of the Millennium Run trees.

J.E. Forero-Romero, The Coarse Geometry of Merger Trees in ΛCDM, MNRAS submitted.

Mon. Not. R. Astron. Soc. **000**, 1–?? (0000) Printed 19 October 2007 (MN LATEX style file v2.2)

The Coarse Geometry of Merger Trees in ΛCDM

Jaime E. Forero-Romero [1,2⋆]
[1] *Astrophysikalisches Institut Potsdam, an der Sternwarte 16, D-14482 Potsdam, Germany*
[2] *Université Claude Bernard Lyon 1, CNRS UMR 5574, ENS Lyon, Centre de Recherche Astrophysique de Lyon, Observatoire de Lyon, 9 Avenue Charles André, 69561 St-Genis-Laval Cedex, France*

19 October 2007

ABSTRACT
We introduce the contour process to describe the geometrical properties of merger trees. The contour process produces a one-dimensional object, the contour walk, which is a translation of the merger tree. We portray the contour walk through its length and action. The length is proportional to to the number of progenitors in the tree, and the action can be interpreted as a proxy of the mean length of a branch in a merger tree.

We obtain the contour walk for merger trees extracted from the public database of the Millennium Run and also for merger trees constructed with a public Monte-Carlo code which implements a Markovian algorithm. The trees correspond to halos of final masses between $10^{11}h^{-1}$ $M_\odot$ and $10^{14}h^{-1}$ $M_\odot$. We study how the length and action of the walks evolve with the mass of the final halo. In all the cases, except for the action measured from Markovian trees, we find a transitional scale around $3 \times 10^{12}h^{-1}$ $M_\odot$. As a general trend the length and action measured from the Markovian trees show a large scatter in comparison with the case of the Millennium Run trees.

Key words: cosmology:theory, cosmology:dark matter

1 INTRODUCTION

The current paradigm of large scale structure formation in the Universe is hierarchical, meaning that structures are formed by merger aggregation.

The dominant driver for this merger dynamic is thought to be the non-baryonic and non-collisional cold dark matter. Therefore, research in large scale structure formation is principally done through the study of dark matter structures, which on large scales are thought to serve as scaffolding for baryonic structures (Springel et al. 2006).

The dark matter merger process has been usually represented by a tree, keeping the analogy with the genealogical tree of an individual. Great effort has been put into the methods of construction of merger trees, as they are a way to understand dark matter aggregation and are a necessary input for codes of semi-analytic galaxy formation. The methods range from the analytical approach of Monte-Carlo methods (Zentner (2007) for a review) passing through hybrid approaches that mix numerical realizations of a density field with analytical approximations for its evolution (Monaco et al. 2007) to the fully numerical approach that identifies the dark matter halos from different snapshots in a N-body simulation to construct the merger history(Hatton et al. 2003).

Usually, the validity of the trees constructed in the analytical and hybrid way is stated from comparisons with trees constructed from numerical simulations (Sheth & Lemson 1999; Parkinson et al. 2007; Neistein & Dekel 2007). Unfortunately, the quantities used to compare trees from two approaches usually sacrifice the complexity inherent to the tree structure in the sake of simple tests. The most common simplification is to select one branch of the tree (the most massive) to make the analysis. Another approach, measures the abundance of structures of a given mass among the halos in all the merger tree. In all these cases the geometrical information is suppressed, mostly because of the lack of simple structures to describe that kind of information.

In this paper we present a new way, in the astrophysical context, to translate the geometrical information from a merger tree into a 1-dimensional structure. The translation is based on the encoding of the tree information into its *contour walk*.

We apply this description to the merger trees from a large dark matter numerical simulation, the Millennium Run (Springel et al. 2005). We use its public database to select halos in different bin masses at redshift $z = 0$ to extract its merger histories and build the contour walk. We analyze each tree in terms of two simple statistics extracted from these walks.

We have also performed this kind of analysis on merger trees obtained with the algorithm described by Neistein &

⋆ E-mail:jforero@aip.de

Dekel (2007), using the source code they kindly made public. With this code we have constructed trees with two resolutions for the minimum mass halo, one mimicking the Millennium Run, and the other with nearly 8 times lower resolution.

This paper is divided as follows. In Section 2 we explain the construction of a contour walk from a merger tree. In Section 3 we present the simulation that produced the public data we used in this paper and the Monte Carlo code for the construction of Markovian merger trees. In Section 4 we perform an immediate implementation of these concepts to the available merger trees. We calculate global statistics from the walks, and discuss its possible physical meaning. In the last section we discuss our results and suggest how the tool we have proposed can be used to tackle more complex questions about merger trees.

2 TREES AND WALKS

The merger trees in the current dark matter paradigm are simple, as only merging is allowed. From the mathematical point of view this tree structure corresponds to a Galton-Watson tree.

Galton-Watson trees are genealogical trees of a population that evolves according to very simple rules. At generation 0, the population starts with one individual, called the ancestor. Then each individual in the population has, independently of the others, a random number of children according to an offspring probability distribution. In the structure formation context we identify the ancestor with the halo at redshift zero, and the offsprings with the parents of the halo.

These trees can be coded by a discrete path around the contour of the tree (Le Gall 2005). This *contour process* is easy to visualize. We imagine the displacement of a particle that starts at time zero from the root of the tree (a halo at redshift zero) and then visit the first not yet visited parent halo, if any, and if none, the successor of the halo. We do so until all the members in the tree have been explored and the particle has come back to the origin.

This process is illustrated in the Fig.1. Where the different values of t correspond to the discrete snapshots for which we have the halo information, and τ_i is the imaginary time giving the pace of the particle around the tree. We point out that one has to define some order between the halos at a given point in the tree in order to define a unique way to walk the tree, a criteria to decide to which parent should jump the particle. This ordering criterium in our case is naturally imposed by the masses of the halos, as we always visit first the most massive progenitor.

We can express this as visiting all the halos in the tree in a depth-first fashion, visiting first the most massive branch at every time.

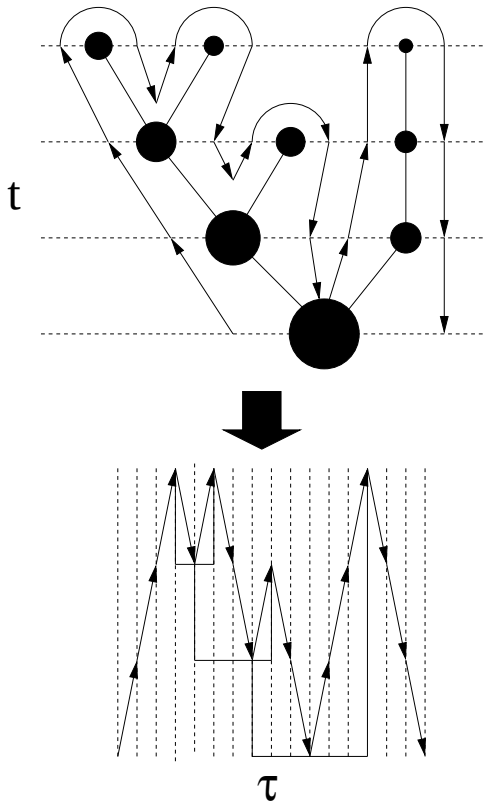


Figure 1. Illustration of the contour process for a simple merger tree. The upper panel represents a merger tree formed by an evolving halo population at different timesteps t (horizontal lines), time flowing from top to bottom. We start from the last halo at the bottom, and visit the halos in the tree at unit timesteps τ. The most massive progenitor is always visited first. We go around until we go back to the initial point. The bottom part of the figure shows the contour walk, its values correspond to the time t where the particle was located at time τ. The tree structure can be recovered completely from the contour walk as indicated by the solid lines in the lower panel.

3 N-BODY SIMULATION AND MONTE-CARLO CODE

3.1 The Millennium Run

The Millennium Run (Springel et al. 2005) (MR hereafter) was carried out by the Virgo Consortium in 2004 at the Max Planck's Society's supercomputer center in Garching. It is a dark matter only simulation containing 2160^3 particles, it was evolved from $z = 127$ to the present-day inside a cubic region of 500 Mpc h^{-1} on a side, the individual particle mass is $8.6 \times 10^8 h^{-1}$ $\mathrm{M}_\odot$. It adopted $\Omega_{dm} = 0.205$, $\Omega_b = 0.045$ for the current densities of dark matter and baryons in a Lambda Cold Dark Matter cosmological model, furthermore it used $h = 0.73$ for the present dimensionless value of the Hubble constant, $\sigma_8 = 0.9$ for the *rms* linear mass fluctuation in a sphere of radius $8h^{-1}$Mpc extrapolated to $z = 0$, and $n = 1$ for the slope of the primordial fluctuation spectrum.

The simulation was run using the TREE-PM N-body code GADGET2 (Springel 2005), 64 outputs were stored at times spaced approximately equally in the logarithm of the expansion factor at early times and at approximately 200 Myr intervals after $z = 1$. At run time all collapsed halos with at least 20 particles ($1.7 \times 10^{10} h^{-1}$ $\mathrm{M}_\odot$) were identified using a friends-of-friends (FOF) (Davis et al. 1985) group-finder with linking length $b = 0.2$. Post processing with the substructure algorithm SUBFIND (Springel et al.

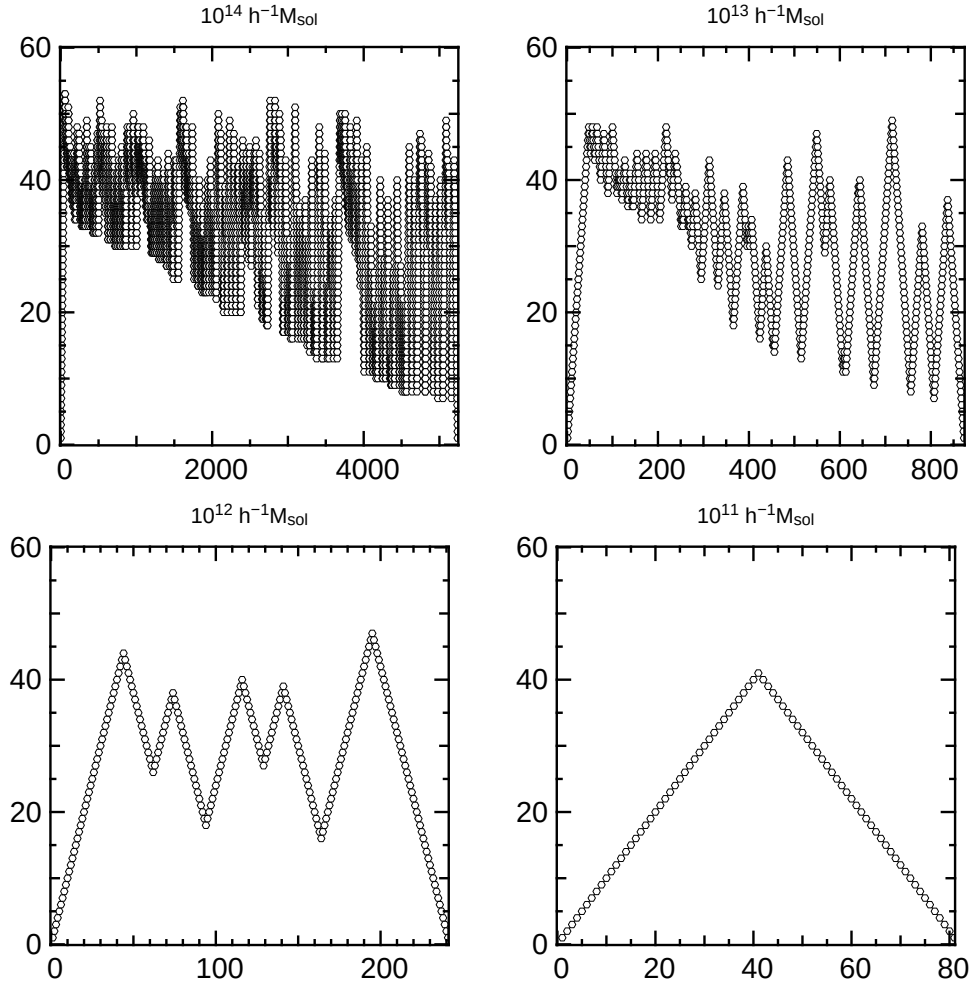


Figure 2. Examples of contour walks constructed from merger trees extracted from the Millennium Run. Each contour corresponds to a halo of different mass. $\sim 10^{14}h^{-1}$ $M_\odot$ (upper left), $\sim 10^{13}h^{-1}$ $M_\odot$ (upper right), $\sim 10^{12}h^{-1}$ $M_\odot$ (lower left), $\sim 10^{11}h^{-1}$ $M_\odot$ (lower right).

2001) allowed a detection and measurement of the resolved sub-halos. This in turn allowed trees to be built with detailed assembly histories for every object and its substructure.

The postprocessed data have been publicly disseminated through an interactive database: `http://www.g-vo.org/Millennium/MyDB`. The particular structure of the database design which allows efficient querying for merger trees was implemented by Lemson & Springel (2006). Conveniently enough, the structure for merger trees in the database is based on a depth-first ordering, making that the output of a tree query is an incomplete version of the contour walk In terms of the lower panel in Fig.1, only the upward arrows offspring-progenitor exist.

3.2 Markovian Trees

The authors Neistein & Dekel (2007) recently proposed and algorithm for the construction of merger trees based on a Markovian approach. This approach to merger trees means that any halo of a given mass M at time t has a progenitor probability distribution depending only on M and t. This scheme is explicitly independent of the large scale environment that could be defined for a halo.

Their merger trees are parameterized by a time variable $\omega \equiv \delta_c/D(z)$ and a mass variable $S(M) = \sigma^2(M)$, where $\delta_c \sim 1.69$ and $D(z)$ is the cosmological linear growth rate and $\sigma^2(M)$ is the variance of the initial density fluctuation field, linearly extrapolated to $z = 0$ and smoothed using a window function that corresponds to a mass M. These two variables are also the natural variables in the Extended Press-Schechter formalism (Zentner 2007).

Their approach is the following. First, from the MR data they find the conditional probabilities for the masses of the main progenitors at a past time t' as a function of halo mass M at a present time t. With this conditional probabilities they build the main progenitor history which by construction reproduce to a good extent the MR data. Then, they extended this approach to the construction of a full merger tree. The extension included some additional heuristic rules based again on the premise of a fair match to the total progenitor number density dN/dM from the Millennium Run data. Nevertheless, the lack of a true physical motivation

cannot ensure the reconstruction of the correct full joint distribution of progenitors.

Even if N-body merger trees are not Markovian, the proposed algorithm manages to reproduce some tree properties, specially those related with the main progenitor but also the total mass distribution in all the progenitors. The detected inaccuracies come from the fact that the estimation of the average mass for the second progenitor is not reliable.

In spite of that, we have decided to use these author's public available code given its explicit effort to reproduce the MR data.

4 EXPERIMENT SETUP AND NOTATION

4.1 Tree Selection

We make use of two kind of trees. Trees extracted from the MR public data base and trees constructed with the public code implementing the Markovian approach. In both cases we concentrate on halos with masses greater than $10^{11}h^{-1}\mathrm{M}_{\odot}$.

For the MR trees we selected all the halos in the simulation box with a given mass $M_H = \log_{10}(M_{200}/10^{10}h^{-1}\ \mathrm{M}_{\odot})$ at redshift $z = 0$, for 28 different values of M_H in bins of width $\Delta M_H = 0.002$ dex, where M_{200} is the halo mass measured within the radius where the halo has an overdensity 200 times the critical density of the simulation. The bins are spaced by 0.1 dex, the least massive bin corresponds to $M_H = 1.3$ and has 6028 dark matter halos and the most massive bin corresponds to $M_H = 4.0$ and has 17 dark matter halos. We also performed measurements for the 15 most massive bins with $\Delta M_H = 0.01$dex, which provided us with nearly five times more halos per bin. The results we obtained in that case are basically the same than in the case $\Delta M_H = 0.002$. In this paper we only report and discuss the results for the selection with the smaller $\Delta M_H = 0.002$ sample.

For the Markovian trees we made two different runs changing the minimal halo mass. The first one mimicked the MR and its minimal mass was $1.7\times10^{10}h^{-1}\ \mathrm{M}_{\odot}$. The second run used a higher minimal mass of $1.0\times10^{11}h^{-1}\ \mathrm{M}_{\odot}$. We will refer to these runs as the high resolution and low resolution runs, respectively. In both cases we also constructed merger trees for halo mass bins spaced by 0.1 dex. The mass bins for the high resolution run are the same as in the MR case, for the low resolution run we can only describe trees from the bin $M_H = 1.6$ up to $M_H = 4.0$. For each bin we construct 1000 trees, each one with 100 steps in ω with $\Delta\omega = 0.1$. Which is equivalent to have trees described from redshifts $z \sim 8$.

4.2 Notation

The snapshots in the dark matter simulation will be labeled by t_i, where i ranges from 0 to 63. The snapshot t_0 corresponds to the $z = 0$. We select a halo in the snapshot t_0 to extract its merger tree. The contour walk, as described in Section 2, can be visualized as a dynamical process of a particle going around the tree, at unit time-steps stopping at each node in the tree and recording the snapshot t_i to which it belongs. The discrete variable counting the imaginary time of the particle walking the tree is noted τ_i.

We write the walk as a sequence of discrete values $\{x_0, \ldots x_N\}$, corresponding to different values of the discrete intermediary times $\{\tau_0, \ldots, \tau_N\}$, N is the total length of the walk, and every x_i can take values from the possible snapshots $t_0 \ldots t_{63}$. In the case of the Markovian trees the values the walk can take vary from ω_0 to ω_{99}, consistent with the fact that we have described these trees at 100 points equally spaced by $\Delta\omega = 0.1$.

For each merger tree we compute two statistics: its length and action. The length corresponds to the number of points in the walk. The action, which will be defined later, is an statistic based on the first derivatives of the walk. We offer in Fig.2, a feeling on the merger walks for halos in four different mass bins.

5 SIMPLE STATISTICS FROM THE WALKS

5.1 Walk Length

The Figure 3 shows the walk length for halo merger trees as a function of halo mass. At low masses where the growth should be dominated by mass accretion and not by mergers, $N \propto M^{0.5}$. At high masses where the growth starts to be controlled by mergers $N \propto M^{0.8}$. The transitional scale corresponds to halos for a halo mass $1.6 \times 10^{12}h^{-1}\mathrm{M}_{\odot}$.

This transitional scale should depend on the resolution of the dark matter simulation. When smaller halos are resolved, the merger trees will be populated with more branches of this lower mass halos, and one could start to see that all the growth is done trough mergers.

Nevertheless, if we intend to study galaxy formation using merger trees, this suggested transitional scale may have some significance. In the favored paradigm of galaxy formation, not every halo can harbor galaxies at a cosmic epoch. Only halos that can cool efficiently the gas may hold star forming galaxies. A lower limit is imposed by the UV background from star formation, which sets a low baryonic fraction in halos of masses below $\sim 1 \times 10^{10}h^{-1}\mathrm{M}_{\odot}$, which is roughly the minimal resolution for halo detection in the Millennium Run. If one aims to study galaxy formation, the UV background sets naturally a minimal mass of the progenitors that should be included in the merger history of a dark matter halo. In any case, we decided to explore the influence of the resolution using the results from the Monte-Carlo code.

In the case of merger trees constructed with a resolution mimicking the MR we find again a transitional scale (Fig.4) around $4.0 \times 10^{12}h^{-1}\mathrm{M}_{\odot}$. The biggest difference in comparison with the N-body results are the exponents describing the length of the walk as a function of halo mass. In the Monte Carlo case both exponents are closer to $\lambda \sim 0.9$. This results would favor a view from which there is still room left for a better description of smooth accretion into the Markov description.

We extend the statistical description of the contour walk by the quantification of the its waviness. We borrow from statistical mechanics the concept of action, which is used in the numerical calculation of path integrals (Krauth 2006).

The action can be defined as the potential energy invested in bending the path, comparing it with a charac-

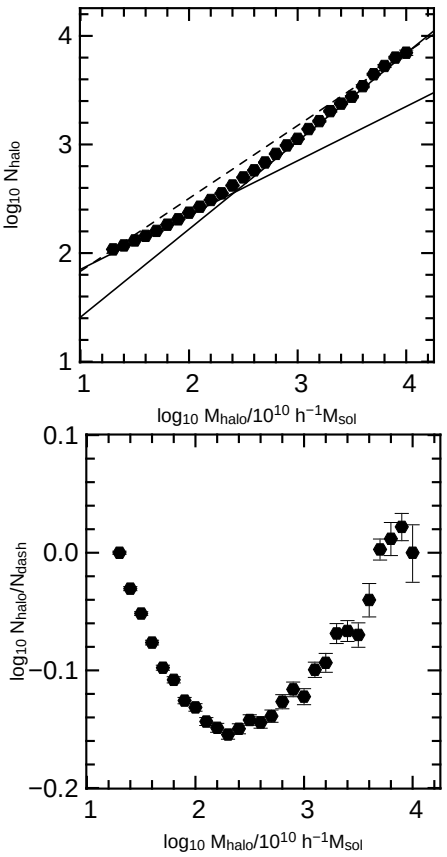


Figure 3. *Upper panel* Walk length N_{halo} for halo merger trees extracted from the Millennium Run as a function of the logarithm of the final halo mass. The solid lines show two different power law trends for the relation $N_{halo} \propto M_{halo}^{\lambda}$, $\lambda = 0.5$ and $\lambda = 0.8$. The intersection of the two lines is located at $1.6 \times 10^{12} h^{-1}\mathrm{M}_{\odot}$. The dashed line simply passes through the two extreme mass points in the plot. *Lower panel* The same results as in the upper panel, but this time the measured values are normalized to the functional dependence of the dashed line in the upper panel. This enhances the features in the curve and allows an easier determination of the transitional mass-scale.

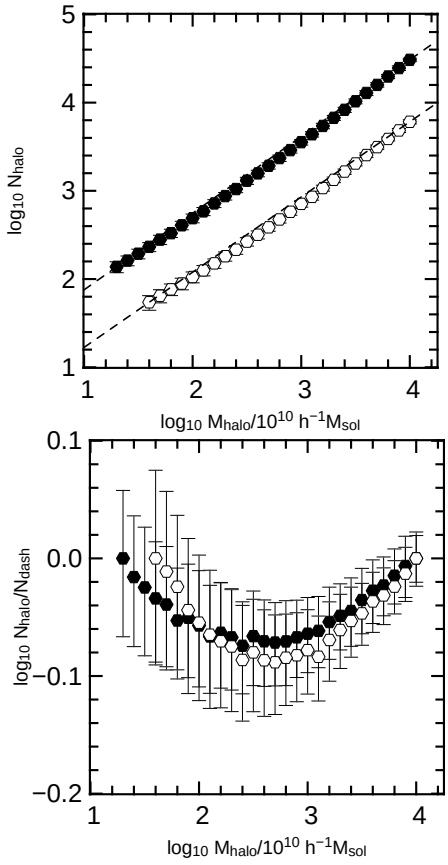


Figure 4. *Upper panel*Same as Fig.3 for merger trees constructed with the Markovian algorithm. Black symbols show the high resolution run, and white symbols the low resolution run. *Upper panel* The lines show again two different power law trends for the relation $N_{halo} \propto M_{halo}^{\lambda}$, $\lambda = 0.80$ and $\lambda = 0.95$. This is more steeper than the $\lambda = \{0.5, 0.8\}$ found in the Millennium merger trees. *Lower panel* The transitional mass-scale is roughly located at $4.0 \times 10^{12} h^{-1}\mathrm{M}_{\odot}$ with a weak dependence on the run's resolution. The dispersion for the measurements from the Markovian trees is noticeably larger than the measurements coming from Millennium Run data.

teristic temperature. If there is a walk defined by points $\{x_1 \dots x_N\}$, the action is usually defined as

$$S = -\sum_{i=1}^{N} \frac{(x_{i+1} - x_i)^2}{2\beta}, \tag{1}$$

where β plays the role of a temperature or imaginary time depending on the context. If we take that definition applied to our case:

$$S = \sum_{i=1}^{N} \frac{(x_{i+1} - x_i)^2}{\tau_{i+1} - \tau_i}, \tag{2}$$

as we have unit steps, $\tau_{i+1} - \tau_i$ is equal to $|x_{i+1} - x_i|$, making the action S equal to the walk length. Therefore, we decide to define a normalized action only from the extreme points in the contour walk. We take this extreme points, y_i, as a sample of the contour walk where the derivative at point x_i, defined as $(x_{i+1} - x_{i-1})/2$, is equal to zero. We note the set of times τ_i corresponding to the vanishing points of the derivative as $\mathcal{E}$ (as in $\mathcal{E}$xtreme).

If we now define the action on the sampled walk $\{y_i |\, \tau_i \in \mathcal{E}\}$

$$S = \frac{1}{N_E} \sum_{i \in \mathcal{E}} \frac{(y_{i+1} - y_i)^2}{\tau_{i+1} - \tau_i} \tag{3}$$

which makes the action S equal to the walk length divided by the number of extreme points. In the context of merger trees, each peak in the walk corresponds to a branch in the tree, making the contribution to the action roughly proportional to the length of that branch. Therefore, the normalized action could be loosely interpreted as a proxy for the mean length of one branch in the tree.

5.2 Walk Action

In Fig.5 we show the normalized action as the function of halo mass for the MR and the Monte-Carlo code. For MR trees we find a distinction between the behavior of the action for low and high mass halos. This time the two mass

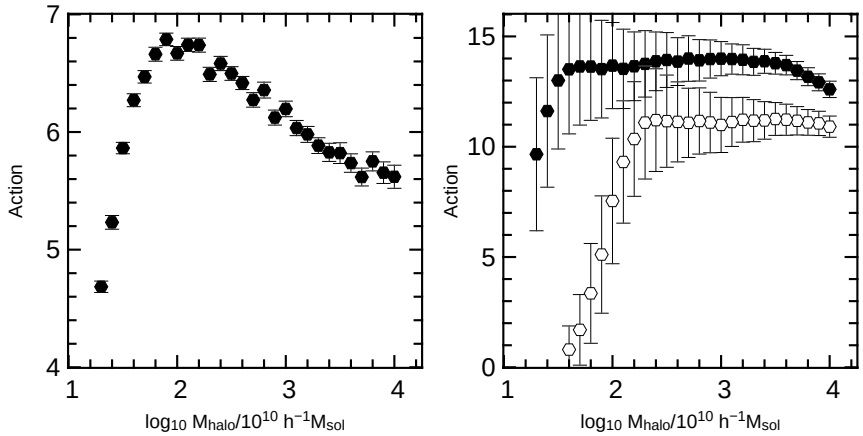


Figure 5. *Left panel* Normalized walk action S, Eq.(3), as a function of halo mass for merger trees extracted from the Millennium Run. We find again a transitional mass-scale around $1.0 \times 10^{12} h^{-1} \mathrm{M}_{\odot}$, which could be interpreted as the mean branch length reaching a maximum for that halo mass. *Right panel* The same as the left panel for merger trees constructed with the Markovian algorithm. Black symbols refer to the high resolution run, white symbols to the low resolution run. The transitional mass-scale is replaced by a plateau. Here the mass resolution plays an important role shaping the general trend for the action S. As in the length case, the dispersion in the measurements from Markovian trees is larger than the measurements coming from Millennium Run data.

extremes share the same action values, meaning that we can identify the same mean branch length for the two extremes masses. For intermediate masses the action increases. In the case of the MR trees, the action achieves a maximum for a mass scale $1.0 \times 10^{12} h^{-1}$ $\mathrm{M}_{\odot}$. For Monte Carlo trees the maximum takes the form of a broader plateau ranging almost two orders of magnitude between $4.0 \times 10^{11} h^{-1}$ $\mathrm{M}_{\odot}$ and $4.0 \times 10^{13} h^{-1}$ $\mathrm{M}_{\odot}$.

Perhaps the most distinguishing factor between the MR and Markovian approach is that the dispersion in the latter is much higher that in the former. A fact that can be interpreted as a higher variability in the geometry of Markovian trees.

6 DISCUSSION

We introduced from the mathematical literature the contour process of a tree, and we applied this concept to the description of merger trees. We used a large dark matter simulation (the Millennium Run) and a Monte-Carlo code (implementing a Markovian approach) to obtain merger trees in these two approximations. Furthermore, the Markovian trees were obtained with two different values for the minimal mass of a parent halo. One resolution mimicked exactly the MR, and the other resolution had a 8 times more massive minimal halo mass. We refer to these Markovian runs as the high resolution and low resolution runs respectively.

We extracted simple statistics from these walk: the length N (proportional to the total number of halos in the tree) and the action S (which can be loosely interpreted as the mean longitude of a branch). We report our results of walk length and action emphasizing its evolution as a function of the physical halo mass, and not its absolute values.

From the length and the action, we found in the Millennium Run a transitional mass scale at $\sim 3.0 \times 10^{12} h^{-1}$ $\mathrm{M}_{\odot}$. In the case of the walk length, N, this transitional mass-scale marks the change between a dependence $N \propto M^{0.5}$ for low halo masses and $N \propto M^{0.8}$ for high mass halos, where M is the mass of the final dark matter halo in the tree. With the action, the scale marks the highest value for the action as function of the halo mass.

For the Markovian trees the dependence of the walk length on the halo mass is almost the same everywhere, although much steeper than the dependence found in the Millennium Run. Nonetheless, we found the same transitional scale from the length statistics for the two resolutions. It does not prove that the transitional scale is independent on the resolution mass (in fact, it should be dependent) but suggests that the scale does not have a strong dependence with the minimal mass resolution used to describe the merger tree. The evolution of the action as a function of halo mass is completely different from the MR case. The Markovian run with mimicking MR resolution does not show a sharp transitional scale, instead it shows a large plateau ranging for two orders of magnitude in mass.

Perhaps the biggest difference between the two ways of constructing the trees is that the Markovian approach always show a bigger dispersion on its contour walk statistics, which seem best defined in the Millennium Run, judging from its low dispersion. This could be interpreted in fact as a higher geometrical variability in the Markovian trees.

We have shown how simple statistics from the contour walk can give a new handle on the description of merger tree, exploring the geometrical information of merger trees. Even if the length of the walk could have been obtained without an intermediating contour process, the approach of extracting information from higher order statistics, as was the case for the normalized action S, can be extended in complexity. For instance, is relevant to point out that the information of the mass in each node is also encoded in the contour walk. The mass information is used as an ordering criteria to walk the tree. One could define sections defined by $\{\tau_i | x(\tau_i = t_1)\}$, i.e. the points where the walk touches the next to last snapshot, and the *length ratio* of the sections delimited by these points would include information of the *mass ratio* of the mergers at time t_1. In general the study of crossing paths, defined as the the section of the contour

walks within some boundary $a < \{x_i\} < b$ might produce useful statistics to further classify the complex merger trees of massive halos.

ACKNOWLEDGMENTS

We thank Jérémy Blaizot (JB) and Gerard Lemson (GL) for early discussions, motivation and criticism around the ideas presented here. We thank JB's suggestion on pointing the discussion towards a comparison with Markovian merger trees. We thank again GL for his amazing day-to-day work on the public Millennium database. We also thank the authors Neistein and Dekel for making public a well documented and easy to use code for merger tree construction. The necessary infrastructure to develop this work was provided in the framework of the HORIZON Project (France).

REFERENCES

Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371

Hatton S., Devriendt J. E. G., Ninin S., Bouchet F. R., Guiderdoni B., Vibert D., 2003, MNRAS, 343, 75

Krauth W., 2006, Statistical Mechanics: Algorithms and Computations. Oxford University Press, September 2006.

Le Gall J F., 2005, Probability Surveys, 2, 245

Lemson G., Springel V., 2006, in Gabriel C., Arviset C., Ponz D., Enrique S., eds, Astronomical Data Analysis Software and Systems XV Vol. 351 of Astronomical Society of the Pacific Conference Series, Cosmological Simulations in a Relational Database: Modelling and Storing Merger Trees. pp 212–+

Monaco P., Fontanot F., Taffoni G., 2007, MNRAS, 375, 1189

Neistein E., Dekel A., 2007, ArXiv e-prints, 708

Parkinson H., Cole S., Helly J., 2007, ArXiv e-prints, 708

Sheth R. K., Lemson G., 1999, MNRAS, 305, 946

Springel V., 2005, MNRAS, 364, 1105

Springel V., Frenk C. S., White S. D. M., 2006, Nature, 440, 1137

Springel V., White S. D. M., Jenkins A., Frenk C. S., Yoshida N., Gao L., Navarro J., Thacker R., Croton D., Helly J., Peacock J. A., Cole S., Thomas P., Couchman H., Evrard A., Colberg J., Pearce F., 2005, Nature, 435, 629

Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726

Zentner A. R., 2007, International Journal of Modern Physics D, 16, 763

Appendix 4
Submitted Paper: Predictability

Abstract

We propose a general framework to scrutinize the performance of semi-analytic codes of galaxy formation. The approach is based on the analysis of the outputs from the model after a series of perturbations in the input parameters controlling the baryonic physics. The perturbations are chosen in a way so that they do not change the results in the luminosity function or mass function of the galaxy population.

We apply this approach to a particular semi-analytic model called GalICS. We chose to perturb the parameters controlling the efficiency of star formation and the efficiency of supernova feedback. We keep track of the baryonic and observable properties of the central galaxies in a sample of dark matter halos with masses ranging from 10^{10} $M_\odot$ to 10^{13} $M_\odot$.

We find very different responses depending on the halo mass. For small dark matter halos its central galaxy responds in a highly predictable way to small perturbation in the star formation and feedback efficiency. For massive dark matter halos, minor perturbations in the input parameters can induce large fluctuations on the properties of its central galaxy, at least ~ 0.1 in (B-V) color or ~ 0.5 mag in U or r filter, in a seemingly random fashion. We quantify this behavior through an objective scalar function we call predictability.

We argue that finding the origin of this behavior needs additional information from other approximations and different semi-analytic codes. Furthermore, the implementation of an scalar objective function, such as the predictability, opens the door to quantitative benchmarking of semi-analytic codes based on its numerical performance.

In a era when sub-samples of galaxy populations issued from semi-analytic models are starting to be used to make predictions about a wide range of

phenomena, it is important to bear in mind that some properties can be subject to a potentially large scatter from mild perturbation in the input parameters of the model.

J.E. Forero-Romero, Predictability in Semi-Analytic Models of Galaxy Formation, MNRAS submitted

Mon. Not. R. Astron. Soc. **000**, 1–?? (0000) Printed 21 October 2007 (MN LATEX style file v2.2)

Predictability in Semi-Analytic Models of Galaxy Formation

Jaime E. Forero-Romero[1,2⋆]
[1] *Astrophysikalisches Institut Potsdam, an der Sternwarte 16, D-14482 Potsdam, Germany*
[2] *Université Claude Bernard Lyon 1, CNRS UMR 5574, ENS Lyon, Centre de Recherche Astrophysique de Lyon, Observatoire de Lyon, 9 Avenue Charles André, 69561 St-Genis-Laval Cedex, France*

21 October 2007

ABSTRACT

We propose a general framework to scrutinize the performance of semi-analytic codes of galaxy formation. The approach is based on the analysis of the outputs from the model after a series of perturbations in the input parameters controlling the baryonic physics. The perturbations are chosen in a way that they do not change the results in the luminosity function or mass function of the galaxy population.

We apply this approach on a particular semi-analytic model called GalICS. We chose to perturb the parameters controlling the efficiency of star formation and the efficiency of supernova feedback. We keep track of the baryonic and observable properties of the central galaxies in a sample of dark matter halos with masses ranging from 10^{10} $M_\odot$ to 10^{13} $M_\odot$.

We find very different responses depending on the halo mass. For small dark matter halos its central galaxy responds in a highly predictable way to small perturbation in the star formation and feedback efficiency. For massive dark matter halos, minor perturbations in the input parameters can induce large fluctuations on the properties of its central galaxy, at least ~ 0.1 in $B-V$ color or ~ 0.5 mag in U or r filter, in a seemingly random fashion. We quantify this behavior through an objective scalar function we call predictability.

We argue that finding the origin of this behavior needs additional information from other approximations and different semi-analytic codes. Furthermore, the implementation of an scalar objective function, such as the predictability, opens the door to quantitative benchmarking of semi-analytic codes based on its numerical performance.

Key words: methods: N-body simulations - galaxies:formation - galaxies:evolution

1 INTRODUCTION

Hierarchical aggregation seems to be at the heart of galaxy evolution. In a cold dark matter universe, as depicted by numerical simulations, its structure grows through subsequent mergers and zero fragmentations. The growth and evolution of galaxies, which are thought to use dark matter as scaffolding, is channeled through this hierarchical aggregation, at least for the most massive structures (Springel et al. 2006).

Notwithstanding all the complexity in the process of galaxy formation and evolution, galaxies still are the most basic population unit in the description of large scale structure in the Universe. And still nowadays much work is being invested in galaxy formation to disentangle the influence of the hierarchical context setup by dark matter from the secular baryonic processes on small scales.

Tackling the problem theoretically implies numerical experiments following large structure dynamics and, at the same time, a description of baryonic processes such hydrodynamics and radiative cooling. This is still very challenging in the usual numerical approach that discretizes space and time, and try to solve a relevant set of equations to capture the physics (Abel et al. 2002; Gottlöber et al. 2006). From the computational point of view it involves achieving an effective resolution spanning at least 5 orders of magnitude in mass, length and time (Norman et al. 2007).

To overcome this barrier the semi-analytic model (hereafter SAM) approach proposes to describe first the nonlinear clustering of dark matter on large scales, and describe later the small scale baryonic physics through analytical prescriptions. The connection between the two scales is provided through the dark matter halo, which is the most basic

⋆ E-mail:jforero@aip.de

unit of non-linear dark matter structure (see Baugh (2006) and references therein).

The non linear clustering of dark matter is described through a merger tree, representing the merging history of a given dark matter halo. The construction methods for merger trees can vary, ranging from Monte-Carlo realizations based on theoretical estimates (Somerville & Kolatt 1999) to the numerical based on N-body simulations (Croton et al. 2006), including also hybrid approaches mixing numerical and analytical techniques (Taffoni et al. 2002).

The different analytic implementations of baryonic process span a wide range of philosophical approaches, physical concepts and numerical implementations. Most of them, nonetheless, constructed from observed correlations in our local patch of Universe.

Regardless of the details of these models, what is general to all of them is the underlying merger trees structure complementary with analytic recipes describing the growth of baryonic structure inside the merger trees. The ignorance respect to the physics included in the analytic recipes is usually represented by scalars, which in most of the cases represent efficiencies of physical processes. This implies that a given realization of a semi-analytic run is completely determined by the dark matter input and the baryonic parameters in the simulation.

Most of the work during the last decade was invested in adding and exploring the effect of these parameters on average quantities, especially the luminosity function. The generic parameters to be used have been more or less settled, and most of the models have achieved a good level of internal consistency by reproducing some key observational features. The confidence on the consistency that can be achieved, and the ease to perform a semi-analytic run, have empowered the modelers to select subsamples and make studies about the most massive galaxies or correlations among populations (De Lucia & Blaizot 2007; Hayashi & White 2007).

As the complexity and interest in semi analytic techniques grow, two relevant issues must be addressed in more detail. First, the issue of error propagation from the incertitude in the input parameters, a factor that might be important in a hierarchical Universe, where the amplification of small initial errors might be important for the most massive and hierarchical objects. Second, the development of objective ways to compare different types of complexity in semi analytic models. This could allow, for instance, the implementation of simple tests with an objective scalar function to measure the model performance.

The objective of this paper is two-fold:

- Propose a methodology to weigh the role of secular baryonic processes in the context of SAMs.
- Propose an objective scalar function that captures the biases and general behavior of semi-analytic models regardless of its detailed implementation.

These two objectives are a result of the same perturbative approach we advocate in this paper. This approach is based on the fact that the only objective information we have to describe the results of a semi-analytic run are the input parameters of the model. In the perturbative approach, we perform semi-analytic runs in the neighborhood of some scalar parameters. This will allow us, as we will show, to get an idea about the limits of our semi-analytic model.

This paper is structured as follows. In Section 2 we describe the structure common to all SAMs and from that point we introduce the concept of perturbations in a semi-analytic model. In Section 3 we introduce the setup for the perturbation experiment of our SAM. We describe in Section 4 the two most relevant qualitative features of the experiment results. We select one of this qualitative results to make a detailed quantitative analysis with three different indices, these results are shown in Section 5. We discuss our results in Section 6.

2 SEMI-ANALYTIC MODELS

2.1 Common features

Semi-analytic models exploit the fact that there are two very different physical scales involved in the process of galaxy formation and evolution. On large scales dark matter and gravity are dominant, while on smaller scales complex radiative processes are central to the development of galactic sized structures (Somerville & Primack 1999; Hatton et al. 2003; Bell et al. 2003; Croton et al. 2006; Monaco et al. 2007).

Inside semi-analytic models all the non-linear dark matter dynamic is described through the merger tree, which represent the process of successive mergers building a dark matter halo. On top of this merger tree, all the complex baryonic physics are implemented through analytic prescriptions derived in most part from observations.

The baryonic processes, in the end, are controlled by a set of scalars, which represent most of the time either an efficiency or a threshold value. From the pure functional point of view, all the baryonic properties $\mathcal{B}$ of a dark matter halo $\mathcal{H}$ are a function of its merger tree $\mathcal{T}$ and the set of scalar parameters controlling the model $\{\lambda_1 \dots \lambda_N\}$.

$$\mathcal{B} = \mathcal{B}(\mathcal{T}, \lambda_1, \dots, \lambda_N). \quad (1)$$

Furthermore, during a semi-analytic run, the set of parameter $\{\lambda_1 \dots \lambda_N\}$ is fixed to be the same for all the halos, all the time. Thus, the trees and the scalar values completely define the outputs.

From the perspective of disentangling the role of different physical elements in the process of galaxy formation, the approach commonly followed is the exploration through a set of different values for the $\{\lambda_i\}$ parameters, taking as a gauge the reproduction of the luminosity function of diverse galaxy populations. This coarse exploration of parameter space have been done until a minimum internal consistency is achieved, a decision based on the success of reproducing a wide set of observational constraints.

Nowadays more and more results of semi-analytic models are being used in the predictive sense, selecting subsamples of galaxies, trying to explain or predict astrophysical quantities of interest based on the results of a semi-analytic model (De Lucia & Blaizot 2007; Hayashi & White 2007). This have been done without an explicit treatment of the potential biases and complications introduced by the semi-analytic model itself.

We are interested in understanding in greater detail the behavior and limits of semi-analytic methods, regardless of its detailed implementation across different codes. Our at-

tempt to deal with the complexity across semi-analytic models, is based on the basic conceptual approach in Eq.1, which is the only structural information we have about how semi-analytic models work.

2.2 Perturbing the model

We intend to measure the effect of perturbations in the model

$$\mathcal{B}' = \mathcal{B}(\mathcal{T}, \lambda_1 + \delta\lambda_1, \ldots, \lambda_N + \delta\lambda_N), \quad (2)$$

where the magnitudes of the perturbations $\delta\lambda_i$ are constrained in such a way that its effect is not significant on the mean population quantities such as the luminosity function, meaning that we are not breaking the broad consistency of the model.

The main objective is measuring the consequences of these perturbation and to use them as a gauge of the model's numerical performance, but also to see how and where are going to emerge the consequences of the perturbations.

If we intend to explore the neighborhood of a given set of parameters $\{\lambda_i\}$ by making runs around $\{\pm\delta\lambda_i\}$, this implies performing 2^N different runs where N is the total number of scalar parameters controlling the model. If we want to explore the neighborhood around m different values for each λ_i, the number of runs becomes m^N.

The number of free parameters in SAMs can be at least 6. Which means that we should deal at least with $2^6 = 64$ different runs to minimally explore the neighborhood of $\{\lambda_i\}$. Performing all these runs over a cosmological volume is unfeasible and perhaps not very useful.

The approach we decided to follow in this paper uses two simplifications. First, we only perform the simulations over a subset of dark matter halos selected at random in a box of cosmological size. Second, we explore only the neighborhood of two scalar parameters.

The size of the halo subsample is about 1% of the total number of halos in the simulated cosmological box, and the parameters we will explore are the star formation efficiency α and the supernova feedback efficiency ϵ.

3 EXPERIMENT

The semi-analytic model we use in this paper is a slightly modified version of that presented in Hatton et al. (2003). As we do not want to compare our results with observations or another models, we briefly review only the elements relevant for our discussion: the dark matter description and the star formation and supernovae feedback implementations.

3.1 Dark Matter

The dark matter simulation was performed using cosmological parameters compatible with a 1st year WMAP cosmology (Spergel et al. 2003) (Ω_m, Ω_Λ, σ_8, h) = (0.30, 0.70, 0.92, 0.70), where the parameters stand for the density of matter, density of dark energy, amplitude of the mass density fluctuations and the Hubble constant in units of 100 km s^{-1} Mpc $^{-1}$. . The simulation volume is a cubic box of side $100h^{-1}$ Mpc with 512^3 dark matter particles, which sets the mass of each particle to $5.16 \times 10^8\ h^{-1}$ M$_\odot$. The simulation was evolved from an initial redshift $z = 32$ down to redshift $z = 0$, keeping the particle data for 100 time-steps.

For each recorded timestep build a halo catalogue using a friends-of-friends algorithm (Davis et al. 1985) with linking length $b = 0.2$. Only the groups with 20 or more bound particles are identified as halos. This sets the minimal mass for a dark matter halo to $1.03 \times 10^{10}\ h^{-1}$ M$_\odot$. These halo catalogues provide the input for the construction of the merger trees used as input for the semi-analytic model.

3.2 Star Formation and SN feedback

The star formation rate is set proportional proportional to the mass of cold gas, and without any other characteristic time scale we impose that the rate at which the gas is consumed to form stars is given by the dynamical time of the disc. This is motivated by the observational correlations observed by Kennicutt-Schmidt (Kennicutt 1998).

Hence, in our model, the global star formation rate $\Psi_\star$ on galactic scales is given by the following equation

$$\Psi_\star = \alpha \frac{m_{gas}}{t_{dyn}}, \quad (3)$$

where α is an efficiency parameter, and t_{dyn} is the dynamical timescale of the component we are interested (disc or bulge). For t_{dyn} we use the time taken for material at the half-mass radius to reach either the opposite side of the galaxy (disc) or its center (bulge), and is given by:

$$t_{dyn} = r_{1/2} \times \pi v^{-1}, \quad (4)$$

where v is a characteristic velocity in the galaxy component and $r_{1/2}$ is the half mass radius. For discs the velocity v is equal to the circular velocity of the disc where the material is assumed to have purely circular orbits. In the case of spheroidal components v is the velocity dispersion, where we assume the matter in the component has only radial orbits.

The star formation is triggered if the column density of the gas is greater that a given threshold constrained by the observations Kennicutt (1998). By simplicity we assume that the initial mass function is universal at all redshift and follows a Kennicutt initial mass function.

Once stars are formed, the massive stars will explode inside the galaxies ejecting hot gas and metals in the interstellar medium. The simple model that we use for this phenomenon is given by the implementation of Silk (2001), where the rate of gas mass loss is written assuming an stationary model

$$\frac{dm_{out}}{dt} = \Psi_\star \times \eta_{SN} \Delta m_{SN} \times (1 + L) \times (1 - e^{-R}), \quad (5)$$

where $\Psi_\star$ is the star formation rate, η_{SN} is the number of supernovae per unit mass of formed stars (fixed number function of the IMF), Δm_{SN} is mean mass loss of one supernova (~ 10 M$_\odot$) and $(1 + L)$ is defined as

$$1 + L \equiv \epsilon \frac{m_{gas}}{m_{gal}}, \quad (6)$$

where m_{gas} and m_{gal} are the gas and total mass of the galaxy component and the parameter ϵ regulates the

efficiency of the feedback. We also define the porosity of the galaxy component as

$$R = \alpha \left(\frac{m_{gal}}{m_{gas}}\right)^{1/2} \left(\frac{17.8}{\sigma}\right)^{2.72}, \quad (7)$$

where σ is a typical dispersion velocity in the interstellar medium (in km/s), which we fix to 10km/s for disks and to the velocity dispersion for the spheroidal components. The parameter α is the star formation efficiency. In this model, usually the ejected amount of gas is of the same order of the mass of formed stars.

3.3 Experiment Setup

From the detected halos at redshift zero we select at random nearly 600 of them, which corresponds to about 1% of the total number of halos in the box. For each halo we have its corresponding merger history, and we are able to run our galaxy formation code on every individual merger tree[1].

For each halo we make 320 runs varying two parameters, the star formation efficiency α and the supernova feedback efficiency ϵ. The first parameter, α, is sampled at 16 points between 0.018 and 0.022, and the second, ϵ, is sampled at 20 points between 0.18 and 0.20. From this we define a 2-dimensional representation to keep track of each run. We setup two coordinate plane. Along the first dimension (x-axis) we vary the star formation efficiency α, and along the second dimension (y-axis) we vary the supernova feedback efficiency ϵ. This defines the α-ϵ plane.

For each run (for a given halo and for given point in the α-ϵ plane) we select the central galaxy in the halo, which is the only galaxy with a clear identity in a hierarchical paradigm. For each central galaxy we track six physical properties: total mass (gas and stars), mass of stars , bolometric luminosity, absolute magnitude in the SDSS_U, SDSS_r filters and the $B-V$ color. We will refer to the values of a given galactic property, for a given galaxy, over the α-ϵ plane as a *landscape*.

4 QUALITATIVE RESULTS

We present in Fig.1 the landscapes for the total galactic mass, stellar mass and SDSS_U absolute magnitude for two galaxies, each column representing one galaxy. Qualitatively speaking, we can spot a striking difference in this figure.

The left column in Fig.1 presents the results for a central galaxy in halo of mass $\sim 10^{11}\mathrm{M}_\odot$. Its growth process have been dominated by what we call smooth accretion, meaning that at our working resolution this halo has not suffered any major merger. The predicted properties vary smoothly over the α-ϵ plane.

On the right column in the same figure, we show the same properties for the central galaxy in a halo of mass $\sim 10^{13}\mathrm{M}_\odot$. In this case the values over the landscape do not follow any pattern. The biggest difference with respect to the previous case is that the halo growth cannot be described by pure accretion, but through repeated mergers.

A second qualitative feature is the emerging bimodality for some landscapes. It is visible in the upper right panel in Fig.1, where it seems that the values over the landscape are oscillating back and forth between two planes. To illustrate better this effect we have constructed the histograms for two kind of landscapes (SDSS_r and SDSS_U magnitudes) for four different halo masses. The results are shown in Fig.2, which shows how the landscapes are not necessarily unimodal. By visual inspection of half of the landscapes for the total mass, bolometric luminosity and SDSS_r, we can report that the non-unimodality is a recurrent landscape feature.

For the rest of the paper we will be concerned with a quantification of the first result, which showed an apparent randomness for the central galaxies in strongly hierarchical halos. We will use three different indicators.

First, we will define a scalar function called predictability, P, for a given galactic property over the α-ϵ plane (Pascual & Levin 1999). The predictability will be almost one for the low mass case, and zero (or even negative) for the case of the massive halo.

The second method of quantification is based on the predictability and the variance over the landscapes. We will calculate a predictability-weighted variance, which is intended to represent a quantitative estimation of the variations we can expect in a galactic property after performing a minimal perturbation $\delta\alpha$-$\delta\epsilon$.

The last method of quantification compares the variance over the landscapes with the variance over a subsample of galaxies hosted by halos of similar mass inside the full cosmological box.

5 QUANTITATIVE RESULTS

5.1 Predictability

We present the first part of the qualitatively results using a scalar function we call predictability. First, we sketch out the general idea behind its definition.

We place ourselves on the α-ϵ plane, and we want to predict the value of some galactic quantity at the point we are standing, we also intend to use the information available in the neighborhood. We have the values of the quantity we want to measure for the four nearest neighbors in the α -ϵ plane. We make a guess for that value by averaging these values, and at the same time we perform the measurement.

We have now two different values at the point in the α -ϵ plane, one is predicted and the other is measured. If the squared difference between these two values is small for each point in the plane, we can be sure that we are over a smooth landscape. If the squared differences over the plane are big, the landscape is not so smooth. The predictability is a measure based on these squared differences.

In practice, we use a discretization of the plane α-ϵ, and we construct two different scalar fields over that plane. The first corresponds to the field measured in the numerical runs, noted L. The second is a predicted version, noted L'.

The values of $L'(\alpha_i, \epsilon_j)$ are calculated from the neigh-

[1] Actually because of technical reasons the code is run over a bundle of merger trees.

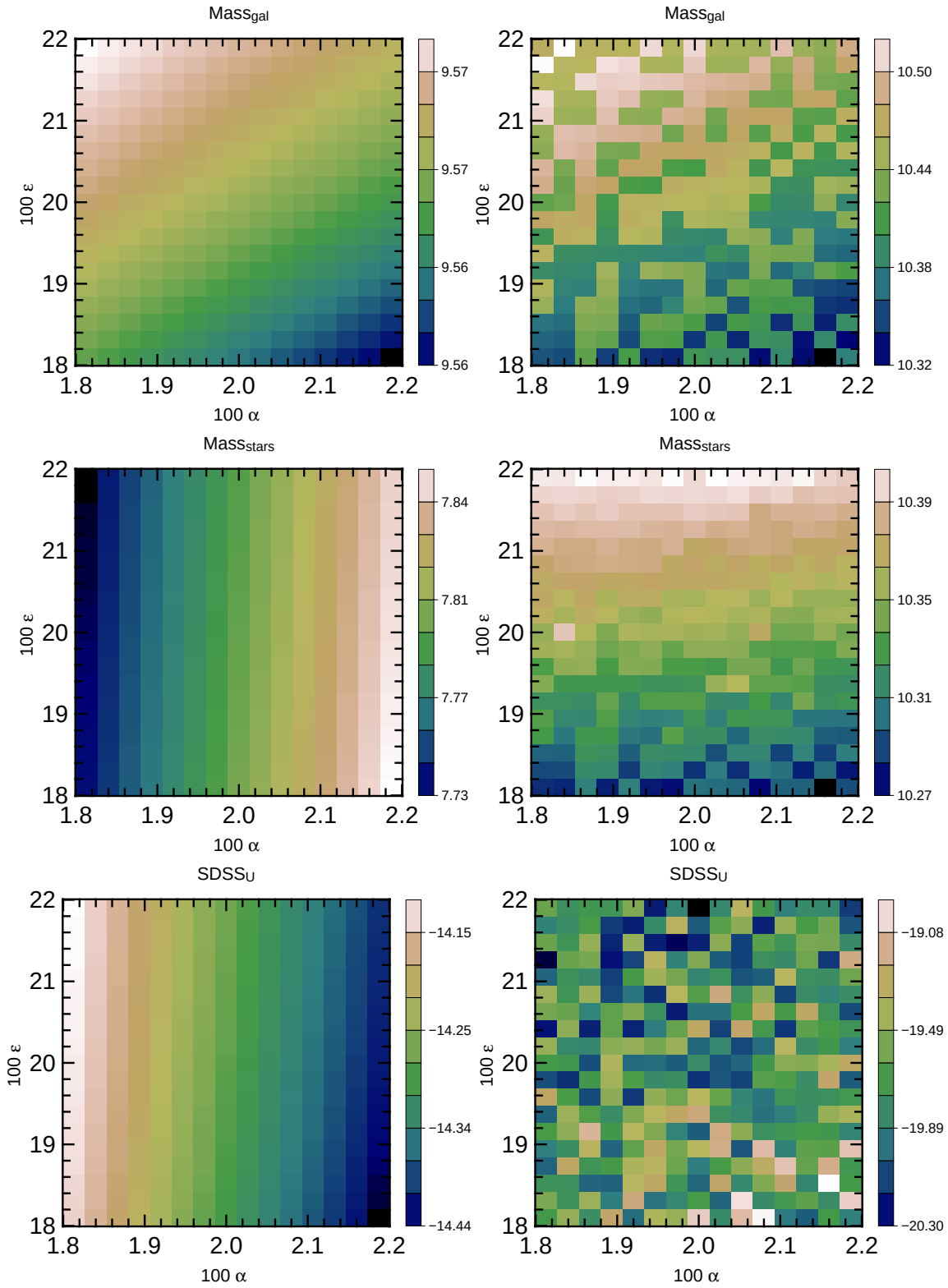


Figure 1. Results for the sampling of the star formation efficiency α and the supernova feedback efficiency ϵ. Each panel represents the results of some galactic property in the central galaxy of a dark matter halo. On the left, the results correspond to the central galaxy in a halo of mass $\sim 10^{10}$ $M_\odot$. On the right, to a more massive halo of $\sim 10^{12}$ $M_\odot$. For each galaxy we show the results concerning the total galaxy mass (upper panels), the stellar mass (middle panels) and the magnitude in the SDSS$_U$ band (lower panels). Every small square in each panel shows the result at redshift $z = 0$ for the run with the corresponding value of α and ϵ. The results for the low mass halo are predictable, for the high mass halo they are almost random. Note that for instance in the case of the SDSS$_U$ filter the values fluctuate over a range of ~ 1.4 mag. The bulk of our paper is devoted to the quantification of this behavior as a function of halo mass.

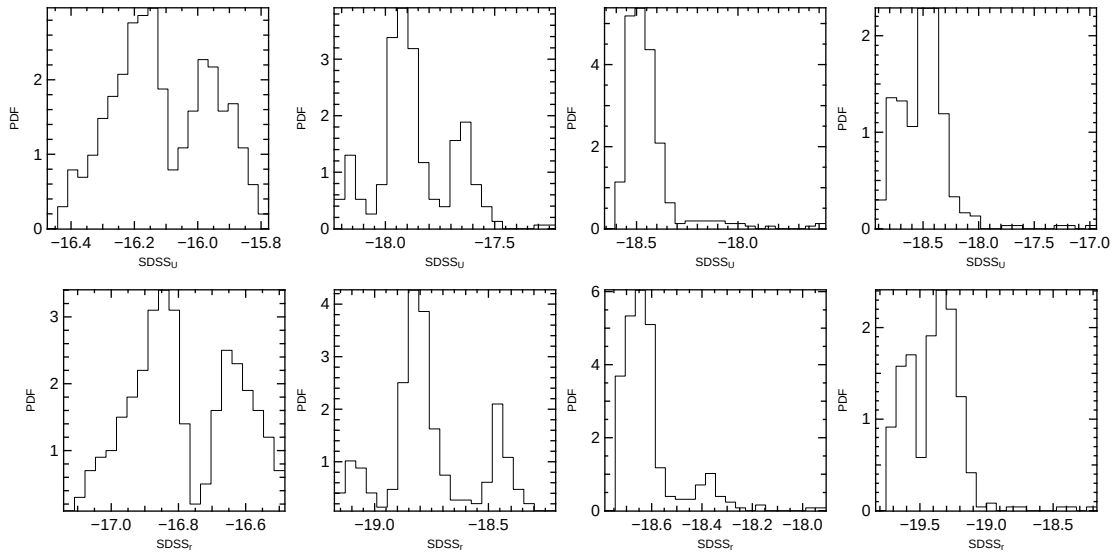


Figure 2. Histograms (normalized to add up to unity) of the values over four randomly selected the landscapes. From left to right the mass of the host dark matter halo increases. Upper row shows the results for the SDSS$_U$ magnitudes. Lower row: SDSS$_r$ magnitudes. This illustrates another qualitative feature of the landscapes, namely that sometimes they are bimodal, for instance in the upper-left panel. We do not try to quantify this behavior in the paper.

boring points in $L(\alpha, \epsilon)$, as follows

$$L'(\alpha_i, \epsilon_j) = \frac{1}{4}[L(\alpha_{i+1}, \epsilon_j) + L(\alpha_{i-1}, \epsilon_j) + L(\alpha_i, \epsilon_{j+1}) + L(\alpha_i, \epsilon_{j-1})]. \quad (8)$$

We construct now the following quantity

$$Q^2 = \frac{1}{N} \sum_{i,j} [L'(\alpha_i, \epsilon_j) - L(\alpha_i, \epsilon_j)]^2, \quad (9)$$

where N is the total number of points in the plane α-ϵ. This quantity help us to define the predictability

$$P = 1 - \frac{Q^2}{\sigma^2}, \quad (10)$$

with σ^2 as the variance of the landscape

$$\sigma^2 = \frac{1}{N} \sum_{i,j} [L(\alpha_i, \epsilon_j) - \bar{L}]^2. \quad (11)$$

The predictability is bounded to $P \leqslant 1$. A value of $P \sim 1$ implies that the landscape is very smooth, while for values $P \leqslant 0$ the changes from neighboring sites can be high.

We now turn to the results of the Fig.3, where we plot the predictability as a function of the logarithm of the mass of the host halo, for all the galaxies in our study. Starting with the total galaxy mass (stars and the gas) we can see that the galaxies have high predictability, $P > 0.9$, in most of the cases. The situation is quite different for the stellar mass and the bolometric luminosity. In these cases the predictability ranges almost evenly between $0 < P < 1$, and we start seeing some fraction of points with negative predictability. In the case of the $(B - V)$ colors and SDSS$_U$, SDSS$_r$ magnitudes we are in a totally different ballpark as most of the landscapes have negative predictability, with a few points over the range $0 < P < 1$.

The conclusion after these results is that we spot a landscape with a very predictability $P < 0.9$ we can be sure that the galaxy is sitting in halo less massive than 3×10^{11} M$_\odot$. In the same vein, when picking the central galaxy in a halo of mass $> 10^{12}$ M$_\odot$, surely the predictability is going to be lower $P < 0.9$, or negative in the case of $(B - V)$ colors and SDSS$_r$, SDSS$_U$ magnitudes.

5.2 P-Weighted Landscape Variance

We explore now the second way of quantification of our results. It is based on the landscape variance $\sigma_{\alpha\epsilon}$ over the the 320 points in the α-ϵ plane (Eq.11) and the predictability P (Eq.10).

We want to weight the landscape variance by the information obtained through the predictability P. Performing a normalization in this way we can have an idea about how much should be expected to vary a given galactic property after performing a perturbation $(\delta\alpha, \delta\epsilon)$.

Using the variance $\sigma_{\alpha\epsilon}$ alone can be misleading in the case of a high predictability landscape, because it could overestimate the variation of performing a $(\delta\alpha, \delta\epsilon)$ perturbation.

We propose then, the P-weighted variance

$$\sigma_{\alpha\epsilon P} = (1 - e^{P-1}) \times \sigma_{\alpha\epsilon}, \quad (12)$$

which has the property of being bound between $0 \leqslant \sigma_{\alpha\epsilon P} \leqslant \sigma_{\alpha\epsilon}$ for the possible values of the predictability $-\infty < P \leqslant 1$.

The general trend (Fig.4) shows a growth in the P-weighted predictability with halo mass, consistent with the fact that the largest values for the predictability come along with large values for the landscape variance. The results for the total mass landscape and the bolometric luminosity stand apart, as this mass trend is less clear than for the other galactic properties.

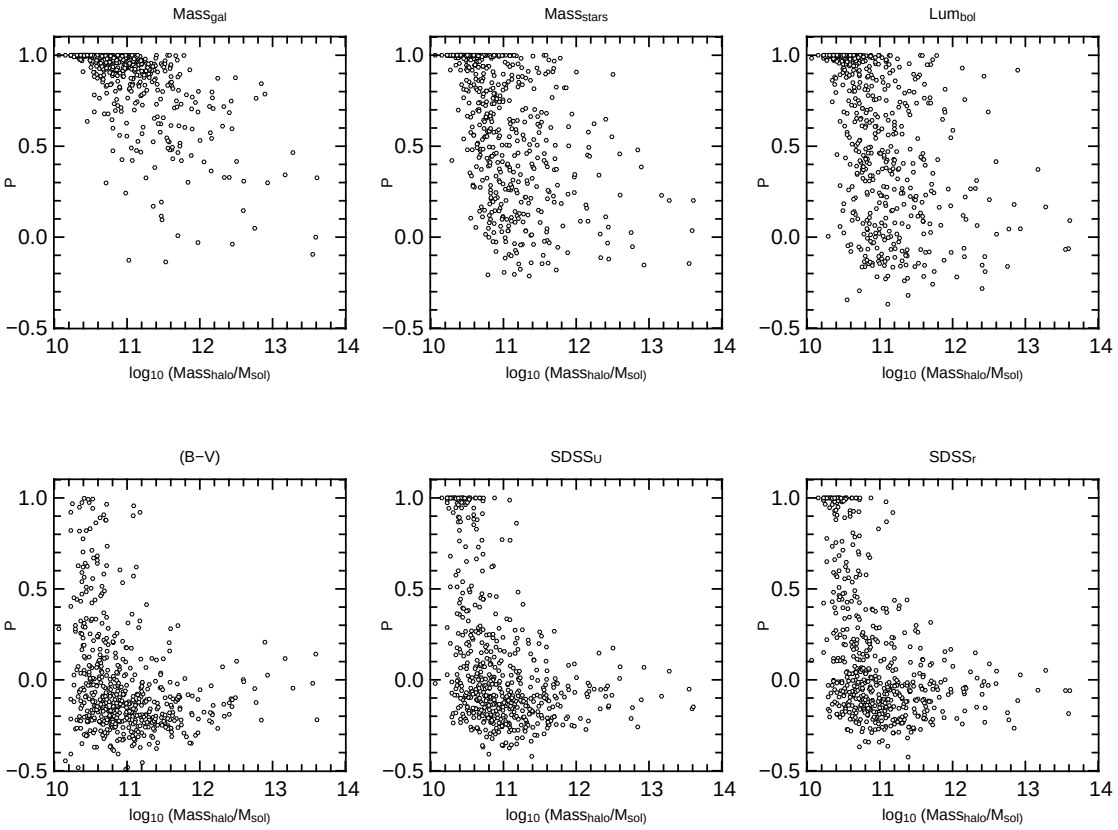


Figure 3. Results for the predictability of different properties. Each point represents a central galaxy. From left to right, top to bottom: total mass of the galaxy, stellar mass, bolometric luminosity, $(B-V)$ color, SDSS_U and SDSS_r magnitudes. The most dependent quantities on the star formation history (colors and magnitudes) are observed in general to have a very low predictability. The most predictable quantity respect to variations in the α-ϵ plane is the total galactic mass.

The values can be used to measure the variation one can expect from a 1% perturbation in (α, ϵ). If we recall that this is calculated from a 1-σ variance, the value of $\sigma_{\alpha\epsilon P}$ that could effectively bracket the fluctuations over the α-ϵ plane would be three times that value. It means that for the most massive halos of mass $\sim 10^{13}$ $\mathrm{M}_\odot$ one can expect, at least, a variation of ~ 0.5 in the SDSS_U magnitudes or ~ 0.1 for the $(B-V)$ colors after 1% variation in (α, ϵ). The variation in the total and stellar mass could achieve at least ~ 0.1 dex.

5.3 Cosmological Variance

Now we compare the landscape variance $\sigma_{\alpha\epsilon}$ with the scale imposed by the cosmological context. We select from the simulated cosmological box (with α and ϵ in the center of the α-ϵ plane) all the dark matter halos with the same mass (within 1%) as the parent halo of the galaxy under study. From this halo population, we calculate the variance σ_{halo} of the galactic properties of our interest for its central galaxies.

We show in Fig.5 the logarithm of the ratio of the two variances $(\sigma_{halo}/\sigma_{\alpha\epsilon})$ as a function of halo mass. For all the cases (except the total galactic mass) we observe the the ratio $\sigma_{halo}/\sigma_{\alpha\epsilon}$ diminishes as the halo mass grows.

The case for the $B-V$ color and the SDSS_r and SDSS_U band magnitude seems special. For high masses the variance ratio is of the same order of magnitude ($\sigma_{halo}/\sigma_{\alpha\epsilon} < 3$, $\log_{10}(\sigma_{halo}/\sigma_{\alpha\epsilon}) < 0.5$) in most of the cases.

This suggests that for central galaxies in low mass halos (where the predictability tends to be high) the variance over its properties is dominated mostly by the possible mass of the host halo. While for high masses (where the predictability tends to be low) the variance for the galactic properties can be equally important if we vary the halo mass or if we make a variation in the star formation and feedback efficiencies.

6 DISCUSSION

We have used N-body simulations and a semi-analytic model of galaxy formation to explore the consequences of small perturbations in the input parameters of the semi-analytic model. Specifically we varied the scalar parameters regulating the star formation α and supernovae feedback efficiency ϵ. We followed some physical properties for 600 central to gauge the effect of the perturbations.

This experiment was motivated both by the interest of performing a description of a semi analytic models, making abstraction of the details in the model , and test the significance of this approach to infer the distinctive footprint of the different baryonic processes.

We find that depending on the halo mass, there are two

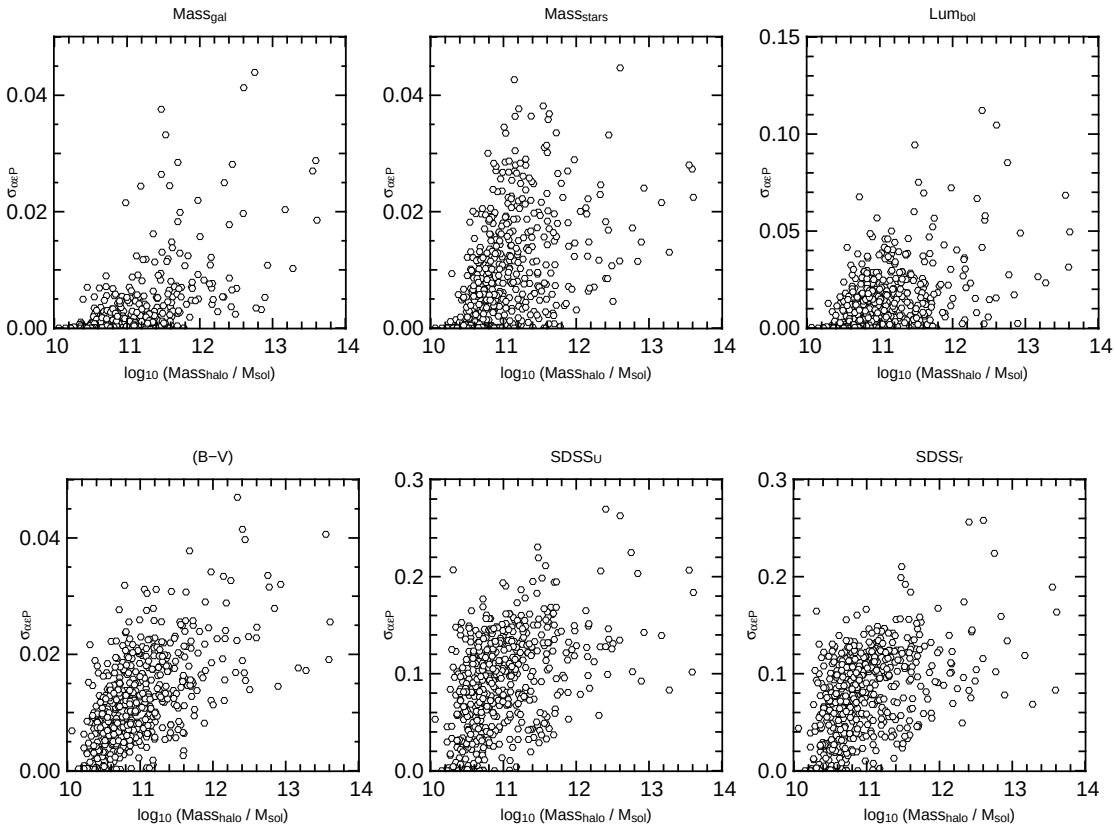


Figure 4. Predictability-weighted variance ($\sigma_{\alpha\epsilon P}$) over the α-ϵ landscape (Eq. 12). This quantity can be used as an estimation of the expected change in a given galactic property when a variation $(\delta\alpha, \delta\epsilon)$ is performed on the input parameters. In the case of magnitudes it can be easily as high as 0.5 for a 1% variation in (α, δ). A trend seems to indicate that this weighted variance increases with the dark matter halo mass.

kinds of different of behavior respect to the change in the input parameters. There is a smooth variation of the galactic properties for low mass halos, and a seemingly random unpredictable behavior limited to high mass halos.

We have quantified this behavior through an objective scalar function called predictability, P, defined in Eq.10. Values $P \simeq 1$ mean a rather smooth and predictable response, while $P \leqslant 0$ point a more unpredictable behavior. The highest predictability is found only in low mass halos while high mass halos present almost always negative values of P. This notion of predictability depends on the property we are looking at. In our case the total galactic mass and the total stellar mass showed the general trend of high predictability on the α-ϵ plane (upper panels in Fig.3). While the $(B-V)$ color and the SDSS_U and SDSS_r bands showed almost always negative predictabilities (lower panel in Fig.3).

Then, we computed the variance $\sigma_{\alpha\epsilon}$ over the α-ϵ, $\sigma_{\alpha\epsilon}$ plane and weighted it by the predictability over the same landscape. This helped us to estimate the possible variation in a given property after performing a change $\delta\alpha - \delta\epsilon$ in the input parameters. Using this measure we found that for high mass halos on should expect rather large variations in the galactic properties from a small perturbation in the baryonic parameters (Fig.4).

In order to give a scale to the landscape fluctuations for a given galaxy, we compare the landscape variance with the variance in a subset of central galaxies of halos taken from the whole cosmological box simulation. The halos are selected to have similar mass as the parent halo of the galaxies we are studying . The general trend showed that at higher halo masses the variance coming from the modulation of the α and ϵ parameters is on the same order of magnitude as the variation on the galaxy properties over the whole box. The opposite trend is only found for low halo masses (Fig.5).

In this particular case of perturbation related to star formation and supernova feedback, it seems that the quantities that exhibit a dependence on the full star formation history (magnitudes and colours) are the most sensitive to variations of the $\alpha - \epsilon$ parameters. For instance, most of the trends we found are not found for the total galaxy mass.

In general, all this evidence seems to point towards a picture where the central galaxies hosted in massive halos, which have grown mainly through mergers, are the most sensitive to small variations of the baryonic parameters in a way that is comparable of doing a significant variation on the mass of the host halo.

From these results one can expect that if the variation of the others scalar parameters in the model is performed, the predictability P, landscape variance σ_{λ_i} and P-weighted landscape variance $\sigma_{\lambda_i P}$ should be higher than the values quoted in this paper. It is very unlikely that the variations of other parameters could cancel out exactly the influence of the efficiencies α ϵ.

There is a hierarchy of causes for this behavior that

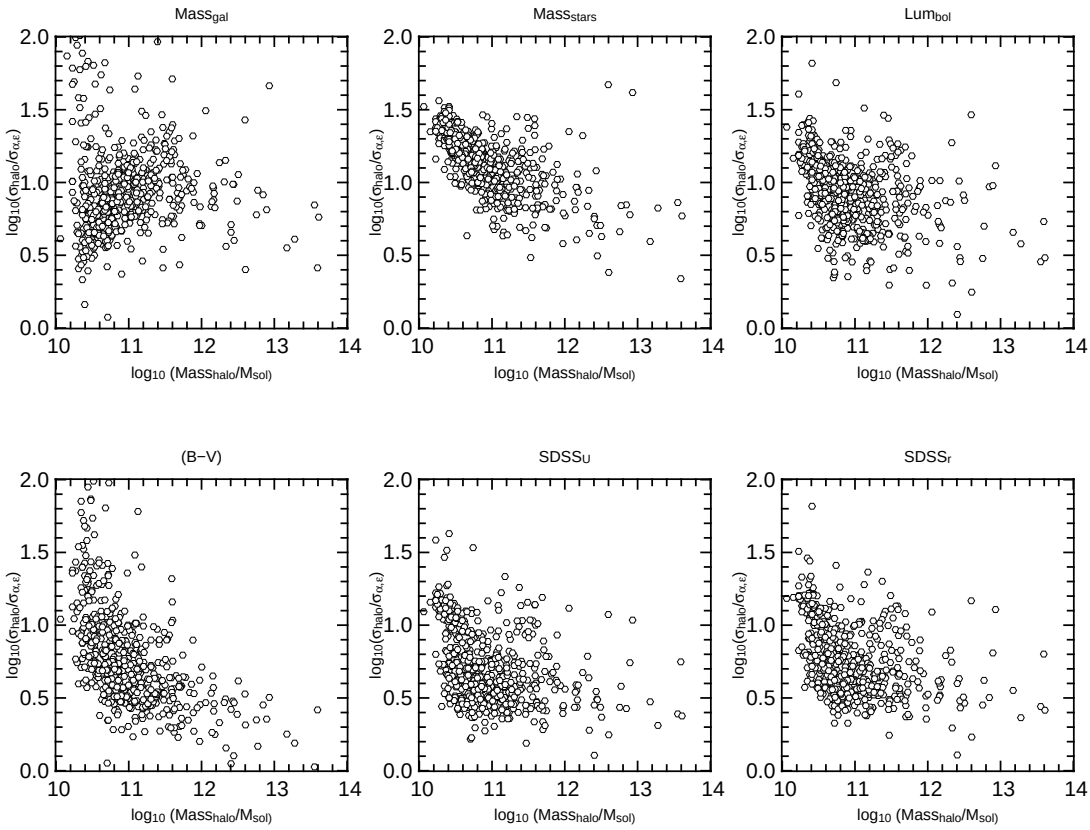


Figure 5. Logarithm of the ratio of the variance calculated over a subset of halos in the simulated cosmological box (σ_{halo}) to the landscape variance $\sigma_{\alpha\epsilon}$. Except for the total galaxy mass, the general trend seems to indicate that the ratio is lower for higher halo masses. For magnitudes and colours, a variation in the α, ϵ input parameters can be as important as a 1% fluctuation in the halo mass. For lower masses, the halo mass fluctuations are dominant.

must be explored. First of all, might this be a signature of the hierarchical build of galaxies? In such a picture it is easy to imagine that mild perturbations at early times might add up to finally yield very different values for very similar initial conditions. This would account for the relatively large values of the variance over the α-ϵ plane compared to the intrinsic variances over the whole population of similar halos in the cosmological volume. But probably not for the low predictability values.

Could this be an artifact coming from the semi-analytic models of galaxy formation? In these models, generally the distinction between what is to be considered as the central galaxy depend on which galaxy is the most massive. This is ambiguous when various galaxies inside a dark matter halo have similar masses, in that case the selection of the central galaxy might be subject to noise. This could explain in part the seemingly random landscapes for high mass haloes.

On the last level of the hierarchy, could this be coming from *our code*? This is impossible to confirm without performing the same kind of experiment with another fully fledged semi-analytic model. Which take us to the issue of comparison between semi-analytic models of galaxy formation. The predictability, as a meaningful scalar objective function, opens the possibility to measure the biases from different semi-analytic codes. This could allow the comparison of different codes based on its numerical performance, going beyond the rather ill-posed strategy of comparison based on astrophysical performance, i.e. reproducing observations.

Finally, the small perturbations we made on the scalar parameters were constructed to not have any effect on the mean properties of the galaxies such as the luminosity function. It means that formally the galaxies we have produced at every perturbation are consistent with the overall galaxy population.

As a consequence of all this, studies making use of selected subpopulations from a wider population generated using semi-analytic models, should bear in mind that this smaller population might not be unique. The dispersion on this subsample of galaxies, coming from the perturbations that can be induced on every parameter in the model, should be explicitly stated. Including that dispersion (in the form of error bars, for instance) seems a necessary condition to make a *fair use* of semi-analytic models, acknowledging in an explicit manner its limitations on predictability.

ACKNOWLEDGMENTS

Many thanks to Dylan Tweed, Julien Devriendt, Jérémy Blaizot, Bruno Guiderdoni and Christophe Pichon for their hard work around the version of the GalICS Code used in this paper. This work was performed in the framework of the HORIZON Project (France).

REFERENCES

Abel T., Bryan G. L., Norman M. L., 2002, Science, 295, 93
Baugh C. M., 2006, Reports of Progress in Physics, 69, 3101
Bell E. F., Baugh C. M., Cole S., Frenk C. S., Lacey C. G., 2003, MNRAS, 343, 367
Croton D. J., Springel V., White S. D. M., De Lucia G., Frenk C. S., Gao L., Jenkins A., Kauffmann G., Navarro J. F., Yoshida N., 2006, MNRAS, 365, 11
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
De Lucia G., Blaizot J., 2007, MNRAS, 375, 2
Gottlöber S., Yepes G., Khalatyan A., Sevilla R., Turchaninov V., 2006, in Manoz C., Yepes G., eds, The Dark Side of the Universe Vol. 878 of American Institute of Physics Conference Series, Dark and baryonic matter in the MareNostrum Universe. pp 3–9
Hatton S., Devriendt J. E. G., Ninin S., Bouchet F. R., Guiderdoni B., Vibert D., 2003, MNRAS, 343, 75
Hayashi E., White S. D. M., 2007, ArXiv e-prints, 709
Kennicutt Jr. R. C., 1998, ApJ, 498, 541
Monaco P., Fontanot F., Taffoni G., 2007, MNRAS, 375, 1189
Norman M. L., Bryan G. L., Harkness R., Bordner J., Reynolds D., O'Shea B., Wagner R., 2007, ArXiv e-prints, 705
Pascual M., Levin S. A., 1999, Ecology, 80, 2225
Silk J., 2001, MNRAS, 324, 313
Somerville R. S., Kolatt T. S., 1999, MNRAS, 305, 1
Somerville R. S., Primack J. R., 1999, MNRAS, 310, 1087
Spergel D. N., Verde L., Peiris H. V., Komatsu E., Nolta M. R., Bennett C. L., Halpern M., Hinshaw G., Jarosik N., Kogut A., Limon M., Meyer S. S., Page L., Tucker G. S., Weiland J. L., Wollack E., Wright E. L., 2003, ApJS, 148, 175
Springel V., Frenk C. S., White S. D. M., 2006, Nature, 440, 1137
Taffoni G., Monaco P., Theuns T., 2002, MNRAS, 333, 623

Appendix 5
Published Paper: Weak Lensing

Abstract

We present the *Lensed Mock Map Facility* (LeMoMaF), a tool designed to perform mock weak lensing measurements on numerically simulated chunks of the universe. Coupling N-body simulations to a semi-analytical model of galaxy formation, LeMoMaF can create realistic lensed images and mock catalogues of galaxies, at wavelengths ranging from the UV to the submm. To demonstrate the power of such a tool we compute predictions of the source-lens clustering effect on the convergence statistics, and quantify the impact of weak lensing on galaxy counts in two different filters. We find that the source-lens clustering effect skews the probability density function of the convergence towards low values, with an intensity which strongly depends on the redshift distribution of galaxies. On the other hand, the degree of enhancement or depletion in galaxy counts due to weak lensing is independent of the source-lens clustering effect. We discuss the impact on the two-points shear statistics to be measured by future missions like SNAP and LSST. The source-lens clustering effect would bias the estimation of σ_8 from two point statistics up to 5% for a narrow redshift distribution of mean $z \sim 0.5$, and up to 2% in small angular scales for a redshift distribution of mean $z \sim 1.5$. We conclude that accurate photometric redshifts for individual galaxies are necessary in order to quantify and isolate the source-lens clustering effect.

Forero-Romero, J. E. and Blaizot, J. and Devriendt, J. and van Waerbeke, L. and Guiderdoni, B., LeMoMaF: Lensed Mock Map Facility, MNRAS, 379, 2007.

Mon. Not. R. Astron. Soc. **379,** 1507–1518 (2007) doi:10.1111/j.1365-2966.2007.12060.x

LEMOMAF: Lensed Mock Map Facility

Jaime E. Forero-Romero,[1]⋆ Jérémy Blaizot,[2] Julien Devriendt,[1]
Ludovic Van Waerbeke[3] and Bruno Guiderdoni[1]

[1]*Université Claude Bernard Lyon 1, CNRS UMR 5574, ENS Lyon, Centre de Recherche Astrophysique de Lyon, Observatoire de Lyon, 9 Avenue Charles André, 69561 St-Genis-Laval Cedex, France*
[2]*Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Strasse 1, 85741 Garching, Germany*
[3]*University of British Columbia, 6224 Agricultural Road, Vancouver, V6T 1Z1, BC, Canada*

Accepted 2007 May 30. Received 2007 May 30; in original form 2006 November 29

ABSTRACT

We present the Lensed Mock Map Facility (LEMOMAF), a tool designed to perform mock weak-lensing measurements on numerically simulated chunks of the Universe. Coupling *N*-body simulations to a semi-analytical model of galaxy formation, LEMOMAF can create realistic lensed images and mock catalogues of galaxies, at wavelengths ranging from the ultraviolet to the submillimetre. To demonstrate the power of such a tool, we compute predictions of the source–lens clustering (SLC) effect on the convergence statistics, and quantify the impact of weak lensing on galaxy counts in two different filters. We find that the SLC effect skews the probability density function of the convergence towards low values, with an intensity which strongly depends on the redshift distribution of galaxies. On the other hand, the degree of enhancement or depletion in galaxy counts due to weak lensing is independent of the SLC effect. We discuss the impact on the two-point shear statistics to be measured by future missions like *SNAP* and *LSST*. The SLC effect would bias the estimation of σ_8 from two-point statistics up to 5 per cent for a narrow redshift distribution of mean $z \sim 0.5$, and up to 2 per cent in small angular scales for a redshift distribution of mean $z \sim 1.5$. We conclude that accurate photometric redshifts for individual galaxies are necessary in order to quantify and isolate the SLC effect.

Key words: gravitational lensing – methods: numerical – large-scale structure of Universe.

1 INTRODUCTION

One of the key goals of modern astrophysical and astronomical research is to map out the spatial distribution of the various matter components of the Universe. In the actual concordance cosmological model, about 85 per cent of the matter in the Universe is thought to be non-baryonic and non-interacting dark matter, while the remaining 15 per cent being composed of baryons (Spergel et al. 2007).

One of the most promising tools to track the distribution of dark matter on cosmological scales is weak gravitational lensing, which has already proven to be well suited for other purposes, such as precision cosmology (Hu 2002; Huterer 2002; Benabed & Van Waerbeke 2004; Bernstein & Jain 2004; Tereno et al. 2005).

Weak gravitational lensing affects the observed galaxy properties such as ellipticities, magnitudes and apparent positions in the sky. In the weak-lensing regime, these effects can only be measured in a statistical sense. Indeed, the first detections of the weak-lensing signal (Bacon, Refregier & Ellis 2000; Kaiser, Wilson & Luppino 2000; Van Waerbeke et al. 2000, 2002; Maoli et al. 2001; Brown et al. 2003; Jarvis et al. 2003) are based on an analysis of the spatial correlation between ellipticities of galaxies. The study of weak lensing through the changes in magnitudes and apparent position in the sky that it causes is an even more challenging measurement (Scranton et al. 2005; Zhang & Pen 2005).

Most of the difficulties in measuring the weak-lensing signal come from observational systematics such as uncertainties in the determination of the point spread function and selection biases (Kaiser 2000; Erben et al. 2001; Vale et al. 2004). There also exist astrophysical errors related to uncertainties in photometric redshift calibration (Ishak & Hirata 2005; Van Waerbeke et al. 2006), intrinsic alignments due to physically associated lens–source pairs (Mandelbaum et al. 2006) or even distant lens–source pairs (Hirata & Seljak 2004). Most of the methods used to assess the importance of all these effects are theoretical or numerical. For the latter, one should also account, in principle, for uncertainties associated with numerical errors inherent to *N*-body simulations, neglected baryonic cooling effects and to a larger extent a poor understanding of the theory of galaxy formation.

The most realistic simulations of weak gravitational lensing performed so far rely on a dark matter distribution taken from an *N*-body simulation and galaxies drawn from a halo occupation

⋆E-mail: forero@obs.univ-lyon1.fr

model (HOM) parametrized by halo mass alone (Heymans et al. 2006; Van Waerbeke et al. 2006), without any information about apparent magnitudes or galaxy colours. Even if this information was available from such models, the mere fact that properties of dark matter haloes depend not only on their masses, but also on their assembly histories – and that the latter have significant effects on galaxy clustering (Croton, Gao & White 2007) – rules out HOMs based on halo mass as precision tools to model effects that rely heavily on a realistic description of the correlations between galaxy and dark matter distributions. The consensus is that an analysis which includes more realistic galaxy populations at various wavelengths is needed.

This paper presents an approach which is a first step to fill this gap and to perform more realistic weak-lensing simulations, thanks to a more realistic description of galaxy populations. It consists in building a tool which we call Lensed Mock Map Facility (LEMOMAF) hereafter, to link together three well-established numerical techniques, each one tackling a different aspect of the problem. More specifically, a state-of-the-art semi-analytic model (SAM) tracks the properties of galaxies within a high-resolution *N*-body simulation as they evolve in time. A light cone is then assembled from the outputs of the simulations to convert 'theoretical' quantities into observables. Finally, a ray-tracing algorithm extracts the weak-lensing signal from the cone. The SAM of galaxy formation that we use in this work is GALaxies In Cosmological Simulations GALICS (Hatton et al. 2003), and mock galaxy maps along with dark matter cones are obtained through a random tiling technique of simulation snapshots, using the Mock Map Facility (MOMAF) (Blaizot et al. 2005) pipeline.

This paper is organized as follows. In Section 2, we review the characteristics of GALICS and MOMAF which are relevant to this study. In Section 3, we explain the weak-lensing formalism and its implementation in LEMOMAF. In Section 4, we present the statistics of the weak-lensing convergence and the effect of weak lensing on differential galaxy counts. Finally, we discuss our results and outline prospects in Section 5.

2 GALICS AND MOMAF

2.1 GALICS

GALICS is a hybrid model of galaxy formation which combines cosmological dark matter *N*-body simulations with a semi-analytic description of baryonic processes. The model is fully described in Hatton et al. (2003), and the version we use here is the same as that used in the previous papers of the GALICS series (Hatton et al. 2003; Blaizot et al. 2004; Devriendt et al., in preparation). We briefly recall the main ingredients in Appendices A and B. Eventually, GALICS outputs are turned into mock catalogues using MOMAF. The following section summarizes how this is achieved.

2.2 MOMAF

MOMAF (Blaizot et al. 2005) is a tool which converts theoretical outputs of hierarchical galaxy formation models into catalogues of virtual observations. The general principle is simple: mock light cones are generated using semi-analytically post-processed snapshots of cosmological *N*-body simulations. These cones can then be projected to synthesize mock sky images.

MOMAF uses a *random tiling* technique described in Blaizot et al. (2005) to build mock observations from the redshift outputs of GALICS. As explained in Blaizot et al. (2005), several different observing cones can be generated from the same set of outputs of GALICS, by changing either the line of sight or the seed for the random tiling. Blaizot et al. (2005) also discuss the bias induced by this random tiling approach on the clustering signal of the final maps.

In this paper, we build 25 cones of galaxies and 25 corresponding cones of dark matter with seeds and lines-of-sight randomly chosen for each observer position, using the same seed for every matching pair of galaxy and dark matter cone. These mock catalogues are $4400\,h^{-1}$ Mpc in depth, and $1\overset{\circ}{.}4 \times 1\overset{\circ}{.}4$ in angular size. The cosmological *N*-body simulation (Ninin 1999) assumes a flat cold dark matter cosmology with a cosmological constant ($\Omega_{\rm m} = 1/3$, $\Omega_\Lambda = 2/3$), and a Hubble parameter $h = H_0/(100\ {\rm km\ s^{-1}\ Mpc^{-1}}) = 0.667$.

These 25 cones allow us to infer an estimate of the dispersion of clustering measurements, that is to say, an estimate of the cosmic variance associated with our mock catalogues. We use galaxy magnitudes in two filters: SDSS *r* and Johnson *K*, because the results in these bands have already been thoroughly examined by Blaizot et al. (2005) and Blaizot et al. (2006). The idea behind this choice is to see the differences between a shallow survey, represented by the SDSS *r* catalogues, and a deep survey illustrated through the Johnson *K* catalogues. The results of the galaxy counts from GALICS–MOMAF and its comparison to observations are shown in Fig. 1, and galaxy redshift distributions for different cuts in magnitude are shown in Fig. 2. In order to make semi-analytic predictions of the weak-lensing statistics (Van Waerbeke et al. 2001) for this broad

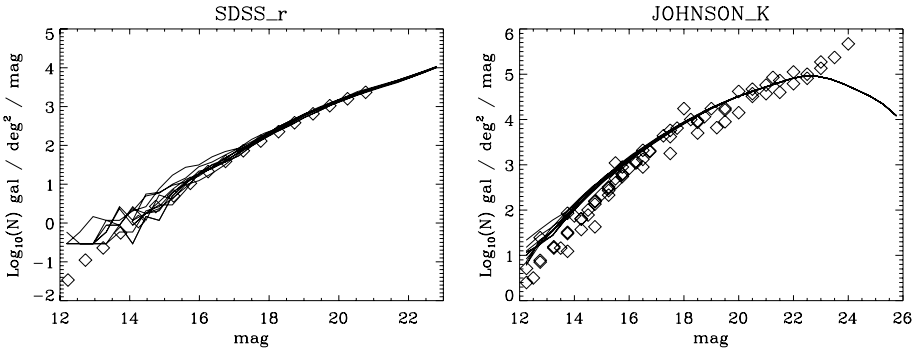


Figure 1. Galaxy counts in the filters SDSS *r* (Yasuda et al. 2001) and Johnson *K* (Djorgovski et al. 1995; Gardner et al. 1996; Moustakas et al. 1997) obtained through the GALICS–MOMAF pipeline. The squares are the observational points, and the lines are the results of the analysis of several virtual light cones built with the simulation. On the right-hand side panel, one can clearly see when the incompleteness in the number of low-luminosity mock galaxies kicks in due to the finite mass resolution of the simulations.

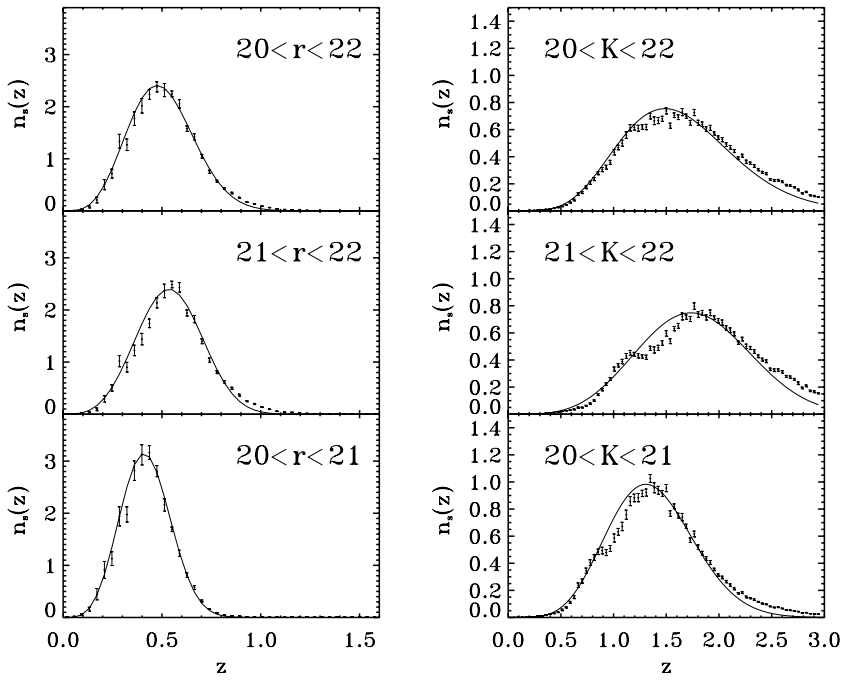


Figure 2. Distributions in redshift for different cuts in r and K magnitude of the galaxy cones. The dots show the mean value of $n_s(z)$ obtained by averaging over the 25 cones, and the error bars show the 1σ error on this mean value. The continuous line is the fit to the function in equation (1).

Table 1. Values of the free parameters in equation (1) for our different redshift distributions.

Filter	Magnitude interval	α	β	z_0
SDSS r	$20 < r < 21$	3.72	2.65	0.36
SDSS r	$20 < r < 22$	3.56	2.20	0.38
SDSS r	$21 < r < 22$	3.24	3.01	0.52
Johnson K	$20 < K < 21$	4.76	2.09	0.88
Johnson K	$20 < K < 22$	5.54	1.41	0.56
Johnson K	$21 < K < 22$	3.08	2.64	1.51

distribution, we describe it with the following functional form:

$$n(z) = \frac{\beta}{z_0\Gamma(1+\alpha/\beta)}\left(\frac{z}{z_0}\right)^{\alpha}\exp\left[-\left(\frac{z}{z_0}\right)^{\beta}\right], \quad (1)$$

where $\Gamma(x)$ is the Gamma function and α, β and z_0 are free parameters to determine for each distribution. The fit values for our mock distributions are in Table 1.

However, we note that given the rather small size of the original N-body simulation box ($100\,h^{-1}$ Mpc on a side) that was used to run GALICS on, our estimate of the cosmic variance is likely to be biased and has to be taken as a lower boundary on the true cosmic errors.

3 LENSED MOMAF

3.1 Weak-lensing formalism

In this section, we provide a summary of the weak-lensing equations relevant for building LEMOMAF and refer the reader interested in a more detailed treatment to Jain, Seljak & White (2000).

We use comoving coordinates and place ourselves in the framework of an isotropic and homogeneous universe that can be described by the Robertson–Walker metric, where in the presence of a perturbative gravitational potential the change in a photon's direction can be written as

$$\mathrm{d}\boldsymbol{\alpha} = -2\nabla_\perp\phi\,\mathrm{d}\chi, \quad (2)$$

where $\mathrm{d}\boldsymbol{\alpha}$ is the photon's deviation, ϕ is the gravitational potential, $\nabla_\perp$ is the gradient in the direction perpendicular to the photon line of propagation and χ is the radial comoving coordinate. A deflection at χ' produces in a plane located at coordinate χ, perpendicular to the line of sight, a deflection of

$$\mathrm{d}\boldsymbol{x} = r(\chi-\chi')\,\mathrm{d}\boldsymbol{\alpha}(\chi'), \quad (3)$$

where $r(\chi)$ is the comoving angular distance, that in a plane universe ($\Omega_\Lambda + \Omega_\mathrm{m} = 1$) is equal to χ. Integrating along the perturbed trajectory of the photon, and then dividing by $r(\chi)$, we obtain the angular position at a position χ:

$$\boldsymbol{\theta}(\chi) = \boldsymbol{\theta}(0) - \frac{2}{r(\chi)}\int_0^\chi \mathrm{d}\chi' r(\chi-\chi')\nabla_\perp\phi. \quad (4)$$

This treatment holds for a single photon. To obtain the effect on an extended object, we calculate the Jacobian of the transformation $(\partial\theta_i(\chi)/\partial\theta_j(0))$, which we call A_{ij} where i and j denote perpendicular directions in the plane perpendicular to the line of sight, in which $\boldsymbol{\theta}$ is also located:

$$A_{ij} = -2\int_0^\chi \mathrm{d}\chi'\frac{r(\chi-\chi')r(\chi')}{r(\chi)}\nabla_i\nabla_j\phi + \delta_{ij}; \quad (5)$$

the Jacobian matrix is usually decomposed as follows:

$$\mathbf{A} = \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2-\omega \\ -\gamma_2+\omega & 1-\kappa+\gamma_1 \end{pmatrix} \quad (6)$$

where κ is the convergence, γ is the shear and ω the rotation.

The relation between the gravitational potential and the mass density perturbation $\delta = \rho/\bar{\rho} - 1$ is given by

$$\nabla^2\phi = \frac{3}{2}\left(\frac{H_0}{c}\right)^2\Omega_\mathrm{m}\frac{\delta}{a}, \quad (7)$$

where H_0 is the Hubble constant at the present, c is the speed of light and a is the expansion factor.

3.2 Multiple plane formalism

The multiple plane formalism consists in dividing the space between the source and the observers into N equidistant (in comoving coordinates) planes perpendicular to the line of sight. Following the convention of Hamana & Mellier (2001), we place ourselves in a cartesian coordinate system noted by (x_1, x_2, y) where the origin is the observer and y indicates the direction of observation. For small angular size fields of view, y can be identified with the radial comoving distance χ. The interplane comoving distance will be called Δy. The projected density contrast on the ith plane located at y_i is given by

$$\Sigma_i(x_1, x_2) = \int_{y_{i-1}}^{y_i} \mathrm{d}y \delta(x_1, x_2, y), \tag{8}$$

which defines a corresponding two-dimensional potential $\Psi = 2 \int \mathrm{d}y \phi$. The position of a light ray in the nth plane can then be found using the multiple-plane lens equation:

$$\boldsymbol{\theta}_n = \boldsymbol{\theta}_1 - \sum_{i=1}^{n-1} \frac{r(\chi_n - \chi_i)}{r(\chi_n)} \nabla \Psi_i, \tag{9}$$

$$\mathbf{A}_n = \mathbf{I} - \sum_{i=1}^{n-1} \frac{r(\chi_n - \chi_i) r(\chi_i)}{r(\chi_n)} \mathbf{U}_i \mathbf{A}_i, \tag{10}$$

where $\mathbf{I}$ is the identity matrix and $\mathbf{U}_i$ is defined by

$$\mathbf{U}_i = \begin{pmatrix} \partial_{11}\Psi_i & \partial_{12}\Psi_i \\ \partial_{21}\Psi_i & \partial_{22}\Psi_i \end{pmatrix}, \tag{11}$$

where ∂_{ij} symbols stand for partial differentiation.

3.3 Ray-tracing algorithm

The algorithm can be divided in to three steps. First, the projection of the dark matter distribution on to the planes. Secondly, the calculation of the potential and its first and second derivatives in these planes, and finally the ray tracing from the observer to the last plane. In practice, things proceed as follows.

(i) Once we have the dark matter particles of the simulation positioned inside an light cone build with MOMAF, we define an orthogonal coordinate system (x_1, x_2, y) at the origin of the cone, where the y-axis is directed from the observer along the symmetry axis of the cone. We then project (along the y-axis) all the particles on to N equidistant planes parallel to the one located at the origin of the cone, which means, for example, that particles originally between (x_1, x_2, y) and $(x_1, x_2, y + \Delta y)$ now sit at (x_1, x_2).

(ii) In each plane we interpolate the overdensity (δ in equation 7), using a cloud-in-cell (CIC) algorithm, over a uniform square grid, G, of N_g cells on a side of comoving size L_side. N_g and L_side are kept identical for all planes. In order to solve the Poisson equation, we pad with zeroes around the grid G, thus creating a grid of size $2 \times N_\mathrm{g}$ in side. We then perform a fast Fourier transform (FFT) of the overdensity grid, and in Fourier space we divide by $-1/k^2$ to inverse transform and obtain the gravitational potential Ψ on the grid. Finally, using finite differences methods we obtain the first and second derivatives of the potential along the x_1 and x_2 directions. This allows us to construct the matrix $\mathbf{U}$ in equation (11) for each point in the plane. A detailed numerical implementation for this step can be found in Premadi, Martel & Matzner (1998).

(iii) We initialize the ray positions over a uniform grid on the nearest plane to the observer. For each ray we interpolate the values of $\boldsymbol{\alpha}$ and the elements of $\mathbf{U}$ using their tabulated values on the CIC grid. Applying equation (9) we then figure out the position of each ray on the next plane, and again interpolate the values of $\boldsymbol{\alpha}$ and $\mathbf{U}$ at this new position to calculate $\mathbf{A}$, and so on and so forth until we reach the final plane. We store all these values, for each ray on every plane.

3.4 Shearing of galaxies

The galaxies in GALICS are represented by three components: disc, bulge and burst. Geometrically speaking, on a mock image, the disc is seen as an ellipse, and the bulge and the burst as circles. The burst is treated as a punctual source and only its magnitude will be modified by weak lensing. The circle and the ellipse can both be parametrized by their shape matrix, according to their weighted quadrupole moments.

For an ellipse of major axis a, minor axis b and with the major axis making an angle β with respect to a horizontal reference axis, the shape matrix can be written as

$$\mathbf{Q} = \frac{1}{\pi a b} \begin{pmatrix} a^2 \cos^2\beta + b^2 \sin^2\beta & (a^2 - b^2)\sin\beta\cos\beta \\ (a^2 - b^2)\sin\beta\cos\beta & a^2 \sin^2\beta + b^2 \cos^2\beta \end{pmatrix}. \tag{12}$$

The case of the circle is recovered by setting $a = b$ and $\beta = 0$. Supposing that the Jacobian matrix does not vary much along the corresponding surface of the galaxy (which is a good approximation in the case of weak lensing caused by large-scale structure), the lensed shape matrix $\mathbf{Q}'$ can be written as

$$\mathbf{Q}' = \mathbf{A}^{-1} \mathbf{Q} \mathbf{A}^{-1}. \tag{13}$$

In this way the lensed properties of the disc and the bulge can be easily found if one knows the Jacobian matrix. In LEMOMAF the $\mathbf{A}$ matrix used to lens a galaxy is the average of four matrices, located at the point of impact of the four closest rays on the nearest plane to this galaxy. We lens bulge and discs separately.

This is used to produce mock lensed images useful for a more realistic treatment of the detection in simulations of the lensing signal through cosmic shear. In this paper, we will not use this capability. We will infer the measurements of κ directly from the numerical values used to modify the galaxy's properties.

3.5 Limits of the method

Most of the limitations of our ray-tracing implementation are resolution issues that come from the use of N-body simulations and the interpolation grids used to trace the light rays. These errors in the calculation of the weak-lensing signal from numerical simulations have been thoroughly quantified before (Jain et al. 2000; Vale & White 2003). We recall these results here, explicitly pointing out the place where the parameters that we use limit the approach the most.

Two relevant scales can be defined to asses the quality of the weak-lensing signal in the simulation. The first one, σ_g, relates to the size of the Fourier grid where the lensing signal is obtained, and the second, σ_n, to the finite resolution of the N-body simulation. We define these quantities as: $\sigma_\mathrm{g} = L_\mathrm{side}/N_\mathrm{g}$, $\sigma_n = L_\mathrm{box}/N_\mathrm{part}^{1/3}$, where L_side is the size of the grid in the plane where we interpolate, N_g its number of cells in one dimension, L_box is the size of the simulation box and N_part is the number of particles inside this box.

If σ_g is larger than σ_n, the power of the signal on small scales will be dominated by the Fourier grid cut-off. If it is smaller, the power of the signal is dominated by noise in the *N*-body simulation. Moreover, if σ_g is not only larger than σ_n but also larger than any significant physical scale in the simulation, the features of the overdensity field that produce the deflection of the rays on that scale are wiped out, and no weak-lensing signal is measured by the ray-tracing method.

We choose a value of $\sigma_g/\sigma_n \sim 1$, guided by the results of previous studies. In our case, since we use a simulation with $L_{box} = 100\,h^{-1}$ Mpc and $N_p = 256^3$, this translates into a resolution of $\sigma_n = 390\,h^{-1}$ kpc. Our ray-tracing simulation uses a grid with $L_{side} = 200\,h^{-1}$ Mpc and $N_g = 800$, which translates into $\sigma_g = 250\,h^{-1}$ kpc.

For a source at a given redshift, the weak-lensing signal probes structures at intermediate redshifts between source and observer. This can be seen from equation (10) where the distance combination $r(\chi_n - \chi_i)r(\chi_i)/r(\chi_n)$ plays the role of an efficiency function for the lensing convergence. This distance combination peaks at an intermediate redshift between the observer and the source as shown in the upper panel of Fig. 3 where it is plotted normalized to the plane distance for three different source positions.

When one takes into account the redshift distribution of the sources used to measure the weak-lensing signal, $n_s(z)$, one realizes that each redshift contributes a different amount to each plane. In order to estimate the joint contribution from the lens efficiency and the redshift distribution of the sources, we consider that each

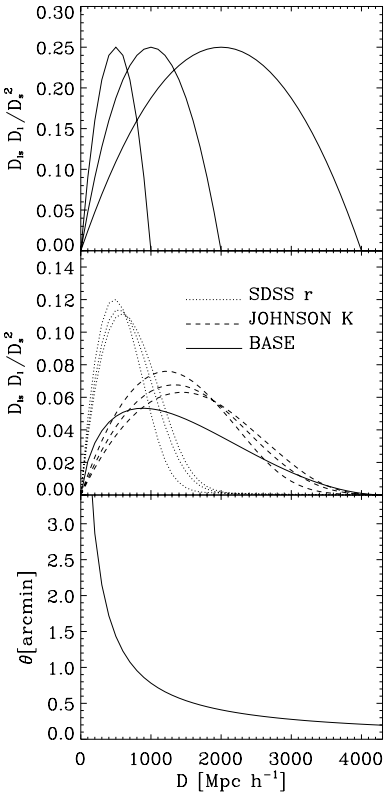


Figure 3. Top panel: distance combination in equation (10) normalized to the distance to the source. The three curves correspond to a distance to the source of 1000, 2000 and 4000 h^{-1} Mpc. Middle panel: sum over the curves of top panel weighted by $n_s(z)$ at the redshift of the source plane. The solid curve, labelled as BASE, shows this sum for a population of galaxies homogeneously distributed in redshift. The curves labelled as Johnson *K* and SDSS *r* show the sum for the respective galaxy distribution in Fig. 2. Lower panel: value of the angular resolution in our ray-tracing simulation as a function of distance from the observer.

efficiency curve is multiplied by the value of $n_s(z_i)$ where z_i is the redshift at which the efficiency curve is calculated, and then we add all the different efficiency curves together. In the middle panel of Fig. 3 we show the results of this calculation for a uniform $n_s(z)$ as well as for the six source distributions presented in Fig. 2.

From the middle panel of Fig. 3, one can see that the distances probed by galaxy populations measured in the SDSS *r* filter lie between 0 and 1000 h^{-1} Mpc (redshifts 0–0.5), while for the Johnson *K* filter this distance interval lies between 1000 and 3000 h^{-1} Mpc (redshifts from 0.5 to 2.0), with a broad peak around 1500 h^{-1} Mpc.

Once the value of σ_g is fixed, we can determine the angular resolution as a function of the distance from the observer, as shown in the lower panel of Fig. 3 for our chosen value of σ_g. From this figure, one can check that in the redshift range probed by galaxies seen in the SDSS *r* filter this angular resolution is of the order of $\sim$1.5 arcmin, and for that probed by galaxies in the Johnson *K* filter of about $\sim$0.5 arcmin.

Discontinuities induced by the random rotations and origin shifts of the tiled boxes in the construction of the cone, using planes that are not perpendicular to the path of the light rays, produce artefacts that were also analysed by Vale & White (2003), concluding that their effects on the lensing signal are negligible. Finally, other errors arise from the translation of dark halo mass resolution into completeness limits at a given magnitude in a given waveband at a given redshift for the galaxy population, or from reduced spatial correlations caused by the random tiling technique. All these limitations have been studied in detail by Hatton et al. (2003), Blaizot et al. (2004, 2005) and Blaizot et al. (2006). In this paper we remain inside these limits to draw our conclusions, but explicitly indicate when they are reached.

4 RESULTS

In this study, we make two kinds of numerical experiments with LEMOMAF: (i) we measure the weak-lensing signal induced by the dark matter cone at galaxy positions, and (ii) we study the change in galaxy counts caused by weak-lensing effects. Before exploring the results of these experiments, we briefly discuss the behaviour of the LEMOMAF ray-tracing module.

4.1 Ray-tracing results

In this section, we focus on the measurements of the moments of convergence (κ) to present the results of the ray-tracing module of LEMOMAF.

The dark matter cones we use to make these measurements are 1.°4 on a side, and have a depth of 4400 h^{-1} Mpc, with a distance of 100 h^{-1} Mpc between lensing planes. In each of these planes, we use a grid with size $L_{side} = 200\,h^{-1}$ Mpc which is split in $N_g = 800$ cells. Weak-lensing properties and galaxy counts are measured only in a central field of 1.°0 on a side. We characterize κ through moments of its distribution function such as its variance $\langle\kappa^2\rangle^{1/2}$, its skewness

$$S_3(\theta) = \frac{\langle\kappa^3\rangle}{\langle\kappa^2\rangle^2} \tag{14}$$

and kurtosis

$$S_4(\theta) = \frac{\langle\kappa^4\rangle - 3\langle\kappa^2\rangle^2}{\langle\kappa^2\rangle^3}. \tag{15}$$

We calculate these moments after smoothing the convergence maps with a circular top-hat filter of angular scale θ.

We are especially interested in the following two results to validate our ray-tracing code.

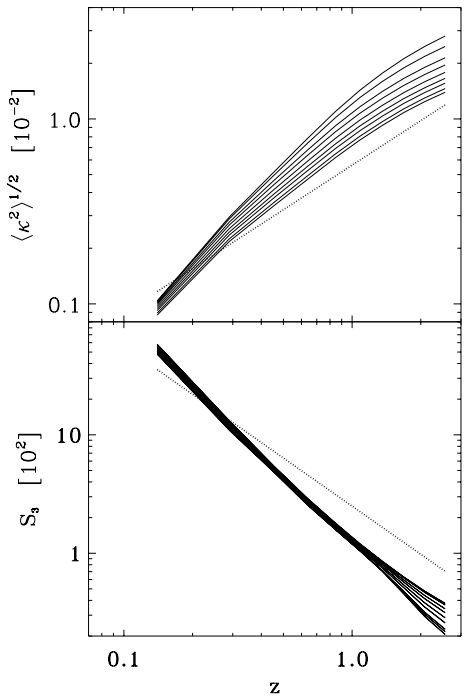


Figure 4. Convergence statistics as a function of redshift. Top panel: variance (divided by 10^{-2}). Bottom panel: skewness (divided by 10^2). The solid lines are the numerical results. Each line represents a different value obtained after smoothing over a different angular scale θ regularly spaced every arcminute from 1 to 9 arcmin. The dotted line shows the theoretical trend predicted by Bernardeau, Van Waerbeke & Mellier (1997). The numerical trend only weakly depends on the smoothing scale θ. The numerical and theoretical results compare fairly well at $z > 1$, but at $z < 1$ the numerical results predict a logarithmic slope about two times steeper than that of analytical theory.

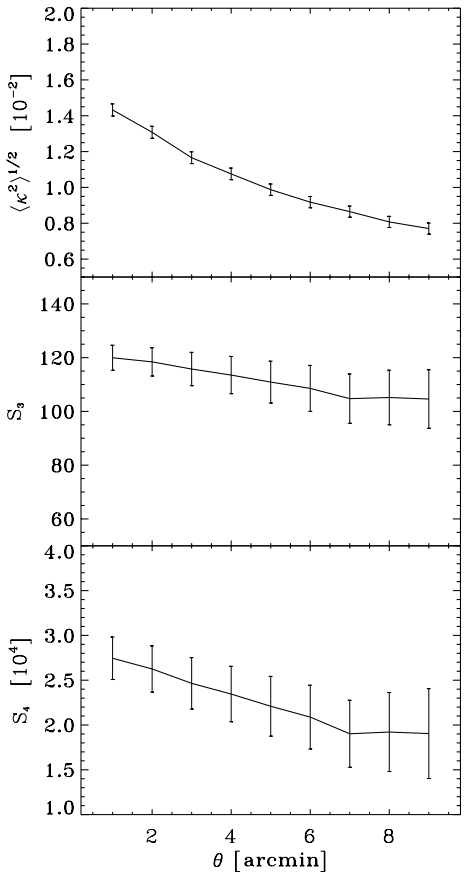


Figure 5. Convergence statistics as a function of smoothing scale. Sources are all located in a single plane at $z = 1$. Top panel: variance (divided by 10^{-2}). Middle panel: skewness (divided by 10^2). Bottom panel: kurtosis (divided by 10^4). The error bars indicate the 1σ dispersion measured around the mean for the 25 cone realizations.

(i) The dependence of $\langle\kappa^2\rangle^{1/2}$ and S_3 as a function of the redshift of the source, for which analytic expressions can be derived.

(ii) The dependence of $\langle\kappa^2\rangle^{1/2}$, S_3 and S_4 on the smoothing scale θ, for a given redshift of the source, for which published values exist in the literature.

In Fig. 4, we show our measurements for $\langle\kappa^2\rangle^{1/2}$ and S_3 as a function of redshift, for nine different smoothing angles θ spaced every arcminute from 1 to 9 arcmin. From this figure, we can see that our computational values follow fairly well the expected theoretical trend. The agreement is better when the source plane is located at redshifts larger than $z = 1$. For closer source planes, located at $z < 1$, the slope of our numerical relation is steeper by about a factor of 2.

In Fig. 5, we plot our results for $\langle\kappa^2\rangle^{1/2}$, S_3 and S_4 as a function of the smoothing angle θ, for sources located in a single plane at $z = 1$ and a field of view of $1^\circ \times 1^\circ$. The order of magnitude of S_3 and S_4 is consistent with a compilation of results for these moments made by Vale & White (2003) for smoothing scales of 4 arcmin. These authors quote an average of $S_3 \sim 135 \pm 10$ and $S_4 \sim (3.5 \pm 0.5) \times 10^4$ for measures obtained using different methods, with error bars reflecting the dispersion amongst reported values. For the same smoothing scale we find $S_3 = 117 \pm 3$, and $S_4 = (2.3 \pm 0.3) \times 10^4$, keeping in mind that uncertainties coming from the numerical simulations themselves can contribute for an extra 10 per cent error in our case, and that theoretical methods compiled in Vale & White (2003) suffer from a similar underestimate.

Being reasonably confident that we are able to produce reliable weak-lensing measurements, we can now couple our ray-tracing code to the outputs of a hybrid model of galaxy formation to study the source–lens clustering (SLC) effect on the convergence statistics, and the effect of weak lensing on galaxy counts.

4.2 Source–lens clustering

In the previous section we measured the statistics of the convergence of rays that uniformly covered a fraction of the sky. In reality, we only have access to the convergence signal measured for galaxies that have specific clustering properties and, even more important, are correlated to the lensing potential. This is known in the literature as the SLC effect.

Theoretical studies of the SLC effect (Bernardeau 1998) predict that it alters the higher order statistics of the convergence. More precisely, while $\langle\kappa^2\rangle^{1/2}$ should remain unchanged according to theory, S_3 should have a lower value. As a matter of fact, S_3 is known to be sensitive to Ω_m almost independently of σ_8, and a combined analysis of the skewness and the variance of the convergence could, in principle, provide new constraints on the values of Ω_m and σ_8. Hence, the cosmological interest to accurately quantify the influence of the SLC effect on the higher order statistics of κ.

Previous numerical work on the SLC effect (Hamana et al. 2002) has focused on its impact on S_3 estimations. This work was carried

out using a simple bias model to paint a population of galaxies on top of the dark matter density field. To our knowledge, our work is the first attempt to use a galaxy population self-consistently derived from an *N*-body simulation to estimate the impact of the SLC effect. More specifically, with MOMAF, we obtain both a dark matter distribution and its matching galaxy distribution in a light cone, this latter being derived by post-processing the dark matter simulation with the GALICS SAM. We then calculate the weak-lensing signal over all the field at all the planes used in the ray-tracing simulation with LEMOMAF, and use this information to shear the properties of galaxies which are in the cone, while storing the value of the convergence computed at each galaxy position. In this section, we therefore return to the analysis of the convergence statistics, but only for the subset of values measured at each galaxy position by interpolation of the values at the neighboring rays.

To quantify the impact of using a self-consistent modelling of the galaxy population on our results, we couple the dark matter and galaxy cones in three different ways. In the first way, that we call MATCH, we shear the galaxies according to their true underlying dark matter distribution. In the second way, called RANDOM, we keep the same pair of galaxy cone/DM cone as in the MATCH case, but erase some of the spatial clustering information by randomizing the positions of galaxies over the sky plane while keeping their distance to the observer constant (i.e. we preserve the galaxy redshift distribution). Finally, in the third case, called NO MATCH, we use the full spatial clustering information for galaxies, but shear them using a different underlying dark matter distribution from the one with which their properties were derived. The idea behind these 'tricks' is that cross-comparisons between the MATCH, NO MATCH and RANDOM methods should provide us with a better understanding of where the impact of the SLC effect on the convergence statistics comes from.

We use two broad-band luminosities for each galaxy, those measured in the SDSS *r*, and Johnson *K* filters. We build 25 different light cones in order to minimize the bias effects induced by the random tiling technique, and maximize the accuracy of the statistics (Blaizot et al. 2005). For each light cone we output both the galaxies and the dark matter distribution. To mimic observational effects as best as we can, we select galaxies on which the lensing signal is to be measured based on their apparent magnitude. The resulting redshift distributions are shown in Fig. 2, where we have arbitrarily split magnitudes in three bins: $20 < m < 21$, $21 < m < 22$ and $20 < m < 22$. Once again, we recall that each light cone is 1.°4 on a side, and has a depth of $4400\,h^{-1}$ Mpc, with a distance between lensing planes of $100\,h^{-1}$ Mpc. In each of these planes we use a grid of size $L_{\rm side} = 200\,h^{-1}$ Mpc split into $N_{\rm g} = 800$ cells. Weak-lensing properties and galaxy counts are measured in a centred field of 1.°0 on a side.

Figs 6 and 8 show the results for the higher order moments of the convergence and the six redshift distributions plotted in Fig. 2. Each linetype corresponds to a different method to construct the maps; RANDOM, NO MATCH and MATCH. We also compare our

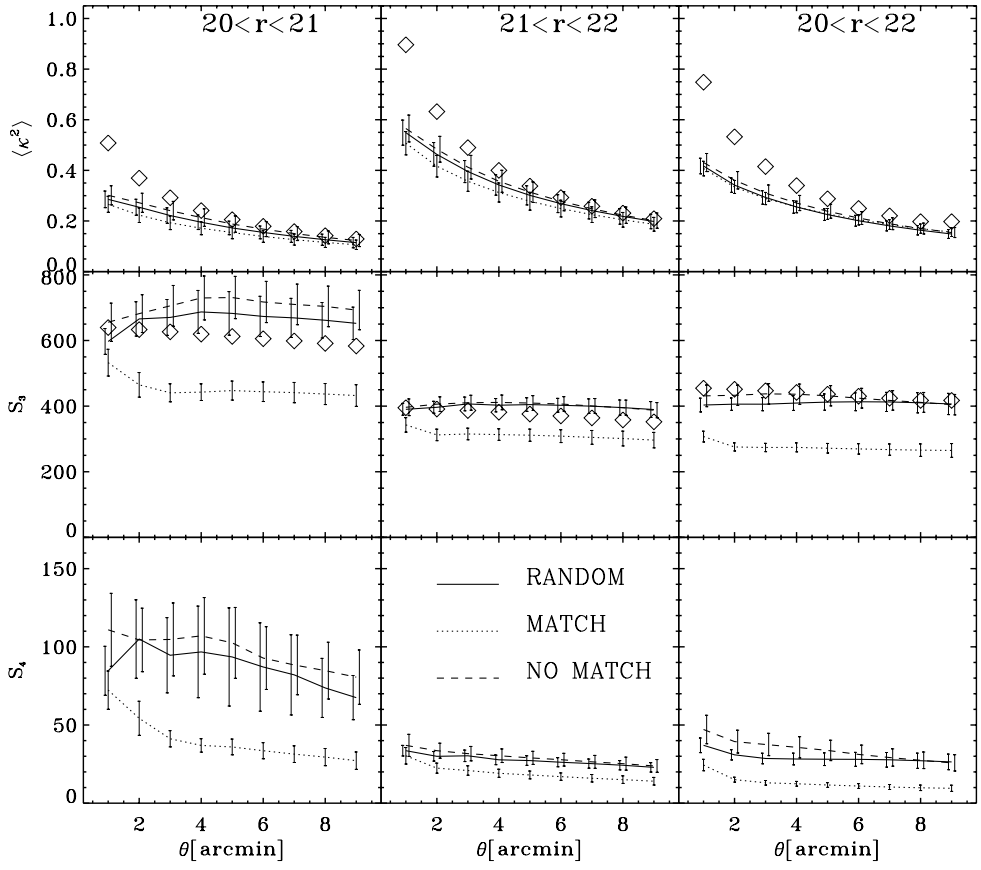


Figure 6. Comparison of the convergence statistics for three different magnitude ranges in the SDSS *r* filter, and for three different methods of measuring the convergence at each galaxy position. The solid line indicates results obtained with the RANDOM method, the dashed line indicates results obtained with the NO MATCH method, and the dotted line indicates results obtained with the MATCH method. $\langle\kappa^2\rangle$ has been divided by 10^{-4} and S_4 has been divided by 10^4. The error bars, shifted horizontally for clarity, indicate the 1σ dispersion measured around the mean for the 25 cone realizations. The overplotted diamonds show values computed with the SAM described in Van Waerbeke et al. (2001).

numerical results to the semi-analytic calculations of Van Waerbeke et al. (2001). Overall, the semi-analytical predictions of these authors show very good agreement with the lensing signal measured in our mock catalogues. Moreover, the better agreement with the RANDOM mocks is somewhat expected: out of the three methods we presented here, this is the one which follows most closely the assumptions made in Van Waerbeke et al. (2001). Most of the small-scale deviations from the semi-analytic trend in the variance plots can be attributed to the finite spatial resolution of our simulation (see Section 3.5).

We introduce the parameter R as done by Hamana et al. (2002) to quantify the amplitude of the SLC effects:

$$R(\mathcal{S}) = \frac{\mathcal{S}_{\rm NO\,SLC} - \mathcal{S}_{\rm MATCH}}{\mathcal{S}_{\rm NO\,SLC}}, \tag{16}$$

where $\mathcal{S}$ can be $\langle\kappa^2\rangle$ or S_3, and NO SLC is one of the methods RANDOM or NO MATCH. Figs 7 and 9 show the results for this expression.

From these figures, it is clear that the results for $\langle\kappa^2\rangle$, S_3 and S_4 in both filters, and in the three different magnitude bins show the

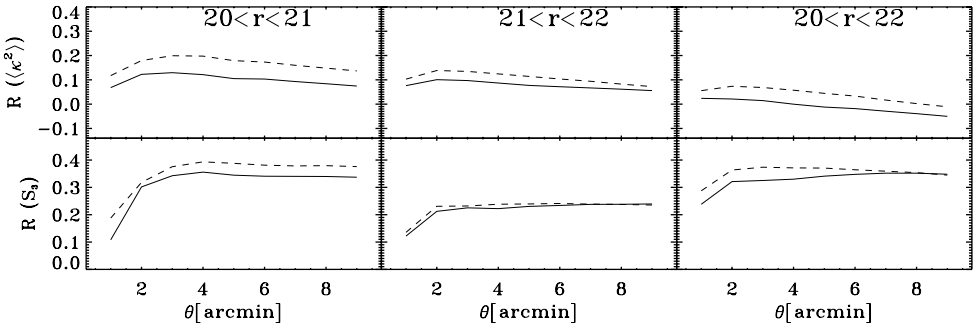


Figure 7. The $R(\mathcal{S})$ factor as defined in equation (16) for $\langle\kappa^2\rangle$ and S_3, for the three different magnitude ranges in the SDSS r filter. The solid line compares the methods MATCH and RANDOM. The dashed line compares the methods MATCH and NO MATCH.

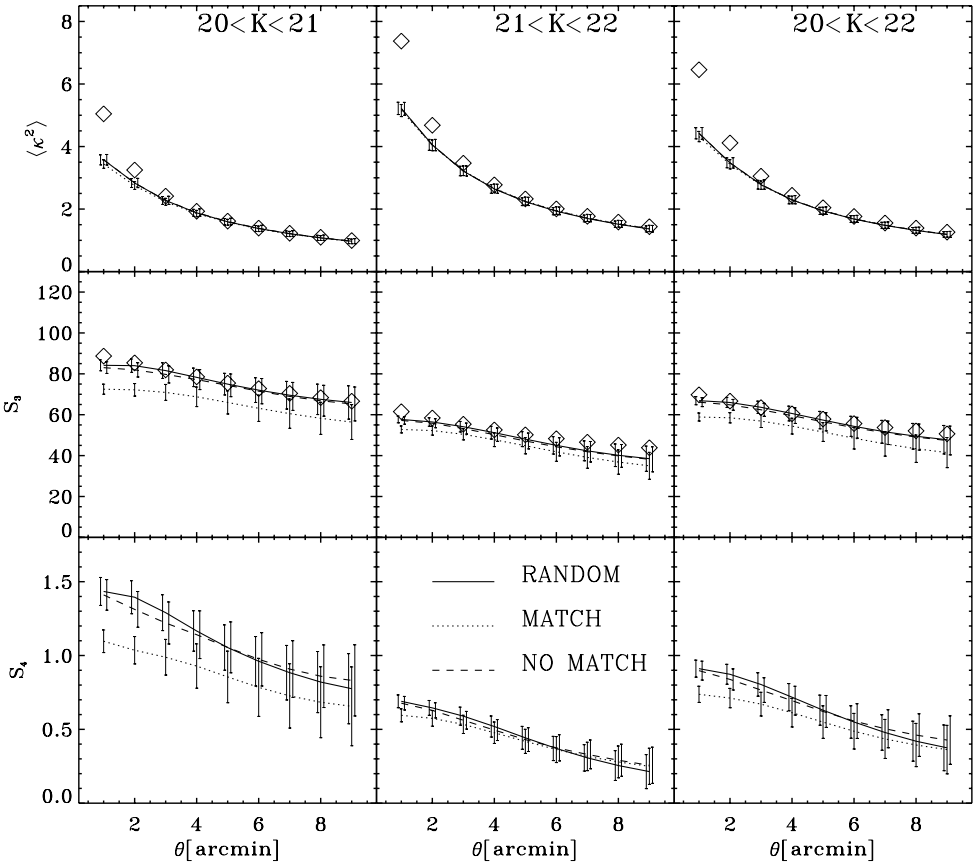


Figure 8. Comparison of the convergence statistics for three different magnitude ranges in the Johnson K filter, and for three different methods of measuring the convergence at each galaxy position. The solid line indicates results obtained with the RANDOM method, the dashed line indicates results obtained with the NO MATCH method, and the dotted line indicates results obtained with the MATCH method. $\langle\kappa^2\rangle$ has been divided by 10^{-4} and S_4 has been divided by 10^4. The error bars, shifted horizontally for clarity, indicate the 1σ dispersion measured around the mean for the 25 cone realizations. The overplotted diamonds show values computed with the SAM described in Van Waerbeke et al. (2001).

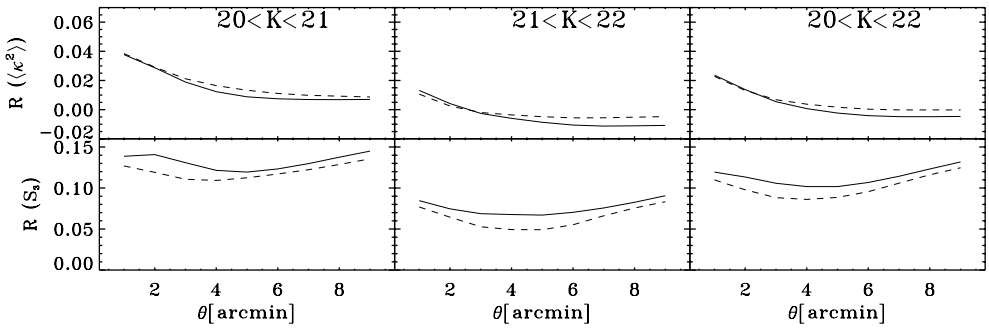


Figure 9. The $R(\mathcal{S})$ factor as defined in equation (16) for $\langle\kappa^2\rangle$ and S_3, for the three different magnitude ranges in the Johnson K filter. The solid line compares the methods MATCH and RANDOM. The dashed line compares the methods MATCH and NO MATCH.

same trend: these statistics are systematically lower for the MATCH case than that for the NO MATCH and RANDOM cases, with the latter being remarkably similar. Furthermore, we clearly see that the effect is more pronounced for the narrower redshift distribution of sources (i.e. SDSS r band). These results, which are obtained using a fairly realistic galaxy distribution, are in broad agreement with the simpler approach advocated by Bernardeau (1998) and Hamana et al. (2002).

However, we also find a significant SLC effect on the two-point statistics, $R \sim 10$ per cent for the SDSS r distributions. The case for the Johnson K distributions is less significant with $R \sim 4$ per cent, being consistent with 0 per cent for smoothing scales larger than 4 arcmin. This is not expected from perturbation theory alone. We suspect that it arises because our simulation probes the highly non-linear regime. A comparison of the MATCH to the NO MATCH run – where the clustering of sources is preserved albeit galaxies positions are not correlated with that of the underlying dark matter distribution – strongly suggests that the signal comes from the adequation of the intra halo galaxy population with the depth of its host potential well.

Finally, we would like to emphasize that such a mock catalogue approach should allow a fast calculation of the SLC effect for future weak-lensing surveys intending to use cosmic shear as a precision cosmology tool.

4.3 Galaxy counts

Weak lensing enlarges the area of the sky which is observed, thus lowering the number density of objects which are detected. In the same time, it causes galaxies to appear brighter, thus increasing their number density for a given apparent magnitude. The net effect depends on the slope $s(m)$ of the number counts of galaxies. Let us write $N_0(m)\,\mathrm{d}m$ as the number of galaxies with magnitudes between m and $m + \mathrm{d}m$, and $s(m) = \mathrm{d}\log N_0(m)/\mathrm{d}m$. If the sources undergo a magnification $\mu = 1/\det \mathbf{A}$ (see Section 3.1 for a definition of the matrix $\mathbf{A}$), and $N'(m)$ is the magnified number of sources corresponding to $N_0(m)$, we can write $N'(m) = \mu^{2.5s-1}N_0(m)$ (Broadhurst, Taylor & Peacock 1995; Jain 2002; Scranton et al. 2005).

The effect of the magnification on the solid angle, which is responsible for a factor μ^{-1} can only be detected if we use a grid that allows us to have a resolution of the order of 0.1 arcmin to make a proper estimation of the deflection angle for a given galaxy. The angular resolution of our simulations is roughly 1 arcmin, which is clearly not enough. Therefore, we expect that only the lensing effects on galaxy magnitudes (which does not require high angular resolution to be seen) will be perceptible. This simplifies the expression for the magnified number of sources to $N'(m) = \mu^{2.5s}N_0(m)$, which in turn reduces, in the weak-lensing regime ($\mu = 1 + 2\kappa$), to

$$\frac{N'(m)}{N_0(m)} - 1 = 5.0\kappa s. \tag{17}$$

The numerical experiments we carry out with LEMOMAF are meant to explore the effect of weak lensing on galaxy counts via the estimate of this ratio, as we vary the size of the field in which the measurements of the counts are performed.

In the 25 original fields of $1^\circ \times 1^\circ$, we compute galaxy counts both for unlensed and lensed fields (using the MATCH and NO MATCH methods). We then cut each of these fields into smaller square patches of angular size 12, 15, 20 and 30 arcmin and 1° on a side, and measure the counts in all of these subfields. We also calculate for each patch the ratio $N(m)/N_0(m)$, and organize our results as follows. For a patch of size θ, taken from a field of original size Θ, we have $N_\mathrm{p} = (\theta/\Theta)^2$ patches. Labelling subfields as $\ell = 1, \ldots, N_\mathrm{p}$ in every original uncut Θ field, we then proceed to stack together the 25 patches which have the same index ℓ. For each patch in the stack, we estimate the ratio of lensed to unlensed integrated counts, and compute the mean for each stack.

In Figs 10 and 11, we plot these means for our N_p stacks, for each value of θ (column panels) and for both methods, NO MATCH and MATCH (top and middle row panels respectively). In the bottom row of these figures, we show $\lambda = 5.0s\kappa$, in order to facilitate the interpretation of the behaviour of the lensed to unlensed counts ratio in terms of equation (17). In this intent, we estimate κ using an average value measured for the galaxies which lie in the magnitude bin of interest in the uncut field of size Θ. From our ray-tracing results (see Fig. 4), we know that we are in the weak-lensing regime where $\kappa \ll 1$, so equation (17) tells us that we can expect a depletion in galaxy counts only when $\langle\kappa\rangle$ and s have different signs.

In the SDSS r band (Fig. 10), λ (in per cent) is restricted to the interval $[-1.0, 0.0]$. High fluctuations in the counts ratio for the low values of the magnitude are due to an ever smaller number of bright galaxies being present in a subfield when the subfield size decreases or the brightness of the source increases. In the magnitude interval $m_r = [19, 22]$ where this effect becomes negligible, we see that galaxy number counts can be depleted down to 1 per cent.

In the Johnson K filter (Fig. 11), λ (in per cent) takes values in the interval $[-0.5, 1.0]$, although its change of sign between the magnitudes 21 and 23 is entirely due to the mass resolution of our N-body simulation which translates into an incompleteness for galaxies fainter than $m_K = 21$ (see also Fig. 1). We therefore restrict ourselves to the magnitude interval $m_K = [16, 21]$ for which there exists a good agreement between modelled and observed counts in terms of slope, and the number of bright galaxies per subfield is high enough.

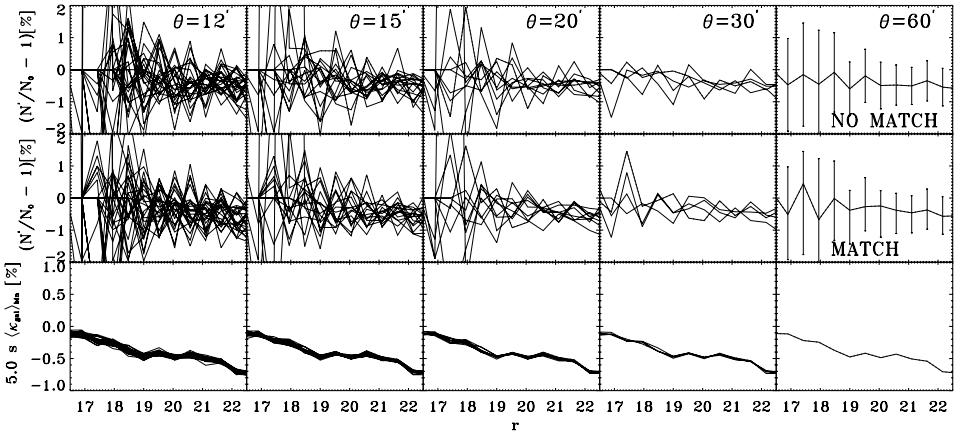


Figure 10. Percentile difference between lensed (N') and unlensed (N_0) galaxy counts in the SDSS r filter, calculated with the NO MATCH (upper panel) and MATCH (middle panel) methods. The lower panel shows the values of $\lambda = 5.0\, s\, \langle\kappa_{\rm gal}\rangle_{\rm bin}$, where s is the logarithmic slope of the counts and $\langle\kappa_{\rm gal}\rangle_{\rm bin}$ is the average value of κ measured from the galaxies in that magnitude bin over the field of size 1° on a side. Each vertical line splits the plot in panels which contain results for different patches of angular size θ on a side, cut inside an original field of size $\Theta = 60$ arcmin on a side. Thus, the number of curves in each row of panels is $N_{\rm p} = (\Theta/\theta)^2$, and each curve represents the mean of the measurement over 25 uncorrelated patches. The rightmost upper and middle panels also show the 1σ dispersion between the 25 Θ fields ($N_{\rm p} = 1$) as error bars.

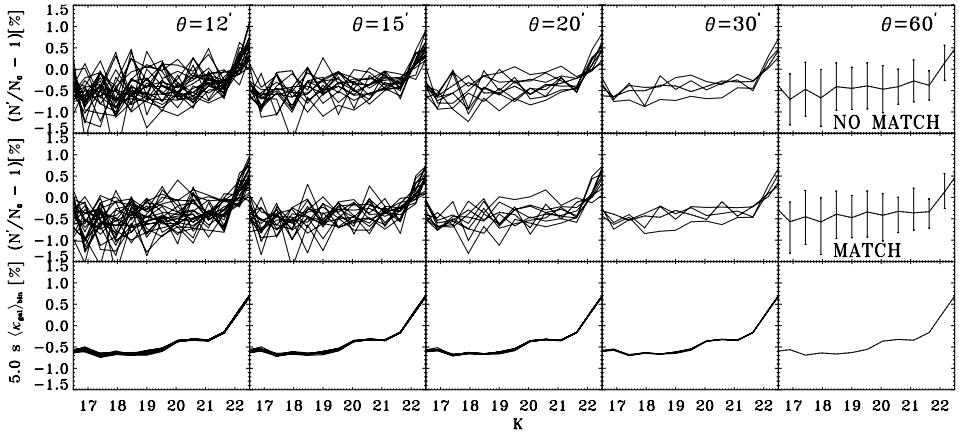


Figure 11. The same as Fig. 10 but for the Johnson K filter. The behaviour for magnitudes $K > 21$ is a numerical resolution artefact which leads to an incompleteness in the number of low-luminosity galaxies (see also the right-hand panel in Fig. 1)

In this magnitude interval, the counts can be depleted down to 0.5 per cent, for low values of m_K.

Unsurprisingly, the dispersion around the theoretically expected ratio increases when the angular size of the field decreases. Finally, Figs 10 and 11 show that the source lens clustering effect does not play an significant role in enhancing galaxy number counts: the NO MATCH and MATCH methods pretty much yield the same quantitative results.

However, from this experiment we confirm our suspicion that the modification of the solid angle is not resolved in our simulations, consequently higher angular resolution must be attained if one hopes to use LEMOMAF in the study of angular correlations induced by cosmic magnification.

5 DISCUSSION AND CONCLUDING REMARKS

In this paper, we presented LEMOMAF which combines dark matter N-body simulations and an hybrid model of hierarchical galaxy formation to make mock lensed images and convergence maps, thanks to a ray-tracing algorithm. More specifically, the results presented here were obtained using a cosmological N-body simulation performed with a treecode, the GALICS model of galaxy formation, the MOMAF pipeline which constructs galaxy and dark matter light cones, and a ray-tracing algorithm through multiple planes to account for the weak-lensing effect. This tool suffers from all the shortcomings inherent to each of these techniques. However, for GALICS and MOMAF these limitations have been carefully identified in a series of papers published over the past four years. As far as the ray-tracing algorithm is concerned, its limitations in terms of the N-body simulation parameters have also been discussed quite thoroughly in the literature. To sum up, these shortcomings impose a limitation on the size of the fields that can be constructed, and the angular resolution that can be reached in the construction of the lensed images and the convergence maps.

Working within the proper interval of validity of these methods, we performed two numerical experiments with LEMOMAF. The first one measured the convergence signal induced by the dark matter density field at galaxy positions in a light cone. Different methodologies for this measurement were implemented with the aim of testing the consequences of the source-clustering effect on the probability density function of the convergence. We found that the SLC effect skews the probability density function (PDF) towards

lower values of the convergence and that, in some cases, it makes this PDF look more Gaussian than that obtained without including SLC, as expected from theoretical considerations. However, even when probing the *same* dark matter distribution, the precise trend of the SLC effect depends sensitively on the specific distribution of the galaxies we consider. For instance, we showed that a narrower redshift distribution of the sources is more sensitive to the SLC effect. This could be problematic for future lensing surveys which intend to perform shear measurement in *thin* redshift slices, a technique called *tomography* (Hu 1999). For the Johnson *K* filter, the SLC effect has an impact at the few per cents level (2 per cent) on the estimations of σ_8 from two-point statistics. For the SDSS *r* filter, the impact is of the order of 5 per cent. This level of contamination was neglected in previous analysis (Bernardeau 1998; Hamana et al. 2002), because its amplitude was well below the largest weak-lensing surveys accuracy at that time (VIRMOS: Van Waerbeke et al. (2000), RCS: Hoekstra et al. (2002)). However, future – nearly full sky – missions like *SNAP* and *LSST* will have to reach a precision of 10^{-3} on the shear measurement, that is, roughly 0.1 per cent from the shear two-point statistics (Van Waerbeke et al. 2006). At this level of precision, the SLC will be a major source of systematics, and the only way to tackle this issue is to have photometric redshifts for each individual galaxy, sources and lenses. This strengthens the requirement that future lensing surveys will have to cover a wide range of the optical spectrum, from *U* band to near-infrared, with narrow-band filters, similar to the COMBO-17 approach (Heymans et al. 2004).

The second numerical experiment measured the lensed to unlensed galaxy counts ratio. The value of this ratio was obtained for various angular sizes of observational fields. The general trend of the results in the simulation can be understood using equation (17), where only magnitude changes played an important role in enhancing the counts in our simulations. We learnt from this experiment that in order to have a realistic treatment of the magnification effects over the change of galaxy positions in the sky we need to go up in angular resolution.

Among the ideas that remain to be investigated using the present configuration of LEMOMAF are those that take advantage of mock images at multiple wavelengths to identify the best strategies for measuring the shear, as well as those which intend to study the bias introduced by intrinsic alignments or other systematic effects on this measurement. We also plan to use LEMOMAF to help design future lensing surveys, which will need to employ a tool taylored to tackle non-linear effects such as SLC.

Future prospects with LEMOMAF include the simulation of galaxy–galaxy lensing and cosmic magnification. This kind of signal demands a resolution in our simulations of the order of 0.1 arcmin. With the *N*-body simulation we used in this paper and FFT methods, we barely achieve a resolution of $\sim$1–2 arcmin. In order to reach higher resolutions, we do not plan to only rely on more resolved *N*-body simulations but also to switch ray-tracing strategy and look in the direction of smooth particle lensing (Aubert, Amara & Metcalf 2007), as we have gained confidence from recent studies with GALICS (Blaizot et al. 2006) that the small-scale distribution of galaxies produced from an adequate-resolution *N*-body simulation and a new positioning scheme of galaxies inside the haloes can be accurate enough to attain this goal.

ACKNOWLEDGMENTS

JEF-R thanks the Ecole Normale Supérieure for financial support at the time of writing the code LEMOMAF through its 'Selection Internationale' project, and Karim Benabed for his much appreciated guidance during the early stages of this project. We thank Brice Ménard for his comments on some of the technical aspects presented in this paper. The *N*-body simulation used in this work was run on the Cray T3E at the IDRIS supercomputing facility. This work was performed in the framework of the HORIZON project.

REFERENCES

Aubert D., Amara A., Metcalf R. B., 2007, MNRAS, 376, 113
Bacon D. J., Refregier A. R., Ellis R. S., 2000, MNRAS, 318, 625
Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15
Benabed K., Van Waerbeke L., 2004, Phys. Rev. D, 70, 123515
Bernardeau F., Van Waerbeke L., Mellier Y., 1997, A&A, 322, 1
Bernardeau F., 1998, A&A, 338, 375
Bernstein G. M., Jain B., 2004, ApJ, 600, 17
Blaizot J., Guiderdoni B., Devriendt J. E. G., Bouchet F. R., Hatton S. J., Stoehr F., 2004, MNRAS, 352, 571
Blaizot J. et al., 2006, MNRAS, 369, 1009
Blaizot J., Wadadekar Y., Guiderdoni B., Colombi S. T., Bertin E., Bouchet F. R., Devriendt J. E. G., Hatton S., 2005, MNRAS, 360, 159
Broadhurst T. J., Taylor A. N., Peacock J. A., 1995, ApJ, 438, 49
Brown M. L., Taylor A. N., Bacon D. J., Gray M. E., Dye S., Meisenheimer K., Wolf C., 2003, MNRAS, 341, 100
Croton D. J., Gao L., White S. D. M., 2007, MNRAS, 374, 1303
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
Devriendt J. E. G., Guiderdoni B., Sadat R., 1999, A&A, 350, 381
Djorgovski S. et al., 1995, ApJ, 438, L13
Eke V. R., Cole S., Frenk C. S., 1996, MNRAS, 282, 263
Erben T., Van Waerbeke L., Bertin E., Mellier Y., Schneider P., 2001, A&A, 366, 717
Gardner J. P., Sharples R. M., Carrasco B. E., Frenk C. S., 1996, MNRAS, 282, L1
Hamana T., Colombi S. T., Thion A., Devriendt J. E. G. T., Mellier Y., Bernardeau F., 2002, MNRAS, 330, 365
Hamana T., Mellier Y., 2001, MNRAS, 327, 169
Hatton S., Devriendt J. E. G., Ninin S., Bouchet F. R., Guiderdoni B., Vibert D., 2003, MNRAS, 343, 75
Heymans C., Brown M., Heavens A., Meisenheimer K., Taylor A., Wolf C., 2004, MNRAS, 347, 895
Heymans C. et al., 2006, MNRAS, 368, 1323
Heymans C., White M., Heavens A., Vale C., Van Waerbeke L., 2006, MNRAS, 371, 750
Hirata C. M., Seljak U., 2004, Phys. Rev. D, 70, 6
Hoekstra H., Yee H. K. C., Gladders M. D., Barrientos L. F., Hall P. B., Infante L., 2002, ApJ, 572, 55
Hu W., 1999, ApJ, 522, L21
Hu W., 2002, Phys. Rev. D, 65, 023003
Huterer D., 2002, Phys. Rev. D, 65, 063001
Ishak M., Hirata C. M., 2005, Phys. Rev. D, 71, 2
Jain B., Seljak U., 1997, ApJ, 484, 560
Jain B., 2002, ApJ, 580, L3
Jain B., Seljak U., White S., 2000, ApJ, 530, 547
Jarvis M., Bernstein G. M., Fischer P., Smith D., Jain B., Tyson J. A., Wittman D., 2003, AJ, 125, 1014
Kaiser N., 2000, ApJ, 537, 555
Kaiser N., Wilson G., Luppino G. A., 2000, preprint (astro-ph/0003338)
Mandelbaum R., Hirata C. M., Ishak M., Seljak U., Brinkmann J., 2006, MNRAS, 367, 611
Maoli R., Van Waerbeke L., Mellier Y., Schneider P., Jain B., Bernardeau F., Erben T., Fort B., 2001, A&A, 368, 766
Moustakas L. A., Davis M., Graham J. R., Silk J., Peterson B. A., Yoshii Y., 1997, ApJ, 475, 445
Ninin S., 1999, PhD thesis, Université Paris 11
Premadi P., Martel H., Matzner R., 1998, ApJ, 493, 10
Scranton R. et al., 2005, ApJ, 633, 589

Spergel D. N. et al., 2007, ApJS, 170, 377
Tereno I., Dore D., Van Waerbeke L., Mellier Y., 2005, A&A, 429, 383
Van Waerbeke L. et al., 2000, A&A, 358, 30
Van Waerbeke L., Hamana T., Scoccimarro R., Colombi S., Bernardeau F., 2001, MNRAS, 322, 918
Van Waerbeke L., Mellier Y., Pelló R., Pen U. L., McCracken H. J., Jain B., 2002, A&A, 393, 369
Van Waerbeke L., White M., Hoekstra H., Heymans C., 2006, ApJ, 26, 91
Vale C., Hoekstra H., Van Waerbeke L., White M., 2004, ApJ, 613, L1
Vale C., White M., 2003, ApJ, 592, 699
Wittman D. M., Tyson J. A., Kirkman D., Dell'Antonio I., Bernstein G., 2000, Nat, 405, 143
Yasuda N. et al., 2001, AJ, 122, 1104
Zhang P. J., Pen U. L., 2005, Phys. Rev. Lett., 95, 241302

APPENDIX A: DARK MATTER

The cosmological *N*-body simulation (Ninin 1999) we use throughout this paper assumes a flat cold dark matter cosmology with a cosmological constant ($\Omega_{\rm m} = 1/3$, $\Omega_\Lambda = 2/3$), and a Hubble parameter $h = H_0/[100\ {\rm km\ s^{-1}\ Mpc^{-1}}] = 0.667$. The initial power spectrum was taken to be a scale-free ($n_{\rm s} = 1$) power spectrum evolved as predicted by Bardeen et al. (1986) and normalized to the present-day abundance of rich clusters with $\sigma_8 = 0.88$ (Eke, Cole & Frenk 1996). The simulated volume is a cubic box of side $L_b = 100\,h^{-1}$ Mpc, which contains 256^3 particles, resulting in a particle mass $m_{\rm p} = 8.272 \times 10^9\,{\rm M}_\odot$ and a smoothing length of 29.29 kpc. The density field was evolved from $z = 35.59$ to present day, and we output about 100 snapshots spaced logarithmically with the expansion factor.

In each snapshot, we identify haloes using a friend-of-friend (FOF) algorithm (Davis et al. 1985) with a linking length parameter $b = 0.2$, only keeping groups with more than 20 particles. At this point, we define the mass $M_{\rm fof}$ of the group as the sum of the masses of the linked particles, and the radius $R_{\rm fof}$ as the maximum distance of a constituent particle to the centre of mass of the group. We then fit a triaxial ellipsoid to each halo, and check that the virial theorem is satisfied within this ellipsoid. If not, we decrement its volume until we reach an inner virialized region. From the volume of this largest ellipsoidal virialized region, we define the virial radius $R_{\rm vir}$ and mass $M_{\rm vir}$. These virial quantities are the ones we use to compute the cooling of the hot baryonic component. Once all the haloes are identified and characterized, we build their merger history trees following all the constituent particles from snapshot to snapshot.

APPENDIX B: LIGHTING UP HALOES

The fate of baryons within the halo merger trees found above is decided according to a series of prescriptions which are either theoretically or phenomenologically motivated. The guideline – which is similar to other SAMs – is the following. Gas is shock-heated to the virial temperature when captured in a halo's potential well. It can then radiatively cool on to a rotationally supported disc, at the centre of the halo. Cold gas is turned into stars at a rate which depends on the dynamical properties of the disc. Stars then evolve, releasing both metals and energy into the interstellar medium (ISM), and in some cases blowing part of the ISM away back into the halo's hot phase. When haloes merge, the galaxies they harbour are gathered into the same potential well, and they may in turn merge together, due either to fortuitous collisions or to dynamical friction. When two galaxies merge, a 'new' galaxy is formed, the morphological and dynamical properties of which depend on those of its progenitors. Typically, a merger between equal mass galaxies will give birth to an ellipsoidal galaxy, whereas a merger of a massive galaxy with a small galaxy will mainly contribute to developing the massive galaxy's bulge component. The Hubble sequence then naturally appears as the result of the interplay between cooling – which develops discs – and merging and disc gravitational instabilities – which develop bulges.

Keeping track of the stellar content of each galaxy, as a function of age and metallicity, and knowing the galaxy's gas content and chemical composition, one can compute the (possibly extincted) spectral energy distribution (SED) of each galaxy. To this end, we use the STARDUST model (Devriendt, Guiderdoni & Sadat 1999) which predicts the SED of an obscured stellar population from the ultraviolet to the submillimetre.

This paper has been typeset from a TEX/LATEX file prepared by the author.

Bibliography

[AcA06] I. Aretxaga, S. consortium, and AzTEC team. The Star Formation History of SHADES Sources. In *Bulletin of the American Astronomical Society*, volume 38 of *Bulletin of the American Astronomical Society*, pages 1008–+, December 2006.

[AHC+07] I. Aretxaga, D. H. Hughes, K. Coppin, A. M. J. Mortier, J. Wagg, J. S. Dunlop, E. L. Chapin, S. A. Eales, E. Gaztañaga, M. Halpern, R. J. Ivison, E. van Kampen, D. Scott, S. Serjeant, I. Smail, T. Babbedge, A. J. Benson, S. Chapman, D. L. Clements, L. Dunne, S. Dye, D. Farrah, M. J. Jarvis, R. G. Mann, A. Pope, R. Priddey, S. Rawlings, M. Seigar, L. Silva, C. Simpson, and M. Vaccari. The SCUBA Half Degree Extragalactic Survey - IV. Radio-mm-FIR photometric redshifts. *MNRAS*, 379:1571–1588, August 2007.

[ATS+96] R. G. Abraham, N. R. Tanvir, B. X. Santiago, R. S. Ellis, K. Glazebrook, and S. van den Bergh. Galaxy morphology to I=25 mag in the Hubble Deep Field. *MNRAS*, 279:L47–L52, April 1996.

[AZZ+05] K. Abazajian, Z. Zheng, I. Zehavi, D. H. Weinberg, J. A. Frieman, A. A. Berlind, M. R. Blanton, N. A. Bahcall, J. Brinkmann, D. P. Schneider, and M. Tegmark. Cosmology and the Halo Occupation Distribution from Small-Scale Galaxy Clustering in the Sloan Digital Sky Survey. *ApJ*, 625:613–620, June 2005.

[Bau06] C. M. Baugh. A primer on hierarchical galaxy formation: the semi-analytical approach. *Reports of Progress in Physics*, 69:3101–3156, December 2006.

[BBM+06] R. G. Bower, A. J. Benson, R. Malbon, J. C. Helly, C. S. Frenk, C. M. Baugh, S. Cole, and C. G. Lacey. Breaking the hierarchy of galaxy formation. *MNRAS*, 370:645–655, August 2006.

[BC03] G. Bruzual and S. Charlot. Stellar population synthesis at the resolution of 2003. *MNRAS*, 344:1000–1028, October 2003.

[BD03] Y. Birnboim and A. Dekel. Virial shocks in galactic haloes?, October 2003.

[BEH+05] M. R. Blanton, D. Eisenstein, D. W. Hogg, D. J. Schlegel, and J. Brinkmann. Relationship between Environment and the Broadband Optical Properties of Galaxies in the Sloan Digital Sky Survey. *ApJ*, 629:143–157, August 2005.

[BEM+04] M. Balogh, V. Eke, C. Miller, I. Lewis, R. Bower, W. Couch, R. Nichol, J. Bland-Hawthorn, I. K. Baldry, C. Baugh, T. Bridges, R. Cannon, S. Cole, M. Colless, C. Collins, N. Cross, G. Dalton, R. de Propris, S. P. Driver,

G. Efstathiou, R. S. Ellis, C. S. Frenk, K. Glazebrook, P. Gomez, A. Gray, E. Hawkins, C. Jackson, O. Lahav, S. Lumsden, S. Maddox, D. Madgwick, P. Norberg, J. A. Peacock, W. Percival, B. A. Peterson, W. Sutherland, and K. Taylor. Galaxy ecology: groups and low-density environments in the SDSS and 2dFGRS. *MNRAS*, 348:1355–1372, March 2004.

[Ber98] F. Bernardeau. The effects of source clustering on weak lensing statistics. *A&A*, 338:375–382, October 1998.

[BGB+04] I. K. Baldry, K. Glazebrook, J. Brinkmann, Ž. Ivezić, R. H. Lupton, R. C. Nichol, and A. S. Szalay. Quantifying the Bimodal Color-Magnitude Distribution of Galaxies. *ApJ*, 600:681–694, January 2004.

[BGD+04] J. Blaizot, B. Guiderdoni, J. E. G. Devriendt, F. R. Bouchet, S. J. Hatton, and F. Stoehr. GALICS- III. Properties of Lyman-break galaxies at a redshift of 3. *MNRAS*, 352:571–588, August 2004.

[BH92] J. E. Barnes and L. Hernquist. Dynamics of interacting galaxies. *ARAA*, 30:705–742, 1992.

[Bin77] J. Binney. The physics of dissipational galaxy formation. *ApJ*, 215:483–491, July 1977.

[BKH05] A. J. Benson, M. Kamionkowski, and S. H. Hassani. Self-consistent theory of halo mergers. *MNRAS*, 357:847–858, March 2005.

[BMdC94] R. Buta, S. Mitra, G. de Vaucouleurs, and H. G. Corwin, Jr. Mean morphological types of bright galaxies. *AJ*, 107:118–134, January 1994.

[BMKW03] E. F. Bell, D. H. McIntosh, N. Katz, and M. D. Weinberg. A First Estimate of the Baryonic Mass Function of Galaxies. *ApJL*, 585:L117–L120, March 2003.

[BSC+06] J. Blaizot, I. Szapudi, S. Colombi, T. Budavàri, F. R. Bouchet, J. E. G. Devriendt, B. Guiderdoni, J. Pan, and A. Szalay. GALICS - V: Low- and high-order clustering in mock Sloan Digital Sky Surveys. *MNRAS*, 369:1009–1020, July 2006.

[BT87] J. Binney and S. Tremaine. *Galactic dynamics*. Princeton, NJ, Princeton University Press, 1987, 747 p., 1987.

[BWG+05] J. Blaizot, Y. Wadadekar, B. Guiderdoni, S. T. Colombi, E. Bertin, F. R. Bouchet, J. E. G. Devriendt, and S. Hatton. MoMaF: the Mock Map Facility. *MNRAS*, 360:159–175, June 2005.

[Car04] S. M. Carroll. Why is the Universe Accelerating? In R. E. Allen, D. V. Nanopoulos, and C. N. Pope, editors, *The New Cosmology: Conference on Strings and Cosmology*, volume 743 of *American Institute of Physics Conference Series*, pages 16–32, December 2004.

[CAV00] P. Colín, V. Avila-Reese, and O. Valenzuela. Substructure and Halo Density Profiles in a Warm Dark Matter Cosmology. *ApJ*, 542:622–630, October 2000.

[CBDG05] A. Cattaneo, J. Blaizot, J. Devriendt, and B. Guideroni. Active Galactic Nuclei In Cosmological Simulations - I. Formation of black holes and spheroids through mergers. *MNRAS*, 364:407–423, December 2005.

[CBIS03] S. C. Chapman, A. Blain, R. Ivison, and I. Smail. The Redshifts of SCUBA Galaxies: Implications for Spheroid Formation. In E. Perez, R. M. Gonzalez Delgado, and G. Tenorio-Tagle, editors, *Star Formation Through Time*, volume 297 of *Astronomical Society of the Pacific Conference Series*, pages 289–+, 2003.

[Cha43] S. Chandrasekhar. Dynamical Friction. I. General Considerations: the Coefficient of Dynamical Friction. *ApJ*, 97:255–+, March 1943.

[CJPS06] T. J. Cox, P. Jonsson, J. R. Primack, and R. S. Somerville. Feedback in simulations of disc-galaxy major mergers. *MNRAS*, 373:1013–1038, December 2006.

[CSW+06] D. J. Croton, V. Springel, S. D. M. White, G. De Lucia, C. S. Frenk, L. Gao, A. Jenkins, G. Kauffmann, J. F. Navarro, and N. Yoshida. The many lives of active galactic nuclei: cooling flows, black holes and the luminosities and colours of galaxies. *MNRAS*, 365:11–28, January 2006.

[DB06] A. Dekel and Y. Birnboim. Galaxy bimodality due to cold flows and shock heating. *MNRAS*, 368:2–20, May 2006.

[DCC+07] P. Di Matteo, F. Combes, I. Chilingarian, A. . Melchior, and B. Semelin. Chemodynamical evolution of interacting galaxies: the GalMer view. *ArXiv e-prints*, 709, September 2007.

[dCMS07] P. di Matteo, F. Combes, A.-L. Melchior, and B. Semelin. Star formation efficiency in galaxy interactions and mergers: a statistical study. *A&A*, 468:61–81, June 2007.

[DD87] S. Djorgovski and M. Davis. Fundamental properties of elliptical galaxies. *ApJ*, 313:59–68, February 1987.

[DEFW85] M. Davis, G. Efstathiou, C. S. Frenk, and S. D. M. White. The evolution of large-scale structure in a universe dominated by cold dark matter. *ApJ*, 292:371–394, May 1985.

[DG00] J. E. G. Devriendt and B. Guiderdoni. Galaxy modelling. II. Multi-wavelength faint counts from a semi-analytic model of galaxy formation. *A&A*, 363:851–862, November 2000.

[DGS99] J. E. G. Devriendt, B. Guiderdoni, and R. Sadat. Galaxy modelling. I. Spectral energy distributions from far-UV to sub-mm wavelengths. *A&A*, 350:381–398, October 1999.

[DH02] D. A. Dale and G. Helou. The Infrared Spectral Energy Distribution of Normal Star-forming Galaxies: Calibration at Far-Infrared and Submillimeter Wavelengths. *ApJ*, 576:159–168, September 2002.

[dL00] R. S. de Jong and C. Lacey. The Local Space Density of SB-SDM Galaxies as Function of Their Scale Size, Surface Brightness, and Luminosity. *ApJ*, 545:781–797, December 2000.

[DLB+87] A. Dressler, D. Lynden-Bell, D. Burstein, R. L. Davies, S. M. Faber, R. Terlevich, and G. Wegner. Spectroscopy and photometry of elliptical galaxies. I - A new distance estimator. *ApJ*, 313:42–58, February 1987.

[Dre80] A. Dressler. Galaxy morphology in rich clusters - Implications for the formation and evolution of galaxies. *ApJ*, 236:351–365, March 1980.

[DS86] A. Dekel and J. Silk. The origin of dwarf galaxies, cold dark matter, and biased galaxy formation. *ApJ*, 303:39–55, April 1986.

[DYB04] J. J. Dalcanton, P. Yoachim, and R. A. Bernstein. The Formation of Dust Lanes: Implications for Galaxy Evolution. *ApJ*, 608:189–207, June 2004.

[ELDM05] D. Elbaz, E. Le Floc'h, H. Dole, and D. Marcillac. Observational evidence for the presence of PAHs in distant Luminous Infrared Galaxies using ISO and Spitzer. *A&A*, 434:L1–L4, April 2005.

[FE80] S. M. Fall and G. Efstathiou. Formation and rotation of disc galaxies with haloes. *MNRAS*, 193:189–206, October 1980.

[Fer96] G. J. Ferland. *Hazy, A Brief Introduction to Cloudy 90.* University of Kentucky Internal Report, 565 pages, 1996.

[FRP+99] S. Folkes, S. Ronen, I. Price, O. Lahav, M. Colless, S. Maddox, K. Deeley, K. Glazebrook, J. Bland-Hawthorn, R. Cannon, S. Cole, C. Collins, W. Couch,

S. P. Driver, G. Dalton, G. Efstathiou, R. S. Ellis, C. S. Frenk, N. Kaiser, I. Lewis, S. Lumsden, J. Peacock, B. A. Peterson, W. Sutherland, and K. Taylor. The 2dF Galaxy Redshift Survey: spectral types and luminosity functions. *MNRAS*, 308:459–472, September 1999.

[FWB+99] C. S. Frenk, S. D. M. White, P. Bode, J. R. Bond, G. L. Bryan, R. Cen, H. M. P. Couchman, A. E. Evrard, N. Gnedin, A. Jenkins, A. M. Khokhlov, A. Klypin, J. F. Navarro, M. L. Norman, J. P. Ostriker, J. M. Owen, F. R. Pearce, U.-L. Pen, M. Steinmetz, P. A. Thomas, J. V. Villumsen, J. W. Wadsley, M. S. Warren, G. Xu, and G. Yepes. The Santa Barbara Cluster Comparison Project: A Comparison of Cosmological Hydrodynamics Solutions. *ApJ*, 525:554–582, November 1999.

[GDS+04] G. L. Granato, G. De Zotti, L. Silva, A. Bressan, and L. Danese. A Physical Model for the Coevolution of QSOs and Their Spheroidal Hosts. *ApJ*, 600:580–594, January 2004.

[GHBM98] B. Guiderdoni, E. Hivon, F. R. Bouchet, and B. Maffei. Semi-analytic modelling of galaxy evolution in the IR/submm range. *MNRAS*, 295:877–898, April 1998.

[GJS+05] J. R. I. Gott, M. Jurić, D. Schlegel, F. Hoyle, M. Vogeley, M. Tegmark, N. Bahcall, and J. Brinkmann. A Map of the Universe. *ApJ*, 624:463–484, May 2005.

[GLS+00] G. L. Granato, C. G. Lacey, L. Silva, A. Bressan, C. M. Baugh, S. Cole, and C. S. Frenk. The Infrared Side of Galaxy Formation. I. The Local Universe in the Semianalytical Framework. *ApJ*, 542:710–730, October 2000.

[GMG+00] S. Ghigna, B. Moore, F. Governato, G. Lake, T. Quinn, and J. Stadel. Density Profiles and Substructure of Dark Matter Halos: Converging Results at Ultra-High Numerical Resolution. *ApJ*, 544:616–628, December 2000.

[GR87] B. Guiderdoni and B. Rocca-Volmerange. A model of spectrophotometric evolution for high-redshift galaxies. *A&A*, 186:1–2, November 1987.

[GSW05] L. Gao, V. Springel, and S. D. M. White. The age dependence of halo clustering. *MNRAS*, 363:L66–L70, October 2005.

[HBE+03] D. W. Hogg, M. R. Blanton, D. J. Eisenstein, J. E. Gunn, D. J. Schlegel, I. Zehavi, N. A. Bahcall, J. Brinkmann, I. Csabai, D. P. Schneider, D. H. Weinberg, and D. G. York. The Overdensities of Galaxy Environments as a Function of Luminosity and Color. *ApJL*, 585:L5–L9, March 2003.

[HCT+02] T. Hamana, S. T. Colombi, A. Thion, J. E. G. T. Devriendt, Y. Mellier, and F. Bernardeau. Source-lens clustering effects on the skewness of the lensing convergence. *MNRAS*, 330:365–377, February 2002.

[HDN+03] S. Hatton, J. E. G. Devriendt, S. Ninin, F. R. Bouchet, B. Guiderdoni, and D. Vibert. GALICS- I. A hybrid N-body/semi-analytic model of hierarchical galaxy formation. *MNRAS*, 343:75–106, July 2003.

[HLF+07] K. Heitmann, Z. Lukic, P. Fasel, S. Habib, M. S. Warren, M. White, J. Ahrens, L. Ankeny, R. Armstrong, B. O'Shea, P. M. Ricker, V. Springel, J. Stadel, and H. Trac. The Cosmic Code Comparison Project. *ArXiv e-prints*, 706, June 2007.

[HRH00] A. Heavens, A. Refregier, and C. Heymans. Intrinsic correlation of galaxy shapes: implications for weak lensing measurements. *MNRAS*, 319:649–656, December 2000.

[HVB+06] C. Heymans, L. Van Waerbeke, D. Bacon, J. Berge, G. Bernstein, E. Bertin, S. Bridle, M. L. Brown, D. Clowe, H. Dahle, T. Erben, M. Gray, M. Hetterscheidt, H. Hoekstra, P. Hudelot, M. Jarvis, K. Kuijken, V. Mar-

goniner, R. Massey, Y. Mellier, R. Nakajima, A. Refregier, J. Rhodes, T. Schrabback, and D. Wittman. The Shear Testing Programme - I. Weak lensing analysis of simulated ground-based observations. *MNRAS*, 368:1323–1339, May 2006.

[HYG07] M. Hoeft, G. Yepes, and S. Gottloeber. Too small to form a galaxy: How the UV background determines the baryon fraction. *ArXiv e-prints*, 708, August 2007.

[ICA+06] I. T. Iliev, B. Ciardi, M. A. Alvarez, A. Maselli, A. Ferrara, N. Y. Gnedin, G. Mellema, T. Nakamoto, M. L. Norman, A. O. Razoumov, E.-J. Rijkhorst, J. Ritzerveld, P. R. Shapiro, H. Susa, M. Umemura, and D. J. Whalen. Cosmological radiative transfer codes comparison project - I. The static density field tests. *MNRAS*, 371:1057–1086, September 2006.

[Jin02] Y. P. Jing. Intrinsic correlation of halo ellipticity and its implications for large-scale weak lensing surveys. *MNRAS*, 335:L89–L93, October 2002.

[KBW+04] A. V. Kravtsov, A. A. Berlind, R. H. Wechsler, A. A. Klypin, S. Gottlöber, B. Allgood, and J. R. Primack. The Dark Side of the Halo Occupation Distribution. *ApJ*, 609:35–49, July 2004.

[Ken89] R. C. Kennicutt, Jr. The star formation law in galactic disks. *ApJ*, 344:685–703, September 1989.

[Ken98] R. C. Kennicutt, Jr. The Global Schmidt Law in Star-forming Galaxies. *ApJ*, 498:541–+, May 1998.

[KGKK99] A. Klypin, S. Gottlöber, A. V. Kravtsov, and A. M. Khokhlov. Galaxies in N-Body Simulations: Overcoming the Overmerging Problem. *ApJ*, 516:530–551, May 1999.

[KHW+03] G. Kauffmann, T. M. Heckman, S. D. M. White, S. Charlot, C. Tremonti, E. W. Peng, M. Seibert, J. Brinkmann, R. C. Nichol, M. SubbaRao, and D. York. The dependence of star formation history and internal structure on stellar mass for 10^5 low-redshift galaxies. *MNRAS*, 341:54–69, May 2003.

[KKS+07] T. Kronberger, W. Kapferer, S. Schindler, A. Böhm, E. Kutdemir, and B. L. Ziegler. Internal Kinematics of Modelled Isolated and Interacting Disc Galaxies. In F. Combes and J. Palous, editors, *IAU Symposium*, volume 235 of *IAU Symposium*, pages 216–216, 2007.

[KKVP99] A. Klypin, A. V. Kravtsov, O. Valenzuela, and F. Prada. Where Are the Missing Galactic Satellites? *ApJ*, 522:82–92, September 1999.

[Kro02] P. Kroupa. The Initial Mass Function of Stars: Evidence for Uniformity in Variable Systems. *Science*, 295:82–91, January 2002.

[KS93] N. Kaiser and G. Squires. Mapping the dark matter with weak gravitational lensing. *ApJ*, 404:441–450, February 1993.

[KWG93] G. Kauffmann, S. D. M. White, and B. Guiderdoni. The Formation and Evolution of Galaxies Within Merging Dark Matter Haloes. *MNRAS*, 264:201–+, September 1993.

[KWH+04] G. Kauffmann, S. D. M. White, T. M. Heckman, B. Ménard, J. Brinchmann, S. Charlot, C. Tremonti, and J. Brinkmann. The environmental dependence of the relations between stellar mass, structure, star formation and nuclear activity in galaxies. *MNRAS*, 353:713–731, September 2004.

[LDP03] G. Lagache, H. Dole, and J.-L. Puget. Modelling infrared galaxy evolution using a phenomenological approach. *MNRAS*, 338:555–571, January 2003.

[LPD05] G. Lagache, J.-L. Puget, and H. Dole. Dusty Infrared Galaxies: Sources of the Cosmic Infrared Background. *ARAA*, 43:727–768, September 2005.

[LPS+04] C. Lonsdale, M. d. C. Polletta, J. Surace, D. Shupe, F. Fang, C. K. Xu, H. E. Smith, B. Siana, M. Rowan-Robinson, T. Babbedge, S. Oliver, F. Pozzi, P. Davoodi, F. Owen, D. Padgett, D. Frayer, T. Jarrett, F. Masci, J. O'Linger, T. Conrow, D. Farrah, G. Morrison, N. Gautier, A. Franceschini, S. Berta, I. Perez-Fournon, E. Hatziminaoglou, A. Afonso-Luis, H. Dole, G. Stacey, S. Serjeant, M. Pierre, M. Griffin, and R. Puetter. First Insights into the Spitzer Wide-Area Infrared Extragalactic Legacy Survey (SWIRE) Galaxy Populations. *ApJS*, 154:54–59, September 2004.

[LS01] T. Lejeune and D. Schaerer. Database of Geneva stellar evolution tracks and isochrones for $(UBV)_J$ $(RI)_C$ JHKLL'M, HST-WFPC2, Geneva and Washington photometric systems. *A&A*, 366:538–546, February 2001.

[LS06] G. Lemson and V. Springel. Cosmological Simulations in a Relational Database: Modelling and Storing Merger Trees. In C. Gabriel, C. Arviset, D. Ponz, and S. Enrique, editors, *Astronomical Data Analysis Software and Systems XV*, volume 351 of *Astronomical Society of the Pacific Conference Series*, pages 212–+, July 2006.

[MBB+07] H. Murakami, H. Baba, P. Barthel, D. L. Clements, M. Cohen, Y. Doi, K. Enya, E. Figueredo, N. Fujishiro, H. Fujiwara, M. Fujiwara, P. Garcia-Lario, T. Goto, S. Hasegawa, Y. Hibi, T. Hirao, N. Hiromoto, S. S. Hong, K. Imai, M. Ishigaki, M. Ishiguro, D. Ishihara, Y. Ita, W. . Jeong, K. S. Jeong, H. Kaneda, H. Kataza, M. Kawada, T. Kawai, A. Kawamura, M. F. Kessler, D. Kester, T. Kii, D. C. Kim, W. Kim, H. Kobayashi, B. C. Koo, S. M. Kwon, H. M. Lee, R. Lorente, S. Makiuti, H. Matsuhara, T. Matsumoto, H. Matsuo, S. Matsuura, T. G. Mueller, N. Murakami, H. Nagata, T. Nakagawa, T. Naoi, M. Narita, M. Noda, S. H. Oh, A. Ohnishi, Y. Ohyama, Y. Okada, H. Okuda, S. Oliver, T. Onaka, T. Ootsubo, S. Oyabu, S. Pak, Y. S. Park, C. P. Pearson, M. Rowan-Robinson, T. Saito, I. Sakon, A. Salama, S. Sato, R. S. Savage, S. Serjeant, H. Shibai, M. Shirahata, J. J. Sohn, T. Suzuki, T. Takagi, H. Takahashi, T. Tanabe, T. T. Takeuchi, S. Takita, M. Thomson, K. Uemizu, M. Ueno, F. Usui, E. Verdugo, T. Wada, L. Wang, T. Watabe, H. Watarai, G. J. White, I. Yamamura, C. Yamauchi, and A. Yasuda. The Infrared Astronomical Mission AKARI. *ArXiv e-prints*, 708, August 2007.

[MCC+03] D. S. Madgwick, A. L. Coil, C. J. Conselice, M. C. Cooper, M. Davis, R. S. Ellis, S. M. Faber, D. P. Finkbeiner, B. Gerke, P. Guhathakurta, N. Kaiser, D. C. Koo, J. A. Newman, A. C. Phillips, C. C. Steidel, B. J. Weiner, C. N. A. Willmer, and R. Yan. The DEEP2 Galaxy Redshift Survey: Spectral Classification of Galaxies at z˜1. *ApJ*, 599:997–1005, December 2003.

[MF05] P. Monaco and F. Fontanot. Feedback from quasars in star-forming galaxies and the triggering of massive galactic winds. *MNRAS*, 359:283–294, May 2005.

[MFD+96] P. Madau, H. C. Ferguson, M. E. Dickinson, M. Giavalisco, C. C. Steidel, and A. Fruchter. High-redshift galaxies in the Hubble Deep Field: colour selection and star formation history to z˜4. *MNRAS*, 283:1388–1404, December 1996.

[MGQ+98] B. Moore, F. Governato, T. Quinn, J. Stadel, and G. Lake. Resolving the Structure of Cold Dark Matter Halos. *ApJL*, 499:L5+, May 1998.

[MH94] J. C. Mihos and L. Hernquist. Triggering of starbursts in galaxies by minor mergers. *ApJL*, 425:L13–L16, April 1994.

[MH96] J. C. Mihos and L. Hernquist. Gasdynamics and Starbursts in Major Mergers. *ApJ*, 464:641–+, June 1996.

[MMP83] J. S. Mathis, P. G. Mezger, and N. Panagia. Interstellar radiation field

and dust temperatures in the diffuse interstellar matter and in giant molecular clouds. *A&A*, 128:212–229, November 1983.

[MMW98] H. J. Mo, S. Mao, and S. D. M. White. The formation of galactic discs. *MNRAS*, 295:319–336, April 1998.

[MQG+99] B. Moore, T. Quinn, F. Governato, J. Stadel, and G. Lake. Cold collapse and the core catastrophe. *MNRAS*, 310:1147–1152, December 1999.

[MR02] E. Moy and B. Rocca-Volmerange. Modelling ionized gas and stellar emissions in starburst components. *New Astronomy Review*, 46:203–205, May 2002.

[MSK+07] S. Matsuura, M. Shirahata, M. Kawada, Y. Doi, T. Nakagawa, H. Shibai, C. P. Pearson, T. Takagi, W.-S. Jeong, S. Oyabu, and H. Matsuhara. AKARI Far-Infrared Source Counts in the Lockman Hole. *ArXiv e-prints*, 707, June 2007.

[MTR+98] J. Magorrian, S. Tremaine, D. Richstone, R. Bender, G. Bower, A. Dressler, S. M. Faber, K. Gebhardt, R. Green, C. Grillmair, J. Kormendy, and T. Lauer. The Demography of Massive Dark Objects in Galaxy Centers. *AJ*, 115:2285–2305, June 1998.

[MVVH06] D. Munshi, P. Valageas, L. Van Waerbeke, and A. Heavens. Cosmology with Weak Lensing Surveys. *ArXiv Astrophysics e-prints*, December 2006.

[ND08] E. Neistein and A. Dekel. Constructing merger trees that mimic N-body simulations. *MNRAS*, 383:615–626, January 2008.

[NFW97] J. F. Navarro, C. S. Frenk, and S. D. M. White. A Universal Density Profile from Hierarchical Clustering. *ApJ*, 490:493–+, December 1997.

[NHv+84] G. Neugebauer, H. J. Habing, R. van Duinen, H. H. Aumann, B. Baud, C. A. Beichman, D. A. Beintema, N. Boggess, P. E. Clegg, T. de Jong, J. P. Emerson, T. N. Gautier, F. C. Gillett, S. Harris, M. G. Hauser, J. R. Houck, R. E. Jennings, F. J. Low, P. L. Marsden, G. Miley, F. M. Olnon, S. R. Pottasch, E. Raimond, M. Rowan-Robinson, B. T. Soifer, R. G. Walker, P. R. Wesselius, and E. Young. The Infrared Astronomical Satellite (IRAS) mission. *ApJL*, 278:L1–L6, March 1984.

[NLB+05] M. Nagashima, C. G. Lacey, C. M. Baugh, C. S. Frenk, and S. Cole. The metal enrichment of the intracluster medium in hierarchical galaxy formation models. *MNRAS*, 358:1247–1266, April 2005.

[NW94] J. F. Navarro and S. D. M. White. Simulations of dissipative galaxy formation in hierarchically clustering universes-2. Dynamics of the baryonic component in galactic haloes. *MNRAS*, 267:401–412, March 1994.

[ORA+00] S. Oliver, M. Rowan-Robinson, D. M. Alexander, O. Almaini, M. Balcells, A. C. Baker, X. Barcons, M. Barden, I. Bellas-Velidis, F. Cabrera-Guerra, R. Carballo, C. J. Cesarsky, P. Ciliegi, D. L. Clements, H. Crockett, L. Danese, A. Dapergolas, B. Drolias, N. Eaton, A. Efstathiou, E. Egami, D. Elbaz, D. Fadda, M. Fox, A. Franceschini, R. Genzel, P. Goldschmidt, M. Graham, J. I. Gonzalez-Serrano, E. A. Gonzalez-Solares, G. L. Granato, C. Gruppioni, U. Herbstmeier, P. Héraudeau, M. Joshi, E. Kontizas, M. Kontizas, J. K. Kotilainen, D. Kunze, F. La Franca, C. Lari, A. Lawrence, D. Lemke, M. J. D. Linden-Vørnle, R. G. Mann, I. Márquez, J. Masegosa, K. Mattila, R. G. McMahon, G. Miley, V. Missoulis, B. Mobasher, T. Morel, H. Nørgaard-Nielsen, A. Omont, P. Papadopoulos, I. Perez-Fournon, J.-L. Puget, D. Rigopoulou, B. Rocca-Volmerange, S. Serjeant, L. Silva, T. Sumner, C. Surace, P. Vaisanen, P. P. van der Werf, A. Verma, L. Vigroux, M. Villar-Martin, and C. J. Willott. The European Large Area ISO Survey - I. Goals, definition and observations. *MNRAS*, 316:749–767, August 2000.

[PCH07] H. Parkinson, S. Cole, and J. Helly. Generating Dark Matter Halo Merger Trees. *ArXiv e-prints*, 708, August 2007.

[Pee93] P. J. E. Peebles. *Principles of physical cosmology*. Princeton Series in Physics, Princeton, NJ: Princeton University Press, —c1993, 1993.

[Pra07] N. Prantzos. An Introduction to Galactic Chemical Evolution. *ArXiv e-prints*, 709, September 2007.

[PS74] W. H. Press and P. Schechter. Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. *ApJ*, 187:425–438, February 1974.

[Ref03] A. Refregier. Weak Gravitational Lensing by Large-Scale Structure. *ARAA*, 41:645–668, 2003.

[RO77] M. J. Rees and J. P. Ostriker. Cooling, dynamics and fragmentation of massive gas clouds - Clues to the masses and radii of galaxies and clusters. *MNRAS*, 179:541–559, June 1977.

[SBD+07] D. N. Spergel, R. Bean, O. Doré, M. R. Nolta, C. L. Bennett, J. Dunkley, G. Hinshaw, N. Jarosik, E. Komatsu, L. Page, H. V. Peiris, L. Verde, M. Halpern, R. S. Hill, A. Kogut, M. Limon, S. S. Meyer, N. Odegard, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Three-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Implications for Cosmology. *ApJS*, 170:377–408, June 2007.

[SBK+92] G. F. Smoot, C. L. Bennett, A. Kogut, E. L. Wright, J. Aymon, N. W. Boggess, E. S. Cheng, G. de Amici, S. Gulkis, M. G. Hauser, G. Hinshaw, P. D. Jackson, M. Janssen, E. Kaita, T. Kelsall, P. Keegstra, C. Lineweaver, K. Loewenstein, P. Lubin, J. Mather, S. S. Meyer, S. H. Moseley, T. Murdock, L. Rokke, R. F. Silverberg, L. Tenorio, R. Weiss, and D. T. Wilkinson. Structure in the COBE differential microwave radiometer first-year maps. *ApJL*, 396:L1–L5, September 1992.

[Sch76] P. Schechter. An analytic expression for the luminosity function for galaxies. *ApJ*, 203:297–306, January 1976.

[SD93] R. S. Sutherland and M. A. Dopita. Cooling functions for low-density astrophysical plasmas. *ApJS*, 88:253–327, September 1993.

[Sil77] J. Silk. On the fragmentation of cosmic gas clouds. I - The formation of galaxies and the first generation of stars. *ApJ*, 211:638–648, February 1977.

[Sil01] J. Silk. The formation of galaxy discs. *MNRAS*, 324:313–318, June 2001.

[SMT01] R. K. Sheth, H. J. Mo, and G. Tormen. Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes. *MNRAS*, 323:1–12, May 2001.

[SNH+84] B. T. Soifer, G. Neugebauer, G. Helou, C. J. Lonsdale, P. Hacking, W. Rice, J. R. Houck, F. J. Low, and M. Rowan-Robinson. The remarkable infrared galaxy ARP 220 = IC 4553. *ApJL*, 283:L1–L4, August 1984.

[Som02] R. S. Somerville. Can Photoionization Squelching Resolve the Substructure Crisis? *ApJL*, 572:L23–L26, June 2002.

[SP99] R. S. Somerville and J. R. Primack. Semi-analytic modelling of galaxy formation: the local Universe. *MNRAS*, 310:1087–1110, December 1999.

[Spr05] V. Springel. The cosmological simulation code GADGET-2. *MNRAS*, 364:1105–1134, December 2005.

[STWS06] C. Scannapieco, P. B. Tissera, S. D. M. White, and V. Springel. Feedback and metal enrichment in cosmological SPH simulations - II. A multiphase model with supernova energy feedback. *MNRAS*, 371:1125–1139, September 2006.

[SWJ+05] V. Springel, S. D. M. White, A. Jenkins, C. S. Frenk, N. Yoshida, L. Gao,

J. Navarro, R. Thacker, D. Croton, J. Helly, J. A. Peacock, S. Cole, P. Thomas, H. Couchman, A. Evrard, J. Colberg, and F. Pearce. Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature*, 435:629–636, June 2005.

[SWTK01] V. Springel, S. D. M. White, G. Tormen, and G. Kauffmann. Populating a cluster of galaxies - I. Results at [formmu2]z=0. *MNRAS*, 328:726–750, December 2001.

[TF77] R. B. Tully and J. R. Fisher. A new method of determining distances to galaxies. *A&A*, 54:661–673, February 1977.

[Tik07] A. V. Tikhonov. Voids in the SDSS galaxy survey. *Astronomy Letters*, 33:499–511, August 2007.

[Tin80] B. M. Tinsley. Evolution of the Stars and Gas in Galaxies. *Fundamentals of Cosmic Physics*, 5:287–388, 1980.

[TSTV02] R. B. Tully, R. S. Somerville, N. Trentham, and M. A. W. Verheijen. Squelched Galaxies and Dark Halos. *ApJ*, 569:573–581, April 2002.

[van98] F. C. van den Bosch. The Formation of Disk-Bulge-Halo Systems and the Origin of the Hubble Sequence. *ApJ*, 507:601–614, November 1998.

[VMR+01] L. Van Waerbeke, Y. Mellier, M. Radovich, E. Bertin, M. Dantel-Fort, H. J. McCracken, O. Le Fèvre, S. Foucaud, J.-C. Cuillandre, T. Erben, B. Jain, P. Schneider, F. Bernardeau, and B. Fort. Cosmic shear statistics and cosmology. *A&A*, 374:757–769, August 2001.

[WR78] S. D. M. White and M. J. Rees. Core condensation in heavy halos - A two-stage theory for galaxy formation and clustering. *MNRAS*, 183:341–358, May 1978.

[YAA+00] D. G. York, J. Adelman, J. E. Anderson, Jr., S. F. Anderson, J. Annis, N. A. Bahcall, J. A. Bakken, R. Barkhouser, S. Bastian, E. Berman, W. N. Boroski, S. Bracker, C. Briegel, J. W. Briggs, J. Brinkmann, R. Brunner, S. Burles, L. Carey, M. A. Carr, F. J. Castander, B. Chen, P. L. Colestock, A. J. Connolly, J. H. Crocker, I. Csabai, P. C. Czarapata, J. E. Davis, M. Doi, T. Dombeck, D. Eisenstein, N. Ellman, B. R. Elms, M. L. Evans, X. Fan, G. R. Federwitz, L. Fiscelli, S. Friedman, J. A. Frieman, M. Fukugita, B. Gillespie, J. E. Gunn, V. K. Gurbani, E. de Haas, M. Haldeman, F. H. Harris, J. Hayes, T. M. Heckman, G. S. Hennessy, R. B. Hindsley, S. Holm, D. J. Holmgren, C.-h. Huang, C. Hull, D. Husby, S.-I. Ichikawa, T. Ichikawa, Ž. Ivezić, S. Kent, R. S. J. Kim, E. Kinney, M. Klaene, A. N. Kleinman, S. Kleinman, G. R. Knapp, J. Korienek, R. G. Kron, P. Z. Kunszt, D. Q. Lamb, B. Lee, R. F. Leger, S. Limmongkol, C. Lindenmeyer, D. C. Long, C. Loomis, J. Loveday, R. Lucinio, R. H. Lupton, B. MacKinnon, E. J. Mannery, P. M. Mantsch, B. Margon, P. McGehee, T. A. McKay, A. Meiksin, A. Merelli, D. G. Monet, J. A. Munn, V. K. Narayanan, T. Nash, E. Neilsen, R. Neswold, H. J. Newberg, R. C. Nichol, T. Nicinski, M. Nonino, N. Okada, S. Okamura, J. P. Ostriker, R. Owen, A. G. Pauls, J. Peoples, R. L. Peterson, D. Petravick, J. R. Pier, A. Pope, R. Pordes, A. Prosapio, R. Rechenmacher, T. R. Quinn, G. T. Richards, M. W. Richmond, C. H. Rivetta, C. M. Rockosi, K. Ruthmansdorfer, D. Sandford, D. J. Schlegel, D. P. Schneider, M. Sekiguchi, G. Sergey, K. Shimasaku, W. A. Siegmund, S. Smee, J. A. Smith, S. Snedden, R. Stone, C. Stoughton, M. A. Strauss, C. Stubbs, M. SubbaRao, A. S. Szalay, I. Szapudi, G. P. Szokoly, A. R. Thakar, C. Tremonti, D. L. Tucker, A. Uomoto, D. Vanden Berk, M. S. Vogeley, P. Waddell, S.-i. Wang, M. Watanabe, D. H. Weinberg, B. Yanny, and N. Yasuda. The Sloan Digital Sky Survey: Technical

Summary. *AJ*, 120:1579–1587, September 2000.

[YMW03] R. Yan, D. S. Madgwick, and M. White. Constraining Evolution in the Halo Model Using Galaxy Redshift Surveys. *ApJ*, 598:848–857, December 2003.