University of Trieste

Faculty of Maths, Physics and Natural Sciences

Master degree in Space Physics and Astrophysics

Thesis of master degree

Fast generation of simulated catalogue of dark matter halos, for future cosmological survey

---

Graduand:
Calcioni Maria Carla

Supervisor:
Prof.
Monaco Pierluigi
Tutor:
Prof. Borgani Stefano

Accademic year 2011-2012

Da piccola mi portavi per mano sotto il "nostro" cielo,
crescendo mi hai aiutato a raggiungerlo.
A mio padre.

ABSTRACT

Il modello ΛCDM o Lambda-CDM è il modello su cui si basa attualmente la cosmologia moderna, esso rappresenta una versione del modello cosmologico del Big Bang per la quale l'universo conterrebbe una costante cosmologica, Λ, materia oscura fredda (Cold Dark Matter), barioni e radiazione. Il modello prevede un evento originario, il "Big Bang", o singolarità iniziale. Questo sarebbe stato seguito da un'accelerazione esponenziale, chiamata inflazione, che avrebbe generato fluttuazioni di densità primordiali. Le previsioni di tale modello sono state confermate dai dati provenienti da accurate misurazioni delle anisotropie nella radiazione cosmica di fondo (CMB) e dalle misurazioni delle fluttuazioni di densità di materia nella struttura su grande scala dell'Universo, quest'ultime ottenute da osservazioni della distribuzione di galassie, delle foreste Lyman-α, del weak lensing gravitazionale e dell'evoluzione della popolazione di ammassi di galassie. Le varie componenti dell'Universo (materia ed energia oscura, fotoni, barioni e neutrini) infatti imprimono sullo spettro delle fluttuazioni alcune impronte caratteristiche legate alle modalità con cui interagiscono e alla diversa storia termica, descritta dall' equazione di stato. Le oscillazioni acustiche nel fluido fotoni-barioni generate prima del disaccoppiamento dei fotoni, a redshift $z = 1088$ (circa 300.000 anni dopo il Big Bang), che sono state osservate sia nello spettro di potenza del CMB che nella distribuzione delle galassie.

Tuttavia questo grande numero di successi conferma ancora una volta la nostra scarsa comprensione dell'Universo. Le conoscenze attuali della cosmologia infatti ci dicono che l'Universo si è evoluto da uno stato omogeneo iniziale fino ad arrivare alle attuali strutture cosmologiche che oggi osserviamo nel cielo, grazie ad un processo di crescita gerarchico. Ad ogni modo questo scenario si basa su due assunzioni che riguardano le condizioni iniziali dell'universo, la natura della gravità stessa e delle componenti di energia e materia oscura. E' per questo motivo che uno degli obiettivi principali della cosmologia moderna, e delle future survey cosmologiche è quello di investigare su queste componenti "oscure". Se da una parte infatti, si è fiduciosi che gli acceleratori di particelle di nuova generazione, come il Large Hadron Collider (LHC), possano aiutare nello studio delle particelle candidate a contribuire alla materia oscura, dall'altra le proprietà dell'energia oscura possono essere vincolate solamente da osservazioni cosmologiche, come l'evoluzione delle strutture cosmiche, o la distanza di luminosità delle supernove di tipo Ia, grazie alla quale è stata scoperta l'energia oscura.

Grandi survey come l'anglo australiana 2dF galaxy redshift survey o la Sloan Digital Sky Survey (SDSS) hanno giocato un ruolo importante nel dare forma ai

vincoli del modello cosmologico. Sono state in grado infatti di dare delle precise stime sulla statistica standard utilizzata per quantificare la struttura a grande scala (funzione di correlazione e spettro di potenza) e hanno quantificato diverse proprietà riguardo la distribuzione di galassie. Le future survey cosmologiche come EUCLID, BOSS (Baryon Oscillation Spectroscopic Survey), VIPERS (VIMOS Public Extragalactic Redshift Survey) e DES (Dark Energy Survey) infieriranno vincoli ancora più stringenti sui parametri cosmologici. La loro strategia è quella di osservare aree di cielo sempre più vaste, ottenendo misure di alta accuratezza a bassi redshift, $z \sim 1$, dove la crescita delle perturbazioni in un universo piatto è rallentata dalla presenza dell'energia oscura. In questo modo la misura della crescita delle perturbazioni nell'intervallo di redshift da 0 a 1-2 permetterebbe di porre stretti vincoli all'equazione di stato dell'energia oscura e darebbe una prova di un'eventuale evoluzione oltre una semplice costante cosmologica. La futura survey EUCLID, ad esempio, coprirà un'area di cielo di 15,000 deg^2 e arriverà ad un livello di profondità tale da raggiungere circa $z \sim 1$ ed oltre. EUCLID è una missione di classe media, recentemente approvata dall'Agenzia Spaziale Europea (ESA) nell'ambito del programma Cosmic Vision 2015-2025, con l'obbiettivo principale di comprendere l'origine dell'espansione dell'Universo. Equipaggiato con due strumenti che lavorano nel visuale e nel vicino infrarosso, questo telescopio sarà in grado di misurare l'effetto della lente debole sulla forma di miliardi di galassie e di misurare con accuratezza il redshift di decine di milioni di galassie. Queste osservazioni, oltre a dare la possibilità di mappare la distribuzione di materia tramite il weak lensing e la distribuzione delle galassie tramite redshift fotometrici e spettroscopici, permetteranno di identificare oltre 60,000 ammassi di galassie. Questi sono gli oggetti gravitazionalmente legati più massivi dell'Universo e la distribuzione della loro popolazione conserva l'impronta della statistica delle fluttuazioni primordiali e della loro successiva crescita. Questo fa degli ammassi di galassie un ottimo strumento per poter porre vincoli ai parametri cosmologici.

Le incertezze sulle stime dei parametri fisici dagli osservabili è però largamente dominata da effetti sistematici legati alla sample variance e al bias con cui le galassie tracciano la loro massa. Un'accurata valutazione di questi errori sistematici richiede l'uso di simulazioni numeriche per poter generare distribuzioni non lineari di materia oscura e modelli per poter popolare i risultanti aloni di materia oscura con cataloghi di galassie. Anche assumendo che la struttura a larga scala possa essere accuratamente descritta dall'evoluzione di pura materia non collisionale e che la generazione di di galassie da aloni di materia oscura sia sotto controllo, i requisiti di una singola simulazione N-body necessari per generare un catalogo simulato del genere di EUCLID, sono piuttosto elevati. Tali simulazioni avrebbero bisogno di più di 10^{10} particelle e più di 100 outputs per generare il merger tree e cataloghi di coni luce. In questo caso le esigenze di hardware in termini di ram e disco fisso diventano onerose e creano maggiorni problemi rispetto a quelli legati ai tempi di calcolo necessari per far generare una singola simulazione. La questione si complica maggiormente quando si ha necessità di un elevatissimo numero di realizzazioni (maggiori di 1000), come

nel caso della stima della matrice di covarianza dello spettro di potenza delle galassie o per statistiche a più alti ordini di grandezza.

Queste problematiche hanno spinto lo sviluppo di approcci approssimati basati sulla Teoria delle Perturbazioni Lagrangiana, ovvero sulle soluzioni perturbative di un sistema di equazioni per il calcolo del displacement di elementi di massa dalle loro posizioni iniziali. Con l'LPT è possibile riprodurre accuratamente il campo di densità di materia su grande scala, ma occorre un approccio diverso per la ricostruzione degli aloni di materia oscura.

Circa 10 anni fa Monaco P. ed altri collaboratori hanno scritto un codice, PINOCCHIO, acronimo per PINpointing Orbit-Crossing Collapsed HIerarchical Objects, capace di generare, con limitate risorse computazionali, cataloghi di aloni di materia oscura di cui sono note le masse, le posizioni e le velocità, a partire dalla realizzazione di un campo di densità gaussiano distribuito su una griglia cubica. Le condizioni iniziali sono quelle utilizzate dalla maggior parte delle simulazioni e sfruttano l'approssimazione di Zel'dovich (primo ordine dell'LPT). L'algoritmo su cui si basa il codice di PINOCCHIO ha radici nell'approccio dell' Extended Press and Schechter e nella sua estenzione al collasso non sferico. PINOCCHIO è in grado di riprodurre la statistica degli aloni di materia oscura (funzione di massa e funzione di correlazione a due punti) con un'accuratezza del $\sim 10 - 20$ per cento, ma è anche in grado di generare aloni di materia oscura in accordo con quelli simulati anche al livello object-by-object. Il codice è stato aggiornato nel tempo. Attualmente la versione più recente è la terza che è stata ottimizzata per l'utilizzo di moderni computer massivamente paralleli, permettendo la generazione di grandi cataloghi di aloni di materia oscura con un tempo di calcolo dell'ordine di qualche frazione di un'ora.

L'obiettivo di questa tesi è sviluppare il codice di PINOCCHIO in due principali direzioni. Nella prima direzione siamo intervenuti con l'intenzione di soddisfarre una particolare esigenza del codice. La versione originale di PINOCCHIO calcolava le grandezze cosmologiche con delle soluzioni analitiche, andando così a considerare solo particolari casi in cui queste erano ammesse. Abbiamo quindi implementato nelle routine cosmologiche le soluzioni numeriche delle diverse grandezze focalizzandoci con maggiore attenzione al calcolo del tasso di crescita lineare della perturbazione al primo ordine e del tempo cosmico. Abbiamo esteso le routine aggiungendo le soluzioni numeriche del tasso di crescita lineare della perturbazione al secondo ordine dell'LPT e della distanza propria. Avendo analizzato l'accuratezza con cui le soluzioni numeriche riproducevano quelle analitiche, questo primo approccio al calcolo numerico, ha comunque contribuito ad una maggiore completezza del codice, garantendo soluzioni per diversi modelli cosmologici, inclusi quelli con un' equazione di stato dell'energia oscura diversa da quella standard. Nella seconda direzione siamo intervenuti nel codice per la produzione in outputs di cataloghi di coni-luce-passati di materia oscura, individuando l'instante in cui gli aloni li attraversano. Il criterio di selezione individua quegli aloni la cui luce emessa ha tempo di raggiungere l'osservatore. Successivamente abbiamo stimato l'incertezza sulla sample variance e sul rumore Poissoniano per una singola realizzazione del cono luce e

per diversi modelli cosmologici, mediante lo studio della statistica di number counts. Abbiamo generato diverse realizzazioni di uno stesso campione di volume dell'Universo di 15^3 Gpc^3 e di volume pari a quello che coprirà la missione EUCLID, cambiando le condizioni iniziali della simulazione. In questo modo PINOCCHIO potrà essere utilizzato per poter dare previsioni sulla distribuzione di aloni in funzione del redshift o per poter dare una stima sulle incertezze degli osservabili di cataloghi reali che possono essere prodotti da future survey cosmologiche come EUCLID.

Entrando più nel dettaglio nel primo capitolo esponiamo delle nozioni base di cosmologia sulla formazione ed evoluzione delle strutture cosmiche, discutendo brevemente del modello cosmologico standard e trattando la teoria lineare e non lineare delle perturbazioni in particolare per la componente di materia oscura. Nel secondo capitolo descriviamo il codice PINOCCHIO, l'algoritmo sui cui si basa, e le sue potenzialità legate alle limitate richieste computazionali, rispetto alle simulazioni N-body. Negli ultimi due capitoli, fulcro della tesi, si descrivono gli sviluppi del codice per la realizzazione di cataloghi di cono luce e i risultati delle stime delle incertezze legate alla sample variance e al rumore Poissoniano.

INTRODUCTION

ΛCDM is the concordance model of the modern cosmology. In this model the Universe is described as made of cold dark matter, baryons, radiation, neutrinos and a component of energy due to the so-called cosmological constant, described by the parameter Λ.

This model has been tested by means of accurate observations of the anisotropies in the cosmic microwave background radiation (hereafter CMB), as well as by means of the measure of the matter density fluctuations on the large scale structure of the Universe. Such fluctuations have been obtained from the distribution of galaxies, from the Lyman-α forest, from weak lensing and from the evolution of the galaxy cluster population.

The different components of the Universe, namely dark matter and energy, photons, baryons and neutrinos, leave their imprint on the power spectrum of the CMB and in the galaxy distribution through the baryonic acoustic oscillations (hereafter BAOs). BAOs are oscillations of the fluids made of photons and baryons, generated before the decoupling of baryons and photons, happened at a redshift $z = 1088$ (approximately 380,000 years after the Big Bang). Such imprint is due to the way these components interact among themselves, as well to the thermal history of the Universe, described by means of the equation of state of the different components.

The modern cosmological model describes an Universe evolved from an initial homogeneous state up to the present cosmic structures that we are today able to observe, through a hierarchical growth process. Such scenario relies upon some assumption concerning the initial conditions of the Universe generated by the inflation and the nature of the gravity itself, as well as the dark component of energy and matter. Understanding these dark components is the goal of modern cosmology, and is the target addressed by future surveys, as described below.

The new generation of particle accelerators, such as the Large Hadron Collider (LHC), can help to study the particles candidate to be the dark matter particles. On the other hand the properties of the dark energy can be constrained only through cosmological observations, such as the growth of cosmic structures or the luminosity distance of the the first place.

Large surveys, like the anglo-australian 2 degree Field Galaxy Redshift Survey and the Sloan Digital Sky Survey (SDSS hereafter), played a fundamental role in putting constraints on the cosmological models. These surveys provided precise estimates of the standard statistics used to characterize the large scale structure of the Universe, namely the correlation function and the power spectrum, as well as the properties of the galaxy distribution. Future surveys like EUCLID, BOSS

(Baryon Oscillation Spectroscopic Survey), VIPERS (VIMOS Public Extragalactic Redshift Survey) and DES (Dark Energy Survey) will put even tighter constraints on the cosmological parameters. In order to achieve such goals, these surveys will observe larger and larger areas of the sky, measuring with higher and higher precision the objects at low redshift, $z \sim 1$, where the growth of perturbations in a flat Universe should be slowed down by the presence of the dark energy. This kind of measures might put strong constraints on the equation of state of the dark energy, eventually probing an evolution of this component, pushing toward more complex models beyond the cosmological constant.

More in detail, Euclid is a middle class mission, recently approved by the European Space Agency (ESA) as part of the Cosmic Vision 2015-2025 program, with the goal of understanding the origin of the expansion of the Universe. It will cover an area of the sky of 15,000 deg^2, going as deep as $z \gtrsim 1$. Equipped with two instruments, working in the visual and near infrared wavelengths, it will be able to measure the weak lensing effect on the shape of billions of galaxies, precisely measuring the photometric and spectroscopic redshift of tens of thousands of galaxies. This way more than 60,000 galaxy clusters will be observed. Galaxy clusters are the most massive gravitationally bound objects in the Universe, and their distribution carries the information on the primordial fluctuations and their subsequent growth. This makes galaxy clusters an ideal tool to constrain cosmological parameters.

Despite the high precision that will be achieved in the future surveys, the reliability on the estimates of physical parameters from observables will be dominated by systematical effects due to the sample variance and the bias with which galaxies trace the gravitational potential. A precise study of these errors requires the use of numerical simulations in order to generate non linear distributions of dark matter, and models to populate the dark matter halos with galaxies. Even assuming that the large scale structure can be accurately described by the evolution of purely non collisional matter, and even assuming to have a good control on the formation of dark matter halos and their galaxies, the requirements for a N-body simulations to generate a catalogue comparable to what Euclid will observe are extremely high. In fact a simulation of this kind would require more than 10^{10} particles, with the outputs stored at more than 100 instants during the evolution of the simulated Universe in order to generate the merger trees and the light cone catalogues. The hardware requirements, such as RAM and hard disk memory, would be extremely high, being even more problematic of the speed of computation itself. The problems become even more problematic if a huge number (more than 1000) of Universe realizations must be generated, like in the case of the computation of the covariance matrix of the power spectrum of galaxies. These issues led to the development of approximated approaches based on the Lagrangian Perturbation Theory (LPT hereafter), that is the perturbative solutions of a system of equations for the computation of the displacement of mass elements from their initial positions. With the LPT it is possible to accurately reproduce the density field of matter on large scale, although it requires a different approach as far as the reconstruction of the dark matter halos is concerned. About 10 years ago, Monaco and

collaborators proposed the PINOCCHIO algorithm, acronyms for *PINpointing Orbit-Crossing Collapsed HIerarchical Objects*. This code is able to generate, with small computational resources, catalogues of dark matter halos, for which the masses, positions and velocities are known, starting from a gaussian density field distributed on a cubic grid. The initial conditions are those used by most of the numerical simulations and use the Zel'dovich approximation (first order of the LPT). The PINOCCHIO algorithm has its origin in the Extended Press and Schechter formalism and its extension to non-spherical collapse. PINOCCHIO is able to reproduce the mass function and the two points correlation function of dark matter halos, with a $\sim 10 - 20\%$ accuracy, as well as to generate dark matter halos in agreement with the simulated ones at a object-by-object level. The latest version of the code, the third one, has been optimized for the use on the modern massively parallel computers, generating large catalogues of dark matter halos in a computing time of the order of fractions of hour. The aim of this Thesis

is the development of the PINOCCHIO code following two main directions of development. The first direction aims at fixing particular limitations of the code. The original version of PINOCCHIO computed cosmological quantities using analytical solutions, therefore considering only particular cases in which these were known. Hence in the cosmological routines we have implemented numerical solutions of the different quantities, focusing in particular on the computation of the first order linear growth rate and on the cosmic time. We have also added the numerical solutions of the linear growth rate of the LPT second order perturbation and the proper distance. These modifications contribute to an improved generality of the code, now able to examine different cosmological models, including those with an equation of state for the dark energy different from the standard one.

The development in the second direction is the production past-light-cones catalogues of dark matter halos. The selection criterion consists in finding the halos whose light has enough time to reach the observer. Afterwards we have estimated the uncertainties on the sample variance and on the Poissonian noise for a single realization of the light cone and for different cosmological models, by using the number count statistics. We have generated many realizations, differing in the initial conditions, of the same volume of the Universe, 15^3 Gpc^3, corresponding to the same volume that will be covered by the Euclid mission. This way PINOCCHIO is able to provide forecasts for the distribution of halos as a function of redshift, as well as to provide estimates on the uncertainties of the observable quantities of the real catalogues that will be produced by future surveys like Euclid.

In the first chapter of this Thesis we will introduce some basics on the cosmological formation and evolution of cosmic structures. The standard cosmological model will be briefly discussed, as well as the linear and non linear theory of perturbation, focusing on the dark matter component. In the second chapter we will present the PINOCCHIO code, the algorithm and its advantages due to the limited required computational resources with respect to the numerical N-body simulations. In the last two chapters the developments of the code for the realization of light cone catalogues will be described, as well as the results

of the estimates and uncertainties dealing with sample variance and Poissonian noise.

CONTENTS

LIST OF TABLES

Table 1: Values of main cosmological quantities

Symbol	Name	Meaning	Value	Unite of measure
H_0	Hubble constant	Present speed of Universe expansion	70.5 ± 2.2	$\frac{km}{(s\,Mpc)}$
t_0	Time since Big Bang	Expansion age of Universe	13.742 ± 0.077	Gyr
t_{dec}	Age of Universe at decoupling	Age of Universe when CMBR was formed	$379^{+8}_{-7} \cdot 10^3$	yr
z_{dec}	Redshift of microwave background	Amount of Universe has grown since CMBR was formed	1089 ± 1	
t_r	Age of Universe at reionization	Likely age of Universe when first stars began to shine	$180^{+330}_{-30} \cdot 10^6$	yr
z_r	Redshift of reionization	Likely redshift of radiation from first stars	10 ± 1.1	
Ω_b	Baryon density parameter	Present density of baryons alone	0.0449 ± 0.0018	
Ω_m	Matter density parameter	Present density of matter alone	0.272 ± 0.019	
Ω_Λ	Dark energy density parameter	Present density of dark energy alone	0.7328 ± 0.019	
Ω_0	Cosmological density parameter	Present density of matter-energy divided by the amount required to make space flat	1.02 ± 0.02	
ρ_{cr}	Critical density	Density for which the Universe has null curvature	$2.7 \cdot 10^{11}$	$\frac{h^2 M_\odot}{Mpc^{-3}}$
σ_8	Sigma8	(see 1.2.2)	0.810 ± 0.017	
n	Spectral index	(see 1.2.2)	$0.951^{+0.015}_{-0.019}$	
A	Amplitude of power pectrum	(see 1.2.2)	$5.9 \pm 1.1 \cdot 10^5$	$\left(\frac{Mpc}{h}\right)^4$

Values of main cosmological quantities given bt WMAP9+spt+act data.

FORMATION AND EVOLUTION OF COSMIC STRUCTURES

The ΛCDM scenario assumes the gravitational instability paradigm: the cosmic structures in the Universe were formed as a consequence of the growth of primordial tiny fluctuations (for example seeded in the inflationary epoch) by gravitational instability in an expanding frame. The fluctuation or perturbation is characterized by its density contrast, with respect to the average density of the Universe.

The early universe, where very high densities and temperatures dominate, can be treated by using fluid thermodynamics. At origin the perturbations of the cosmic fluid have small amplitudes, with a density contrast relative to the background much smaller than unity: we are in the so called linear regime. This is well described by the Newtonian perturbation theory when structures have sizes much smaller than the horizon size, so that causality can be considered instantaneous. The relativistic theory of small perturbations is required when considering perturbations larger than the horizon size or when the matter content of the perturbations cannot be treated as a non-relativistic fluid. We will not treat this last one, because we will focus on the behavior of perturbations inside the horizon, in particular on the dark matter perturbations.

Nevertheless the linear regime is not enough to explain the formation of many objects in the present-day Universe, including galaxies and clusters of galaxies. In fact they have densities orders of magnitude higher than the average density of the Universe. These objects are thus in the highly nonlinear regime. Therefore to complete our description of structure formation in the Universe,we need to go beyond perturbation growth in the linear regime and address the gravitational collapse of overdensities in the nonlinear regime.

It's possible to find a deeper discussion of these arguments, in tex book of cosmology like Principles of Physical Cosmology by Peebles (1993), Cosmological Physics by Peacock (1999), Galaxy Formation and Evolution by Mo et al. (2010), Galactic dynamics by Binney and Tremaine (1987) and The Origin and evolution of cosmic structures by P.Coles (2002)).

In this chapter, in section §1.1 we briefly describe the cosmological standard model (ΛCDM), in the sections§1.2 and §1.3 we examine respectively the linear

regime and the non-linear regime for the component of Dark matter perturbations.

1.1 STANDARD COSMOLOGICAL MODEL

The ΛCDM is a version of the Big Bang cosmological model in which the universe contains a cosmological constant, denoted by Λ, cold dark matter, baryons and radition.

The model deals with a single originating event, the "Big Bang" or initial singularity. This was immediately followed by an exponential expansion of space, called inflation that generated fluctuations. The early universe remained hot for several hundred thousand years, a state that is detectable as a residual cosmic microwave background or CMB, a very low energy radiation emanating from all parts of the sky. The "Big Bang" scenario, with cosmic inflation, is the current cosmological model consistent with the observed continuing expansion of space, the observed distribution of lighter elements in the universe (hydrogen, helium, and lithium), and the spatial texture of minute irregularities (anisotropies) in the CMB radiation.

The model is based on the cosmological principle. It is the assertion that: on sufficiently large scale (beyond those traced by the large scale structure of the distribution of galaxies), the Universe is both homogeneous and isotropic. The Universe is clearly not exactly homogeneous, so cosmologists define homogeneity in an average sense: the Universe is taken to be identical in different places when one looks at sufficiently large pieces. This amounts to the strongly philosophical statement that the part of the Universe which we can see is a fair sample, and that the same physical laws apply throughout.

The assumptions of homogeneity and isotropy permit to define the comoving system: a "comoving" coordinate, the coordinate of observers moving along with the Hubble flow perceiving so the universe as isotropic; and a comoving time coordinate, the elapsed time since the Big Bang according to a clock of a comoving observer (it is the measure of cosmological time). Together, they form a complete coordinate system, giving us both the location and time of an event. The dependence of time of the mutual physical distance (proper distance) is given by the scale factor $a(t)$.

The model assumes General Relativity is the correct theory of gravity on cosmological scale, it uses the Einstein field equations (EFE), the Friedmann–Lemaître–Robertson–Walker metric (FLRW) metric, the Friedmann equations and the cosmological equations of state to describe the observable universe from right after the inflationary epoch to present and future time (qui).

So the model includes an expansion of metric space that is well documented both as the red shift of spectral absorption or emission lines in the light from distant galaxies and as the time dilation in the light decay of supernova luminosity curves.

The dark energy is the name that has been attributed to explain the force that causes accelerated expansion. Two proposed forms for dark energy are:
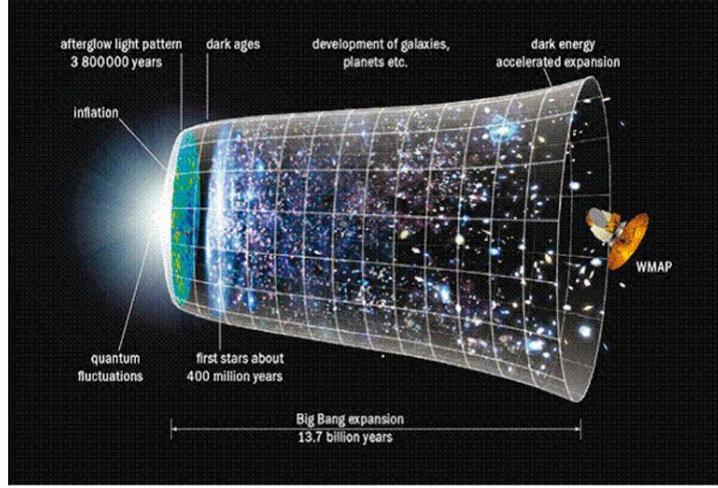
Figure 1: Time line of Universe (NASA/WMAP).

the cosmological constant,Λ, a constant term inserted by Albert Einstein in his relativity equations of fields interpreted as a constant energy density filling space homogeneously; scalar fields such as quintessence, dynamic quantities whose energy density can vary in time and space. Actually there is no observational evidence against a cosmological constant, for that reason the model is ΛCDM.

Another component of the energy density of the Universe is dark matter. Dark matter is described as being cold (i.e. its velocity is non-relativistic at the epoch of decoupling), non-baryonic (consisting of matter other than protons and neutrons), dissipationless (i.e. does not involve the dissipation of energy) and collisionless. It is introduced to explain the low anisotropy level of the CMB and permits to describe the structure of the universe on large scales and how galaxies form within that structure.

The remaining component of matter is the ordinary matter observed and almost 0.1 per cent of relic neutrinos.

Also, the radiation component is today negligible but, it was much more important in the distant past, dominating the matter at redshift > 3200.

1.1.1 *The EFE, the FLRW metric, the Friedmann equations and the cosmological equations of state*

The simplifying assumption that the universe is approximatively spatially homogeneous and isotropic, i.e. the Cosmological Principle, empirically justified approximatively on scales larger than ~ 100Mpc, implies that the metric of universe background must be of the form of Friedman Lamatre Robertson Walker (FLRW) metric:

$$ds^2 = c^2 dt^2 - a^2(t) \left[\frac{dr^2}{1 - Kr^2} + r^2(d\vartheta^2 + \sin^2\vartheta d\varphi^2) \right] \tag{1.1}$$

where c is the speed of light, (r, ϑ, φ) the comoving coordinates and t the proper time or cosmic time. The costant K, can assume values $0, 1, -1$, linked to the Universe geometry; respectively: for $K = 0$ the space is flat, for $K = 1$ the space

has positive curvature and for $K = -1$ the space has negative curvature. Finally the function $a(t)$ represent the cosmic scale factor on cosmic time t. This one contains the information of the expansion of Universe in the different epochs. This factor is linked to the redshift by the relation

$$a(t) = a_0/(1+z). \tag{1.2}$$

with $a_0 = 1$, while the cosmic time t is given by:

$$t(z) = \int_z^\infty \frac{dz}{(1+z')H(z')} = \int_\infty^{(1-1/a)} \frac{da}{H(a)a} \tag{1.3}$$

where H is the Hubble parameter. The EFE equate spacetime curvature (expressed by the Einstein tensor) are written in the form:

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu} R + g_{\mu\nu}\Lambda = \frac{8\pi G}{c^4}T_{\mu\nu} \tag{1.4}$$

where $R_{\mu\nu}$, is the Ricci curvature tensor, R, the scalar curvature, $g_{\mu\nu}$, the metric tensor, Λ, is the cosmological constant, G, is Newton's gravitational constant, c, the speed of light in vacuum, and $T_{\mu\nu}$, the stress–energy tensor.

Alexander Friedmann in 1922 got a set of equations obtained by solving the Einstein field equations, considering the energy tensor of a perfect fluid. The Friedmann equations describe therefore the expansion of space of the universe within the context of general relativity for which the structure of space time is determined on the distribution of the matter-energy inside it.

There are two independent Friedmann equations for modeling a homogeneous, isotropic universe. The first is given by:

$$H^2(t) = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2} \tag{1.5}$$

or, expliciting the different component of density by:

$$H^2(t) = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\left[\rho_{mo}\left(\frac{a_0}{a}\right)^3 + \rho_{ro}\left(\frac{a_0}{a}\right)^4 + \rho_{\Lambda0}\right] - \frac{Kc^2}{a^2} \tag{1.6}$$

which is derived from the oo component of Einstein's field equations and where the parameter ρ_i is the i-th component of the total density of matter-energy: m is the non-relativistic matter, r for relativistic matter, and Λ for the cosmological costant.

The second Friedmann equation is given by:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda c^2}{3} \tag{1.7}$$

or by:

$$\dot{H}(t) + H^2(t) = \frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right). \tag{1.8}$$

$H \equiv \frac{\dot{a}}{a}$ is the Hubble parameter, G, Λ, and c are universal constants (G is Newton's gravitational constant, Λ is the cosmological constant, c is the speed

of light in vacuum). $\frac{k}{a^2}$ is the spatial curvature in any time-slice of the universe. The subscript "o" near a letter indicates the quantity measured at the actual time t_o.

The Hubble parameter can change over time if other parts of the equation are time dependent (in particular the mass density, the vacuum energy, or the spatial curvature). Evaluating this parameter at the present time yields Hubble's constant which is the proportional constant of Hubble's law i.e. $v = H_0 D$ where v is the the recessional velocity, and D the obseerved distance of structure. For the parametration of the uncertainty on the experimental measure, it's used to express the quantities that depend of the value of H_o, as a function of $h \equiv H_o/(100 \mathrm{km\,s}^{-1}\mathrm{Mpc}^{-1})$.

It is used to express the Friedmann equations as a function of the parameter density of the different components of matter-energy, Ω_{io} given by:

$$\Omega_{io} \equiv \frac{\rho_{io}}{\rho_{cr}}, \qquad \qquad \mathrm{con}\ i = r, m, \Lambda. \qquad (1.9)$$

where ρ_{cr} is the critical density:

$$\rho_{cr} = 3H^2/8\pi G \qquad (1.10)$$

Therefore the first Friedmann equation can be recast in the form:

$$E^2(a) = \frac{H^2}{H_0^2} = \Omega_{ro}\left(\frac{a_o}{a}\right)^4 + \Omega_{mo}\left(\frac{a_o}{a}\right)^3 + \Omega_{ko}\left(\frac{a_o}{a}\right)^2 + \Omega_{DE}\left(\frac{a_o}{a}\right)^3 \exp\left(-3\int_1^a \frac{dP}{d\rho}\frac{da\prime}{a\prime}\right).$$
$$(1.11)$$

where Ω_r is the radiation paramenter, Ω_m is the matter (dark plus baryonic) parameter, $\Omega_k = 1 - \Omega$ is the "spatial curvature" parameter and Ω_{DE} is the dark energy paramenter. The value of the factor scale at $z = 0$, a_o is usually fixed to 1, normalizing each quantities with respect to present time. The integrand contains the equation of state of the dark energy, $\frac{dP}{d\rho}$, where P is the pressure and ρ is density. It is usually parameterized as:

$$\frac{dP}{d\rho} = w_0 + w_1(1 - a) \qquad (1.12)$$

where w_0 and w_1 are parameters, w_1 takes into account the dipendence of the time dependence of the equation. The case with $w_1 = 0$, $w_0 = -1$ correspond to assume as the dark energy component, the cosmological constant ($\Omega_{DE} = \Omega_\Lambda$). Considering the flat concordance model of ΛCDM, the equation (1.11) becames for the different epochs:

RADIATION DOMINATED ERA (RD). For $z \gg z_{eq}$ the radiation dominates, we have:

$$\frac{a}{a_o} = \left(\frac{32\pi G\rho_r}{3}\right)^{1/4} t^{1/2}. \qquad (1.13)$$

MATTER DOMINATED ERA (MD). After the equivalence ($z \ll z_{eq}$), the radiation doesn't influence very much the global dynamic, and we can neglect the $\Omega_{\Lambda 0} = 0$ contribution:

$$\left(\frac{\dot{a}}{a}\right)^2 = H_0^2 \Omega_{m0} \left(\frac{a_0}{a}\right)^3, \tag{1.14}$$

that has as solution:

$$\frac{a}{a_0} = \left(\frac{3}{2} H_0 t\right)^{2/3}. \tag{1.15}$$

This is the solution for the Einstein-de Sitter (EdS) model of Universe.

DARK ENERGY ERA (DED). Per $1 + z \lesssim \left(\frac{\Omega_{m0}}{\Omega_{\Lambda 0}}\right)^{-1/3}$ the cosmological constant is not negligible anymore and we have:

$$\left(\frac{\dot{a}}{a}\right)^2 = H_0^2 \left[\Omega_{m0} \left(\frac{a_0}{a}\right)^3 + \Omega_{\Lambda 0}\right]. \tag{1.16}$$

Because $0 < \Omega_{m0} < 1$ and $H_0 \equiv \dot{a}/a > 0$, we obtain the solution:

$$\frac{a}{a_0} = \left(\frac{\Omega_{m0}}{\Omega_{\Lambda 0}}\right)^{1/3} \left[\sinh\left(\frac{3}{2} \Omega_{\Lambda 0}^{1/2} H_0 t\right)\right]^{2/3}, \tag{1.17}$$

For early time the solution becomes the Eds solution, $a \propto t^{2/3}$, while for large t it approximates the solution for a De Sitter model, $a \propto \exp(\Omega_{\Lambda 0}^{1/2} H_0 t)$.

1.2 LINEAR THEORY OF PERTURBATIONS

The astrophysical objects we observe today are the result of the evolution of the primordial fluttuations. A key concept in the theory of gravitational instability is the Jeans length. Fluctuations on scales larger than the Jeans length are unstable against contraction under gravity, while fluctuations on smaller scales are supported against further collapse by pressure forces. Fluctuations on scales larger than the size of the horizon grow through self-gravity. As the universe grows older, the horizon expands and encompasses density fluctuations on progressively larger scales.

Fluctuations that enter in the horizon when the density of radiation is driving the expansion of the universe are effectively frozen until the matter and radiation densities become equal. This effect is called Meszaros effect. The radiation density effectively acts as a pressure that prevents the further collapse of any perturbation in the matter density (barion and dark matter). Due to the Meszaros effect, the spectrum of perturbations entering the horizon before the redshift of equivalence between radiation and matter z_{eq}, is distorted.

Perturbations which always stay outside the horizon don't produce distortion of the power spectrum.
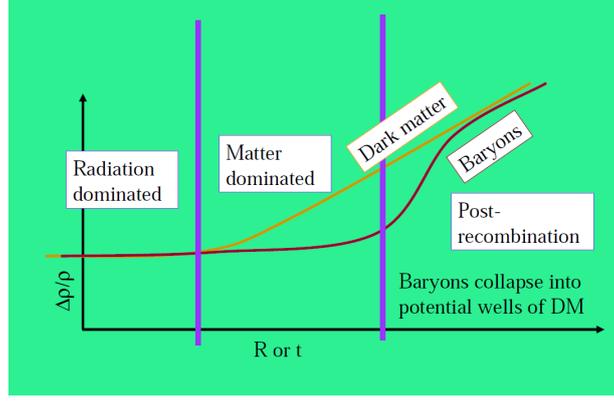
Figure 2: Evolution of the matter components as a function of time. The perturbations of dark matter start to grow after the redshift of equivalence, the barionic ones only after the redshift of decoupling (that is very closed to the redshift of recombination) falling later in the dark matter potential holes (http://www.merate.mi.astro.it).

After the redshift of equivalence, i.e. during the matter dominated era, the fluctuations made of collisionless matter, i.e. nonbaryonic dark matter, that entered in the horizon, can gravitationally grow, to collapse then later into a complex network of dark matter halos. While the baryonic or collisional matter, can grow only after the decoupling (when the Jeans mass in baryons falls by many orders of magnitude at a redshift of around $z \sim 1100$) finding the potential wells of the dark matter, it falls inside them (see figure 2).

1.2.1 *Newtonian perturbation theory: equation of evolution of perturbations for the CDM component*

In this section we describe the equation of the evolution of perturbations of the CDM component, i.e. the cold, non relativistic, collisionless, pressurless, component of the matter of the Universe, inside the horizon, until the amplitude of the perturbations are small ($\delta \ll 1$). The Newtonian approach considers the dark matter component as a fluid. This assumption is valid as long as the local velocity dispersion of the dark matter particles is sufficiently small respect the scale of interest (particle diffusion can be neglected on the scales of interest). We are under these conditions.

So consider the density ρ and velocity field **u** under the influence of a gravitational field with potential ϕ. The time evolution of the fluid is given by the equation of continuity (which describes mass conservation), the Euler equations (the equations of motion) and the Poisson equation (describing the gravitational field):

$$\frac{\partial \delta}{\partial t} + \frac{1}{a}\nabla \cdot [(1+\delta)\mathbf{v}] = 0 \quad \text{(continuity)}, \tag{1.18}$$

$$\frac{\partial \mathbf{v}}{\partial t} + \frac{\dot{a}}{a}\mathbf{v} + \frac{1}{a}(\mathbf{v} \cdot \nabla)\mathbf{v} = -\frac{\nabla \Phi}{a} \quad \text{(Euler)}, \tag{1.19}$$

$$\nabla^2 \Phi = 4\pi G \bar{\rho}_m a^2 \delta \quad \text{(Poisson)} \tag{1.20}$$

where $\nabla \equiv \nabla_{\mathbf{x}}$ and $\mathbf{v} = a\dot{\mathbf{x}}$, with $\mathbf{x} = \mathbf{r}/a(t)$ comoving coordinates. The equations are expressed as a function of the density contrast δ, defined as:

$$\delta(\mathbf{x}, t) = \frac{\rho(\mathbf{x}, t) - \bar{\rho}(t)}{\bar{\rho}(t)} \qquad (1.21)$$

with $\bar{\rho}$ the average density of the Universe*. Combining them, in the hypotesis of small perturbations ($\delta \ll 1$), it is possible to linearize the equations by neglecting all the second order terms in δ and $\mathbf{v}$. In this case, we obtain:

$$\frac{\partial^2 \delta}{\partial t^2} + 2\frac{\dot{a}}{a}\frac{\partial \delta}{\partial t} = 4\pi G \bar{\rho}_m \delta \qquad (1.22)$$

The term on the right-hand side is the gravitational term, which causes perturbations to grow via gravitational instability.

In the linear regime, the equations governing the evolution of the perturbations of dark matter, are all linear in perturbation quantities. It is then useful to expand the perturbation fields in some suitably chosen mode functions. If the curvature of the Universe can be neglected, as is the case when the Universe is flat or when the scales of interest are much smaller than the horizon size, the mode functions can be chosen to be plane waves and the perturbation fields can be represented by their Fourier transforms. For example for the density field δ, we have:

$$\delta(\mathbf{x}, t) = \sum_{\mathbf{k}} \delta_{\mathbf{k}}(t) \exp(i\mathbf{k} \cdot \mathbf{x}); \qquad (1.23)$$

where V_u is a portion of space large enough within which the fluctuations are assumed periodic that in the limit of $V \to \infty$ gives:

$$\delta(x, t) = V \int \frac{d^3k}{(2\pi)^3} \delta_k(t) exp(ik \cdot x). \qquad (1.24)$$

If we calculate the Fourier trasformation of the equation (1.22), we obtain the equation of the evolution of each of the different modes $\mathbf{k}$; where ∇ is replaced by $i\mathbf{k}$, and ∇^2 by $-k^2$, the equation of evolutionof dark matter becomes:

$$\frac{d^2\delta_{\mathbf{k}}}{dt^2} + 2\frac{\dot{a}}{a}\frac{d\delta_{\mathbf{k}}}{dt} = 4\pi G\bar{\rho}_m \delta_{\mathbf{k}} \qquad (1.25)$$

We will mainly focus on the evolution of this component, that is particulary important for this work.

We will find that some modes are amplified during the linear evolution while others are damped. The evolution therefore acts as a filter of the primordial density perturbations generated at some time in the early Universe.

We can demostrate that if $\delta_1(t)$ and $\delta_2(t)$ are two solutions of the differential equation 1.25 then:

$$\delta_2\dot{\delta_1} - \delta_1\dot{\delta_2} \propto a^{-2}, \qquad (1.26)$$

---

* The presence of a relativistic background has effect only on the variation of the expanding Universe rate, i.e.$a(t)$

Thus, if one solution of equation (1.25) is known, the other one can be obtained by solving this first-order differential equation. To solve (1.25), we recall that the Hubble constant, obeys:

$$\frac{dH}{dt} + H^2 = -\frac{4\pi G}{3}(\bar{\rho}_m + 2\rho_\Lambda),$$ (1.27)

given by the second Friedmann equation (1.8); because $\bar{\rho}_m \propto a^{-3}$ and ρ_Λ are costant, when we derive with respect to t the relation above we obtain:

$$\frac{d^2H}{dt^2} + 2\frac{\dot{a}}{a}\frac{dH}{dt} = 4\pi G\bar{\rho}_m H,$$ (1.28)

i.e. both H(t) that $\delta_\mathbf{k}(t)$ obbey to the same equation. Since H(t) decreases with time (see equation (1.6)),

$$\delta_- \propto H(t)$$ (1.29)

represent the decreasing solution. If we replace this solution in the equation (1.25), the growing mode can be written as

$$\delta_+ \propto H(t)\int_0^t \frac{dt'}{a^2(t')H^2(t')} \propto H(z)\int_z^\infty \frac{1+z'}{E^3(z')}dz',$$ (1.30)

where $E(z)$ is given by equation (1.11), expressed as a function of the redshift z (see equation (1.2) for the relation beween the scale factor and the redshift). In general the growing mode can be obtained by the equation (1.30) numerically for a generic cosmology in which the Dark energy term is described by the cosmological constant.

In figure 3 is shown the linear growing mode as a function of the scale factor for different models calculated numerically by PINOCCHIO. We notice that the growth is more rapid in the EDS model, while the development of the perturbations is slower in the case of open Universe ($\Omega_\Lambda = 0$). The model with cosmological costant gives an intermediate evolution between the other ones.

1.2.2 *Linear power spectrum, transfer function and correlation function*

The amplitude of fluctuations at different wave numbers (k) is described by the power spectrum. In section 1.2.1 we have seen that it is possible to expand the perturbation as a superposition of modes with different wave numbers:

$$\delta(x,t) = V\int \frac{d^3k}{(2\pi)^3}\delta_k(t)exp(ik\cdot x).$$ (1.31)

At each mode k it is possible to associate a length scale (λ), $k = 2\pi\lambda^{-1}$. The power spectrum is defined as:

$$P(k) = \left\langle |\delta_k(t)^2|\right\rangle$$ (1.32)

in the limit in which the volume V_u tends to inifinity (see previous section), $V_u \to \infty$ The power spectrum so describes the amplitude of the perturbations.
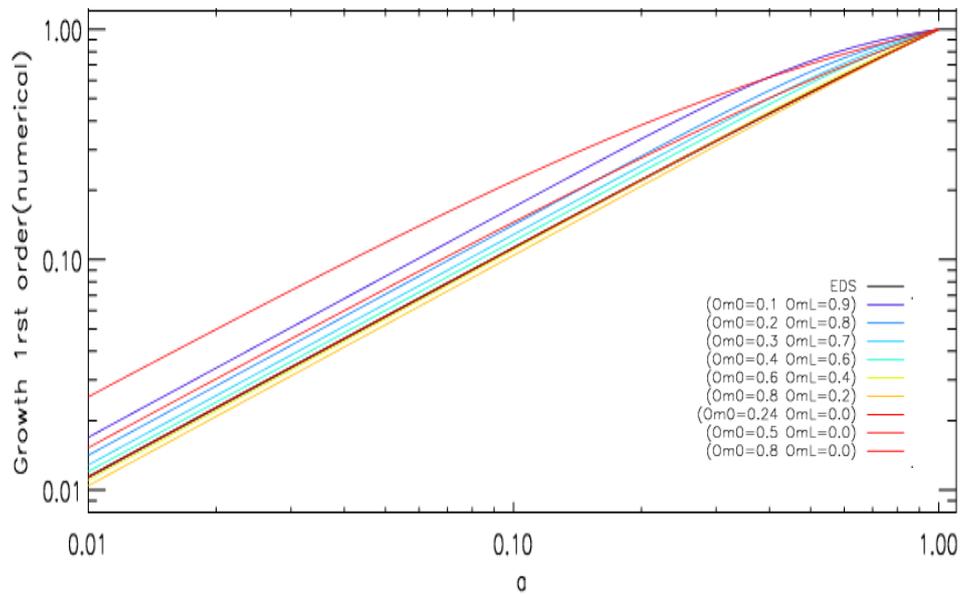
Figure 3: The top panel shows the linear growth factor, here noted as D, as a function of scale factor for different cosmological models, calculated numerically by PINOCCHIO.
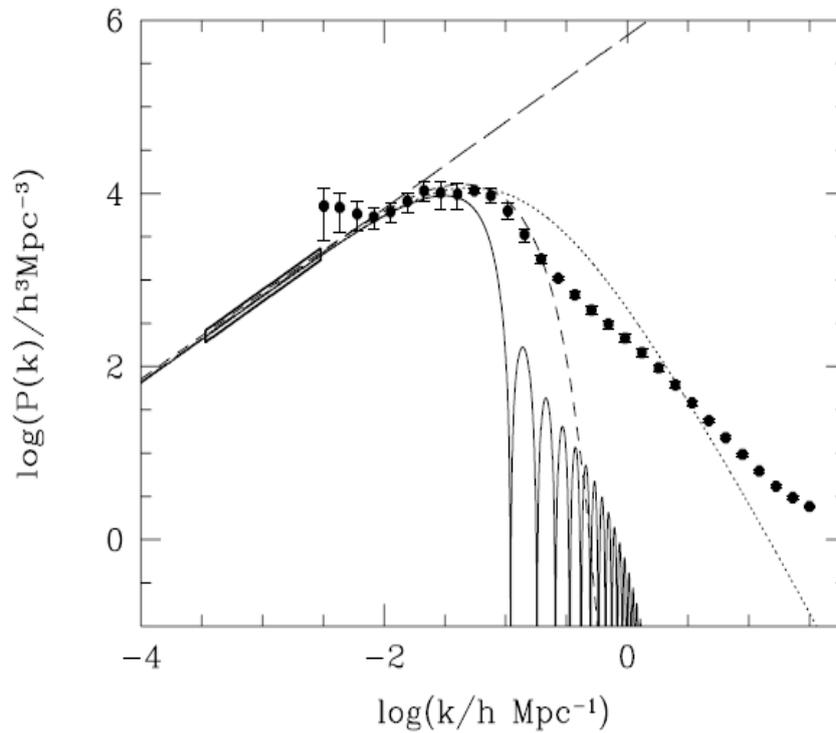
Figure 4: Examples of power spectra for universes with the critical density in mass. Long dashed line: the Harrison Zeldovich form of the primordial power spectrum; dotted line: power spectrum in a universe with the critical density in cold dark matter; solid line: power spectrum when baryons contribute all the critical density; short dashed line: universe in which all the mass is in the form of massive neutrinos. The points show a measurement of the power spectrum of galaxy distribution (C. Baugh, http://www.astro.caltech.edu/ george/ay21/eaa/eaa-powspec.pdf.

The power spectrum is the quantity predicted directly by theories for the formation of large scale structure. In the case of a density field in which the fluctuations are drawn from a Gaussian distribution, the power spectrum gives a complete statistical description of the fluctuations. The power spectrum generated by the inflaction at a given k is usually assumed to have a power law dependence on scale:

$$|\delta(k,t)|^2 \propto P(k) = Ak^n \tag{1.33}$$

A popular choice is the scale-invariant spectrum with spectral index $n = 1$, proposed independently by Zeldovich and Harrison, while A is a constant.

The rate at which fluctuations grow on different scales is determined by an interplay between self-gravitation, pressure support and damping processes. These effects lead to a modification of the form of the primordial power spectrum that is expressed in terms of a transfer function $T(k,z)$:

$$P(k,z) = Ak^n T(k,z). \tag{1.34}$$

where A is a costant of normalisation. To determine the shape and amplitude (A) of the linear power spectrum is one of the most important tasks of observational cosmology. One historical prescription for normalizing a theoretical power spectrum involves the variance of the galaxy distribution when sampled with randomly placed spheres of radii R. The predicted variance of the density field for a given radius R is related to the power spectrum by

$$\sigma^2(R) = \sigma_R^2 \frac{1}{2\pi} \int dk k^2 \hat{W}_R^2(k) P(k) \tag{1.35}$$

where

$$\hat{W}_R^2(k) = \frac{3}{(kR)^2} \left[\sin(kR) - kR\cos(kR) \right] \tag{1.36}$$

is the Fourier transform of the spherical top-hat window function:

$$W_R(r) = \frac{3}{4\pi R^3} \Theta(R - r) \tag{1.37}$$

where $\Theta(R - r)$ is the Heaviside step function. The value of $\sigma(R)$ derived from the distribution of galaxies, is about unity for $R = 8h^{-1}\text{Mpc}$. Thus, one could in principle normalize the theoretical power spectrum by requiring $\sigma_8 \sim 1$ at $R = 8h^{-1}\text{Mpc}$. However, there are several problems with this approach. First of all, since $\sigma(R) \sim 1$ we are not accurately probing the linear regime for which $\delta \ll 1$. Secondly, this normalization is based on the assumption that galaxies are accurate tracers of the fluctuations in the mass distribution. This may not be true if, for example, galaxies formed preferentially in high density regions. Indeed, if we adopt the less restrictive assumption that the fluctuations in the galaxy distribution are proportional (but not necessarily equal) to the fluctuations in the mass distribution, then:

$$\delta_{gal} = b\delta_m \tag{1.38}$$

where b = constant is a *bias* parameter whose value depends on how galaxies have formed in the mass density field.

In this case $\sigma_m(8h^{-1}Mpc) = \frac{\sigma_{gal}(8h^{-1}Mpc)}{b} = 1/b$. The value of b is still uncertain, since an accurate theory for galaxy formation is still lacking at the present time. In fact b is found to be a function of various galaxy properties, such as luminosity and color. To accurately normalize the linear power spectrum thus requires a method that is not affected by nonlinear evolution and that does not depend on the assumption of galaxies tracing the mass distribution. It is important to realize that σ_8 is evaluated from the initial power spectrum evolved to the present time according to linear theory. Since perturbations on scales of $8h^{-1}Mpc$ may well have gone nonlinear by the present time, this is not necessarily the same as the variance of the actual, present-day mass distribution.

The power spectrum has another important caracteristics: it is linked by the Fourier transform to the correlation function $\xi(r)$, that is a measure of the degree of clustering in the spatial distribution of structures. The spatial two-point or autocorrelation function is defined as the excess probability, compared with that expected for a random distribution, of finding a pair of objects (galaxies, cluster of galaxies . . .) at a given separation r_{12}. On small scales, $0.1h^{-1}Mpc \leqslant r \leqslant 10h^{-1}Mpc$, the spatial correlation function is well described by a power law form $\xi(r) = (r_0/r)^\gamma$ with $\gamma \sim 1.8$ and a correlation length $r_0 \sim 5h^{-1}Mpc$ The power spectrum P(k) is related to the two-point correlation function by:

$$\xi(r) = \frac{1}{2\pi^2} \int dk k^2 P(k) \frac{\sin(kr)}{kr} \tag{1.39}$$

1.3 NON LINEAR THEORY OF PERTURBATIONS

The evolution of a small amplitude density fluctuation on a particular scale can be followed independently of fluctuations that may exist on other length scales using linear perturbation theory. In the later stages of collapse, fluctuations on different length scales become coupled and the subsequent evolution is non-linear. While the linear theory provides exact solutions, it's impossible to follow the non-linear stage in an analytical way. So one is forced to use approximated tecniques like higher order perturbation methods or numerical N-body simulations. Non-linear evolution of density fluctuations, changes the shape of the power spectrum. To complete our description of the role of dark matter in structure formation in the Universe, we therefore need to go beyond perturbation growth in the linear regime, discussed in the previous section, and address the gravitational collapse of overdensities in the non-linear regime.

1.3.1 *Lagrangian perturbation theory (LPT)*

In the previous section we have described the evolution of linear perturbations in an Euleurian approach, i.e. we have considered volumes of fluid fixed in space, of which we have studied the variations of the density and the velocity at fixed spatial position. A first possible extension of the linear Eulerian theory would be increasing the order in expansion of $\delta(\mathbf{x})$ and $\mathbf{v}(\mathbf{x})$. However while this approach

is cleary wrong for $\delta(\mathbf{x}) \sim 1$, the series of expansion are not guaranted to converge already at $\delta(\mathbf{x}) \lesssim 1$. Zeldovich in 1970 proposed a kinematical approach as an alternative description to the formation of structure. The approximation proposed, was just the first order of the Lagrangian perturbation theory. Later other authors (Buchert et al., 1994; Moutarde et al., 1991; Buchert and Ehlers, 1993), considered higher orders of the Lagrangian perturbation theory. The advantage of the Zeldovich approximation is that it normally breaks down later than Eulerian linear theory.

In the Lagrangian perturbation theory the dynamical variable is the Lagrangian displacement field $\mathbf{s}(\mathbf{q}, t)$ of a mass element:

$$\mathbf{r}(\mathbf{q}, t) = a(t)[\mathbf{x}(\mathbf{q}, t)] = a(t)[\mathbf{q} + \mathbf{s}(\mathbf{q}, t)]. \tag{1.40}$$

where, in fluid-mechanical terminology, $\mathbf{x}$ is said to be the Eulerian position, and $\mathbf{q}$ the Lagrangian position, $\mathbf{r}$ is the proper coordinate. Initially $\mathbf{s}(\mathbf{q}, t) = 0$ so that $\mathbf{q}$ is the same as the usual comoving cordinate at initial time $t = 0$ The particle trajectory in the expanding universe is:

$$\frac{d^2\mathbf{x}(t)}{dt^2} + 2H(t)\frac{d\mathbf{x}(t)}{dt} = -\frac{\nabla \Phi}{a} \tag{1.41}$$

This is the Eulerian equation (1.19) reported as function of the comoving cordinate $\mathbf{x}$ and considering the total time derivative instead of the partial one. We find the same equation of motion by a Lagrangian approach, using the Lagrangian equation for a particle and applying the variational principle of Hamilton. Taking the divergence of the previous equation, using the Poisson equation (1.20), expressing the density as a function of the critical one (see equation (1.10)) and as a function of the density contrast (equation (1.21)) we obtain:

$$\nabla_x \cdot \left[\frac{d^2\mathbf{x}(t)}{dt^2} + 2H(t)\frac{d\mathbf{x}(t)}{dt} \right] = -\frac{\nabla_x^2 \Phi}{a} = \frac{3a^2(t)H^2(t)\Omega_m \delta(\mathbf{x}(t))}{2} \tag{1.42}$$

The particle density in Lagrangian coordinates, of course is the same as the average density of the Universe, so for mass conservation we have:

$$\bar{\rho}(t)d^3q = \rho(\mathbf{x}(t))d^3x = \bar{\rho}(t) = \bar{\rho}(t)[1 + \delta(\mathbf{x}, t)]d^3x \tag{1.43}$$

$\mathbf{x}$ is always function of $\mathbf{q}$ and t in this approach. In this way taking into account of the equation (1.40) and (1.42), we can relate the eulerian density constrast to the lagrangian displacement:

$$1 + \delta(\mathbf{x}, t) = |\frac{d^3q}{d^3x}| = 1/[\det(\delta_{ij} + s_{i,j}(\mathbf{q}, t)] = 1/J(\mathbf{q}, t) \tag{1.44}$$

where $J(\mathbf{q}, t)$ is the jacobian of the Lagrangian to Eulerian coordinate mapping and, $s_{i,j}(\mathbf{q}, t) = \frac{ds_i(\mathbf{q}, t)}{dq_j}$.

Now replacing the equation (1.44) for the density contrast in the equation of motion, the equation of motion for the displacement $\mathbf{s}(\mathbf{q}, t)$ is:

$$J(\mathbf{q}, t)[\delta_{ij} + s_{i,j}(\mathbf{q}, t)]^{-1} \left[\frac{d^2s_{i,j}(\mathbf{q}, t)}{dt^2} + 2H(t)\frac{ds_{i,j}(\mathbf{q}, t)}{dt} \right] = \frac{3a^2(t)H^2(t)\Omega_m[J(\mathbf{q}, t) - 1]}{2}.$$

$$(1.45)$$

This last equation is the master equation of the Lagrangian perturbation theory. To have the perturbative solution, in $\mathbf{s}(\mathbf{q}, t)$:

$$\mathbf{s}(\mathbf{q}, t) = \mathbf{s}^{(1)}(\mathbf{q}, t) + \mathbf{s}^{(2)}(\mathbf{q}, t) + \dots \qquad (1.46)$$

1.3.1.1   *First order LPT: Zeldovich approximation*

In the Lagrangian scenario a simple and elegant approximation to describe the non-linear stage of gravitational evolution has been developed by Zel'dovich (1970) (see the review by Shandarin and Zeldovich, 1989, for an exhaustive description of the Zeldovich approximation). In this approach, the initial matter distribution is considered to be homogeneous and collisionless.

Let's consider the initial displacement of particles and that they continue to move in the initial direction. Then the comoving cordinate is given by:

$$\mathbf{r}(\mathbf{q}, t) = a(t)[\mathbf{x}(\mathbf{q}, t)] = a(t)[\mathbf{q} + \mathbf{s}^{(1)}(\mathbf{q}, t)]. \qquad (1.47)$$

This corresponds to a truncation at the first order of the perturbative expansion of equation (1.40). The Zeldovich approximation consists in assuming small values for the displacement vector $\mathbf{s}(\mathbf{q}, t)$, therefore only the first term of equation (1.46) is considered. Under this condition is possible to factorize the displacement in the time dependence and in the $\mathbf{q}$-dependence:

$$\mathbf{s}^{(1)}(\mathbf{q}, t) = D^{(1)}(t)\mathbf{p}^{(1)}(\mathbf{q}) \qquad (1.48)$$

The inverse of the Jacobian we find in equation (1.45) becomes:

$$[\delta_{ij} + s_{i,j}(\mathbf{q}, t)]^{-1} \simeq \delta_{ij} - s_{i,j}(\mathbf{q}, t) = \delta_{ij} - D^{(1)}(t)\frac{dp_j^{(1)}(\mathbf{q})}{dq_i} \qquad (1.49)$$

where the term $\frac{dp_j^{(1)}(\mathbf{q})}{dq_i}$ is called deformation tensor. The master equation of the motion becomes:

$$1 + s_{k,k}(\mathbf{q}, t))[\delta_{ij} - s_{i,j}(\mathbf{q}, t)\left[\frac{d^2 s_{i,j}(\mathbf{q}, t)}{dt^2} + H(t)\frac{ds_{i,j}(\mathbf{q}, t)}{dt}\right] =$$
$$= \frac{3a(t)^2 H(t)^2 \Omega_m s_{k,k}(\mathbf{q}, t)}{2} \text{ i.e.} \qquad (1.50)$$

$$\left[\frac{d^2 s_{i,i}^{(1)}(\mathbf{q}, t)}{dt^2} + H(t)\frac{ds_{i,i}^{(1)}(\mathbf{q}, t)}{dt}\right] = \frac{3a(t)^2 H(t)^2 \Omega_m s_{i,i}^{(1)}(\mathbf{q}, t)}{2} \qquad (1.51)$$

Because of the factorization and the curl free condition of the Lagrangian displacement, that is irrotational, it's possible to write the spatial function $\mathbf{p}^{(1)}(\mathbf{q}) = -\nabla\phi^{(1)}(\mathbf{q})$ and $\mathbf{s}^{(1)}(\mathbf{x}, t) = -D^{(1)}(t)\nabla\phi^{(1)}(\mathbf{q})$. Using the previous expression and the relation for the density contrast (see equation (1.44)) $1 + \delta^{(1)}(\mathbf{x}, t) = J^{-1} \simeq 1 - s_{i,i}(\mathbf{q}, t)$. The linear solution becomes:

$$\nabla_q \cdot \mathbf{s}^{(1)}(\mathbf{x}, t) = -\delta^{(1)}(\mathbf{x}, t) = -D^{(1)}(t)\nabla_q^2\phi^{(1)}(\mathbf{q}) \qquad (1.52)$$

where the $\nabla\phi^{(1)}$ is is the linear Lagrangian potential, which is related to the linear density field. The time evolution of $\delta^{(1)}(\mathbf{x}, t)$ obeys to the equation:

$$\frac{d^2 D^{(1)}(t)}{dt^2} + 2H(t)\frac{dD^{(1)}(t)}{dt} = \frac{3a(t)^2 H(t)^2 \Omega_{mo}}{2} D^{(1)}(t) \qquad (1.53)$$

that is equivalent to the equation (1.22), of the eulerian theory, so, with the same solutions. $D^{(1)}$ corresponds to the linear growing mode. The particle position and the corrisponding velocity in the ZA is given by:

$$\mathbf{x} = \mathbf{q} - D^{(1)}(t)\nabla\phi^{(1)}(\mathbf{q}) \qquad (1.54)$$

$$\mathbf{v} = a(t)\frac{d\mathbf{x}}{dt} = -a(t)\frac{dD^{(1)}(t)}{dt}\nabla\phi^{(1)}(\mathbf{q}) \qquad (1.55)$$

$$(1.56)$$

which are also written respectively as:

$$\mathbf{x} = \mathbf{q} - \mathbf{v}^{(1)}t \qquad (1.57)$$

$$\mathbf{v} = a(t)f^1 H(t)\frac{dD^{(1)}(t)}{dt} \qquad (1.58)$$

$$(1.59)$$

where f^1 is $d\ln D^{(1)}/d\ln a$. The second expression for the position of the particle approximates the motion of a fluid element in a laminar flow, with no intersection of the trajectories. Under this approximation, the fluid element feels an initial acceleration proportional to the gradient of the potential $\phi^{(1)}(\mathbf{q})$ only at the beginning. The density diverges as the Jacobian determinant vanishes, forming a caustic. In this case the trasformation between the Eulerian coordinate $\mathbf{x}$ and the Lagrangian one $\mathbf{q}$ becomes multivalued, i.e particles undergo orbit crossing (OC). The OC defines the time in which the collapse of the fluid element occurs. In this condition the Zel'dovich approximation breaks down. When overdensities grow, the non-linear effect due to the selfgravity of the perturbation, is not properly reproduced. The particles continue their initial motion determined by the equation (1.55), instead of experiencing the effect determined by the generated overdesities, then crossing the caustics. The first perturbations to collapse and to condense as a selfgravitating structure, are the small scale ones, later become larger ones. The advantage of this approximation is that it normally breaks down later than Eulerian linear theory. As a consequence of such an approximation, the theory predict what's called Zel'dovich pancake. The deformation tensor is diagonalizable, because $\mathbf{p}(\mathbf{q})$, as we've seen, is the gradient of a scalar function, so it is real and symmetric. Let $\alpha(\mathbf{q})$, $\beta(\mathbf{q})$, $\gamma(\mathbf{q})$ be the eigenvalues associated to the eigenvectors of the strain tensor that define a set of three principal (orthogonal) axes, which give the contraction or expansion along the three principal axes. The density can be written:

$$\rho(\mathbf{q}, t) = \frac{\bar{\rho}}{[1 - D^{(1)}(t)\alpha(\mathbf{q})][1 - D^{(1)}\beta(\mathbf{q})][1 - D^{(1)}\gamma(\mathbf{q})]}. \qquad (1.60)$$

If the eigenvalues are ordered such that $\alpha(\mathbf{q}) \geqslant \beta(\mathbf{q}) \geqslant \gamma(\mathbf{q})$, when $D^{(1)}(t_1) = \alpha_{max}^{-1}$, a first singularity in equation (1.60) occurs. This corresponds to the

formation of a pancake (sheet-like structure) by contraction along one of the principal axes. For this reason, Zeldovich argued that pancakes are the first structures formed by gravitational clustering. Other structures like filaments and knots come from simultaneous contractions along two and three axes, respectively. The Zeldovich approximation predicts the first non-linear structure to arise in correspondence of the high peaks of the $\alpha(\mathbf{q})$ field and represents a significant step forward with respect to linear theory. It has been successfully applied for example to generate the initial conditions of simulations (see e.g. Villumsen, 1989; Pauls and Melott, 1995; Seto, 1998).

1.3.1.2 *Second order LPT*

It has been proven that high-order corrections of the Lagrangian perturbation theory improve the result obtained from first order solution up to OC (Buchert et al., 1994; Moutarde et al., 1991; Buchert and Ehlers, 1993), . The second order Lagrangian solution has been derived for specific cosmologies: for generic non flat Friedmann models without cosmological constant by Buchert et al. (1994), with the same cosmology but with a different formalism by Catelan (1995), for flat model with no cosmological constant by Bouchet (1996), with a relativistic approach in an expanding universe by Matarrese et al. (1994). We are here more interested for this work more at the general equation of evolution of the grow at the second order rather than to the solutions for a particular cosmology.

Similarly to first order, the the displacement of the second order can also be factorized:

$$s^{(2)}(\mathbf{q}, t) = D^{(2)}(t)\mathbf{p}^{(2)}(\mathbf{q}, \mathbf{t_i}) \tag{1.61}$$

Proceeding for the first order LPT, we consider for the second order the Jacobian of equation (1.44) and the equation of motion (1.45). Using the solution for $\mathbf{s}^{(1)}(\mathbf{q}, t)$, and the simmetry $s_{i,j}^{(1)}(\mathbf{q}, t) = s_{j,i}^{(1)}(\mathbf{q}, t)$, it's possible to verify that the time dependent part of $\mathbf{s}^{(2)}$, $D^{(2)}(t)$, that is the growing mode of the perturbation at the second order, solution of:

$$\frac{d^2 D^{(2)}(t)}{dt^2} + 2H(t)\frac{dD^{(2)}(t)}{dt} - \frac{3a(t)^2 H(t)^2 \Omega_{mo}}{2}D^{(2)}(t) = -\frac{3a^2(t)H^2(t)\Omega_m}{2}[D^{(1)}(t)]^2 \tag{1.62}$$

Buchert et al. (1994) found the analytical solution for a flat ΛCDM universe $D^{(2)}(t) \simeq -3(D^{(1)}(t))^2 \Omega_m^{1/143}/7$ that approximates the exact solution for $D^{(2)}$ betten than 6 per cent. Because of the factorization and the curl free condition of the Lagrangian displacement $\mathbf{s}^{(2)}(\mathbf{q}, t) = D^{(2)}(t)\mathbf{p}^{(2)}(\mathbf{q}, t_i)$, it's possible to write the spatial function $p^{(2)}(\mathbf{q}, t_i) = -\nabla\phi^{(2)}(\mathbf{q}, t_i)$. The spatial part of the second order solution describes the effect of the second order gravitational tidal field:

$$\mathbf{p}^{(2)}(\mathbf{q}, t_i) = \frac{1}{(D^{(1)}(t_i))^2}\frac{1}{2}\sum_{i \neq j}\left\{s_{i,i}^{(1)}(\mathbf{q}, t_i)s^{(1)}{}_{j,j}(\mathbf{q}, t_i) - s^{(1)}{}_{i,j}(\mathbf{q}, t_i)s^{(1)}{}_{j,i}(\mathbf{q}, t_i)\right\} \tag{1.63}$$

We recall that $s_{j,i}^{(1)}(\mathbf{q}, t_i) = D^{(1)}(t_i)\frac{dp_i^{(1)}(\mathbf{q})}{dq_i}$. Multiplying this last equation by $D^{(2)}(t)$ we can have an approximation for the displacement to second order, using the approximation for $D^{(2)}(t)$ by Bouchet (1996). Finally, the solution for the position and for the velocity up to second order is given by:

$$
\begin{aligned}
\mathbf{x}(\mathbf{q}, t) &= q + s^{(1)}(\mathbf{q}, t) + s^{(2)}(\mathbf{q}, t) \\
&= q - D^{(1)}(t)p^{(1)}(\mathbf{q}) + D^{(2)}(t)p^{(2)}(\mathbf{q})
\end{aligned}
\tag{1.64}
$$

$$
\begin{aligned}
\mathbf{v}(\mathbf{q}, t) &= a(t)\frac{d\mathbf{x}}{dt} = \\
&= a(t)f_1 H(t)\frac{dD^{(1)}(t)}{dt} + a(t)f_2 H(t)\frac{dD^{(2)}(t)}{dt}
\end{aligned}
\tag{1.65}
$$

where $f_i = dlnD_i/dlna$ that is well approximated by: $f_1 \simeq [\Omega_m(z)]^{5/9}$ and $f_2 \simeq 2[\Omega_m(z)]^{6/11}$.

1.3.2 *The collapse stage: spherical and ellipsoidal collapse*

As already emphasized, nonlinear gravitational dynamics is difficult to deal with analytically. However, if simple assumptions are made about the symmetry of the system, analytical models can still be constructed. Although these models are not expected to give accurate descriptions of the true nonlinear problem of gravitational collapse, they provide valuable insight into the complex processes involved.

1.3.2.1 *Spherical collapse*

Spherical symmetry is one of the few cases in which gravitational collapse can be solved exactly (Peebles, 1980). In fact, as a consequence of Birkhoff's theorem, a spherical perturbation evolves as a FRW Universe with density equal to the mean density inside the perturbation. The simplest spherical perturbation is the top-hat one, i.e. a constant overdensity δ inside a sphere of radius R; to avoid a feedback reaction on the background model, the overdensity has to be surrounded by a spherical underdense shell, such to make the total perturbation vanish. This approach is based on assuming the perturbation as a closed Universe in itself and imposing that the border speed of the perturbation is zero at an initial time t_i.

In EDS Universe ($\Omega_m = 1$), the above assumptions give the relation $D(t_i) = (3/5)\delta(t_i)$, between the linear growing mode and the initial overdensity. In order for the perturbation to collapse, it's necessary that its inizial density parameter $\Omega_p(t_i) = \Omega(t_i)(1 + \delta_i)$ is larger than 1. The evolution of a the perturbation radius depends therefore only on its initial overdensity and is then given by a Friedmann equation. In the Einstein-de Sitter background, any spherical overdensity reaches a singularity (collapse) at a final time:

$$
t_c = \frac{3\pi}{2}\left(\frac{5\delta t_i}{3}\right)^{-3/2} t_i
\tag{1.66}
$$

By that time its linear density contrast reaches the value:

$$\delta_l(t_c) = \delta_c = \frac{3}{5}\left(\frac{3\pi}{2}\right)^{3/2} = 1.686 \tag{1.67}$$

In an open Universe not all overdensities will collapse: for this to happen the initial density contrast has to be such that the total density inside the perturbation overcomes the critical density. This can be quantified (not exactly but very accurately) as follows: the growing mode saturates at $D(t) = 5/2(\Omega_m^{-1})$, so that a perturbation ought to satisfy $\delta_l > 1.69 \cdot 2(\Omega_m^{-1})/5)$ to be able to collapse. Of course, collapse to a singularity is not what really happens in reality. It is typically supposed that the structure reaches virial equilibrium at that time. In this case, arguments based on the virial theorem and on energy conservation show that the structure reaches a radius equal to half its maximum expansion radius, and a density contrast of about $\delta_{vir} \sim 178$.

Equation (1.67) gives a quite important quantity that is used also to describe the mass function of the virialized halos. In fact it gives the density contrast for a perturbation in the initial density field to evolve in a collapsed virialized structure. One of the advantages of spherical symmetry is that, because of Birkhoff's theorem, it is possible to introduce in a background metric a perturbation without affecting the rest of the Universe, provided any positive perturbation is compensated for by an (outer) negative one, such to make the total mass perturbation vanish. This is necessary to ensure the self-consistency of the problem: the background has to evolve as if it were unperturbed. This reasoning is not valid any more when introducing a triaxial perturbation in an unperturbed background: this is going to influence the background, through non-linear feedback effects.

1.3.2.2 *Ellipsoidal collapse*

The convenience in using the homogeneous ellipsoid collapse model resides in the fact that it can easily be solved by means of a numerical integration of a system of three second-order ordinary differential equations.

A homogeneous triaxial ellipsoid is characterized by its mean overdensity and its axial ratios; it can experience a global expansion, a deformation or a global rotation. A homogeneous ellipsoid possesses a "minimal" geometric complexity which makes its structure analogous to that of a mass element. As a matter of fact, it is possible to write down evolution equations for a Newtonian homogeneous ellipsoid.

The fundamental difference between a homogeneous ellipsoid and a generic mass element is in the role of the potential, a quadratic form in the first case and a whole random (Gaussian) field in the other. To extract an ellipsoid from a perturbed potential field (Bond and Myers, 1996a) in a point, it suffices to expand the potential around that point in a Taylor series.

To use ellipsoidal collapse in a cosmological context, the correct strategy is not to try to insert an ellipsoid in a uniform background, but to extract an ellipsoid from a perturbed FRW Universe (Monaco, 1998).

The dynamical variables of ellipsoidal collapse are the three axes $a_i(t)$ of the ellipsoid; they are normalized as the scale factor: $a_i(t) = a(t)$ if the ellipsoid is a sphere with null density contrast. Their evolution equations are, in a open and flat Universe with cosmological constant (the EdS is obtained considering the present parameter density $\Omega_{mo} = 1$):

$$\frac{d^2 a_i}{da^2} - (2a(1+\Omega_{mo}^{-1})a)^{-}1\frac{da_i}{da} + (2a^2(1+\Omega_{mo}^{-1})a)^{-}1a_i\left[\frac{1}{3} + \frac{\delta}{3} + \frac{b\prime_i}{2} + \lambda\prime_{vi}\right] = 0;$$

(1.68)

$$\frac{d^2 a_i}{da^2} - \frac{1 - 2\Omega_{mo}^{-1}a^3}{2a(1+(\Omega_{mo}^{-1}-1)a^3)}\frac{da_i}{da} + (2a^2(1+\Omega_{mo}^{-1})a)^{-}1a_i\left[\frac{1}{3} + \frac{\delta}{3} + \frac{b\prime_i}{2} + \lambda\prime_{vi}\right] = 0.$$

(1.69)

respectively, where the density contrast δ is given by:

$$\delta = \frac{a^3}{a_1 a_2 a_3} - 1,$$

(1.70)

while the quantities b'_i and λ'_{vi} are defined as:

$$b'_i = \frac{2}{3}[a_i a_j a_k R_D(a_i^2, a_j^2, a_k^2) - 1] \qquad i \neq j \neq k$$

(1.71)

(where the R_D is the Carlson's elliptical integral

$$R_D(x,y,z) = \frac{3}{2}\int_0^\infty \frac{d\tau}{(\tau+x)^{1/2}(\tau+y)^{1/2}(\tau+z)^{3/2}},$$

(1.72)

and

$$\lambda'_{vi} = -\frac{a}{a_0}\left(\frac{\delta}{3} - a_0\lambda_i\right)..$$

(1.73)

The three coupled second-order ordinary differential equations, given by the (1.68) or (1.69), can be solved numerically, by means of standard routines.

The ellipsoidal collapse is used to describe the evolution of a mass element at early times. Until the time the collapse happens on its shortest axis, its principal axis are given by the deformation tensor $\frac{dp_j^{(1)}(\mathbf{q})}{dq_i}$ (see 1.3.1.1). Under the assumption that the evolution of the ellipsoid is independent of the cosmology, a good approximation of its evolution it is given by the LPT up to third order (Monaco, 1998).

The homogeneous ellipsoid collapse model has been used in different cosmological context to estimate collapse times: by Bond and Myers (1996a), Bond and Myers (1996b), Bond and Myers (1996c)), White and Silk (1979)), Monaco (1998).

1.3.3   Statistic of the hierarchical clustering

The scenario presented in the previous chapters describes the current theory of hierarchical clustering, according to which structures grow hierarchically from initial gaussian density fluctuations. We now turn to the statistical treatment of hierarchical clustering. A first attempt to analytically describe complessively the statistic of the hierarchical clustering, was made by Press and Schechter (1974) (PS). Despite its semplicity, the PS formalism allows one to calculate many properties of the population of dark matter halos, such as their mass function, the mass distribution of their progenitors, their merger rate. The analytic nature of the extended Press–Schechter (EPS) formalism helps us to understand how the properties of the halo population are related to the cosmological framework.

   Its not-rigorous nature implies, however, that its predictions should always be checked using other methods, primarily numerical N-body simulations of cosmic structure formation. This approach has developed in the two last decades thanks to huge advancement in computer tecnology. The approach consists in reproduce a simulation of a dynamical system of particles under the gravity force of gravity.

1.3.3.1   The Mass Function: the Press-Schechter formalism

In this section we focus on the statistical properties of the dark matter halos. Clearly, since dark matter halos are the hosts of galaxies, these properties will have a direct link to the clustering properties of galaxies.

   Let's consider a density field $\delta(\mathbf{x}, t)$, that in linear regime evolves as $\delta(\mathbf{x}, t) = \delta_0(\mathbf{x})D(t)$, where $\delta_0(\mathbf{x})$ is the overdensity linearly extrapolated until today, and $D(t)$ the linear growing mode nomalized to 1 at t_0. According to the spherical collapse model, described in in section 1.3.2, regions with $\delta(\mathbf{x}, t) > \delta_c \simeq 1.69$, or similarly with $\delta_0(\mathbf{x}) > \delta_c/D(t) \equiv \delta_c(t)$, collapse, forming virialized objects. We want to assing a mass to these regions. Let be the smoothed density field be defined as

$$\delta_S(\mathbf{x}; R) \equiv \int \delta_0(\mathbf{x}')W(\mathbf{x} + \mathbf{x}'; R)d\mathbf{x}, \tag{1.74}$$

where $W(\mathbf{x}; R)$ is a filter with radius R, corrisponding to a mass $M = \gamma_f \bar{\rho} R^3$, with the parameter γ_f that depends on the filter form. The ansatz of the PS formalism consists in assuming the probability of $\delta_s > \delta_c(t)$ to be the same of the mass elements fraction that at time t are in halos of mass bigger than M. If $\delta_0(\mathbf{x})$ is a Gaussian field, the same is $\delta_s(\mathbf{x})$, and the probability of $\delta_s > \delta_c(t)$ is given by:

$$\mathscr{P}[> \delta_c(t)] = \frac{1}{\sqrt{2\pi}\sigma(M)} \int_{\delta_c(t)}^{\infty} \exp\left[-\frac{\delta_S^2}{2\sigma^2(M)}\right] d\delta_s = \frac{1}{2}\text{erfc}\left[\frac{\delta_c(t)}{\sqrt{2}\sigma(M)}\right], \tag{1.75}$$

where $\sigma^2(R = (M/\gamma_f \bar{\rho})^{1/3}) = \sigma^2(M)$ is defined by the equation (1.35). According to the PS ansazt, the probability (1.75) is equal to $F(> M)$, the mass fraction of collapsed object with mass larger than M. The Press-Schechter has a problem; for $M \to 0$ ($\sigma(M) \to \infty$) one expects that $\mathscr{P}[> \delta_c(t)] \to 1$, i.e all

the matter of the Universe is contained in arbitrary small mass halos. Instead equation (1.75) $\mathscr{P}[> \delta_c(t)] \to 1/2$, gives only half of the total mass is part of collapsed objects. This is because it's given null probability at a point with $\delta_s(\mathbf{x}, M) < \delta_c$ for a given smoothing scale M, to have $\delta_s(\mathbf{x}, M') > \delta_c$ for a bigger smoothing scale $M' > M$. In other words this approach does not consider the probability for that point to be included in a larger mass halo. A more rigorous derivation for the mass function, is based on the *excursion-set formalism* (Bond et al., 1991) (EPS), that explains the missing factor of two, adopting a spherical top-hat filter in Fourier space. The mass function that is obtained at least when $F(> M) = 2\mathscr{P}[> \delta_c(t)]$ is:

$$
\begin{aligned}
n(M, t)dM &= \frac{\bar{\rho}}{M} \frac{\partial F(> M)}{\partial M} dM = 2 \frac{\partial \mathscr{P}[> \delta_c(t)]}{\partial \sigma_M} \left| \frac{d\sigma_M}{dM} \right| dM \\
&= \sqrt{\frac{2}{\pi}} \frac{\bar{\rho}}{M^2} \frac{\delta_c}{\sigma_M} \exp\left[-\frac{\delta_c^2}{2\sigma_M^2} \right] \left| \frac{d \ln \sigma_M}{d \ln M} \right| dM.
\end{aligned}
\tag{1.76}
$$

This is known as the PS mass function Press and Schechter (1974). From equation (1.76), halos with mass M can form in a significative number only when $\sigma_M \gtrsim \delta_c(t)$. Defining the caracteristic mass $M^*(t)$ as

$$
\sigma(M^*) = \delta_c(t) = \delta_c/D(t),
\tag{1.77}
$$

only the halos with $M \lesssim M^*$ will form in a significative number at time t.

Because in the hierarchical model $D(t)$ is a growing function of time, while σ_M decreases with the mass, the caracteristic mass M^* grows with time. So progressively larger halos form as time goes by.

Qualitatively, the prediction is that the mass distribution is a power law for small masses, with an exponential cutoff above characteristic mass, M^*, whose value increases with time.

The mass function of equation (1.76) is sensitive to the values of the cosmological parameters that enter in the equation (1.76) for σ_M, that depends on the power spectrum, on the density parameters and on the linear growing.

In figure 5 we show the cumulative mass function of the halos for different cosmological parameters as function of the redshift. The different models are normalized to the mass function at $z = 0$.

1.3.4 *Beyond the PS formalism*

An alternative way to express the mass function is given by

$$
f(\sigma_m, z) = \frac{M}{\bar{\rho}} \frac{dn(M, z)}{d \ln \sigma_m^{-1}},
\tag{1.78}
$$

defined as the mass fraction associated to halos for unit range of $\ln \sigma^{-1}$. With this definition the PS mass function is expressed as

$$
f_{PS}(\sigma_m, z) = \sqrt{\frac{2}{\pi}} \frac{\delta_c}{\sigma_M} \exp\left[-\frac{\delta_c^2}{2\sigma_M^2} \right].
\tag{1.79}
$$
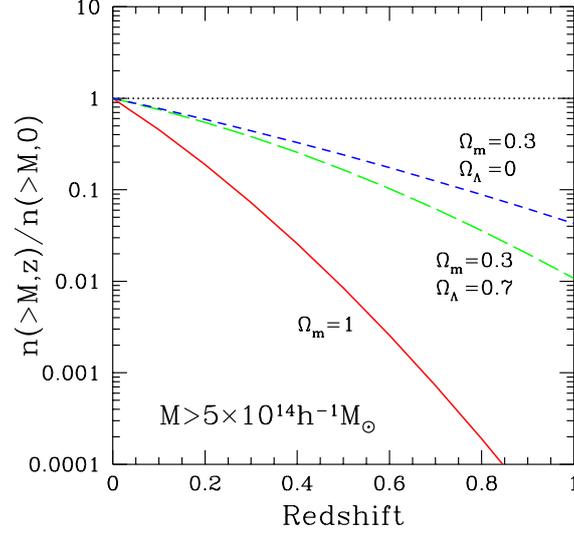
Figure 5: Cumulative mass function $n(> M, z)$ as a function of the redshift with $M = 5 \times 10^{14} h^{-1} M_{\odot}$ for three different cosmological model (Rosati, Borgani & Norman (2002) Rosati et al. (2002)): (*solid line*) $\Omega_m = 1$ e $\sigma_8 = 0.5$, (*long dash line*) $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$ e $\sigma_8 = 0.8$, (*short dash line*) $\Omega_m = 0.3$, $\Omega_\Lambda = 0$ e $\sigma_8 = 0.8$

The Press & Schechter mass function, despite its semplicity, has been used for more than a decade. Only recently, thanks to N-body simulations, able to cover a wide dynamical range in halo mass, significative deviations from the PS expression are appreciated. These deviations are usualy interpretated as corrections to the PS approach.

One of the first steps in deriving more accurate expression for the mass function is to consider an ellipsoidal collapse rather than a sperical one, as assumed in the PS approach. Including their effects for this kind of collapse the mass function can be generalized to

$$f(\sigma_m, z) = \sqrt{\frac{2a}{\pi}} C \left[1 + \left(\frac{\sigma_M^2}{a \delta_c^2} \right) \right] \frac{\delta_c}{\sigma_M} \exp \left[-\frac{a \delta_c^2}{2 \sigma_M^2} \right] \tag{1.80}$$

(Sheth and Tormen, 1999). Comparing this expression with the results obtained from N-body simulations, and imposing that $\int_0^\infty f(\sigma_M) d\nu = 1$, the following parameters are obtained: $a = 0.707$, $q = 0.3$, e $C = 0.3222$.

Others have found an alternative expression for the mass function, as a fitting function to the mass function measured from N-body simulations. Warren et al. (2006) have proposed the following:

$$f(\sigma_m, z) = A(\sigma^{-a} + b) \exp^{-\frac{c}{\sigma^2}}), \tag{1.81}$$

where A, a, b, c are the fit parameters (see list (8) in the paper of Warren et al., 2006). The function $f(\sigma_m, z)$ turns out to be sensitive to the algorithm used in the simulations to identify halos and to estimate their mass (Evrard et al., 2002).

The PS expression for the mass function understimates the number of massive halos and overstimate the ones below M^*.

Intermediate methods between simulations and analytical approaches, are represented by perturbative approaches that describe the growth of haloes in a given numerical realisation of a linear density field, such as the (1truncated Zel'dovich approximation (Borgani et al., 1994), the peak-patch algorithm (Bond and Myers, 1996b), (Bond and Myers, 1996a)) and the merging cell model (Lanzoni et al., 2000). A new algorithm PINOCCHIO, to compute the formation and evolution of dark matter haloes in a given linear density field is proposed in 1999, by Monaco and Murante (1999). PINOCCHIO is acronym for PINpointing Orbit Crossing Collapsed HIerarchical Objects. In common with the other perturbative approaches, it uses a local description of the dynamics, in order to identify collapsed haloes. The Lagrangian perturbation theory is used to displace the haloes to their final positions (Monaco, 1994, 1997a,b). The innovation of thit algorithm lies in the several orders of magnitude less computer time and post analysis than the corresponding full blown numerical simulations with an accurate description of the detailed clustering and merger history of haloes. In addition, the successful reproduction of the merger history of halos it provides, demonstrates that it identifies the key processes that govern halo formation, and that these can be described with a perturbative approach. In the next chapter we will describe in detail PINOCCHIO.

THE PINOCCHIO CODE

<div style="text-align: right; font-size: 3em;">*2*</div>

Today numerical simulations are the tool for describing and investigating the non-linear evolution and the hierarchical clustering. N-body simulations in particular generate the non-linear distribution of dark matter.

Recent works are based on the Lagrangian perturbation theory (LPT), that is the perturbative solutions of a system of equations for the computation of the displacement of mass elements from their initial positions.

A powerful algorithm that operate in this line and that we present in this chapter is PINOCCHIO (PINpointing Orbit-Crossing Collapsed HIerarchical Objects). PINOCCHIO is able to identify dark matter haloes in a given numerical realization of the linear density field in a hierarchical universe (Monaco et al., 2002b). It exploits the Lagrangian perturbation theory, and in particular its ellipsoidal truncation which can accurately predict the collapse, in the orbit-crossing sense, of generic mass elements. Some points, undergoing orbit crossing, are assigned to the network of filaments and sheets that connects haloes; this network resembles closely that found in N-body simulations. The code generates a catalogue of dark matter haloes with known mass, position, velocity and merging history. The predictions of the code are very accurate when compared with the results of large N-body simulations that cover a range of cosmological models, box sizes and numerical resolution.

In this chapter, in section § 2.1, we present a brief discussion about N-body simulations and in section § 2.2 we present PINOCCHIO, its algorithm and its performances.

2.1 N-BODY SIMULATIONS

In a N-body simulation, the mass distribution is usually represented by particles distributed within a periodic box. The motion of each particle is computed numerically by taking into account its interactions with other particles. The number of particles in the most advanced simulations can reach several ten of billions. An example is the Millennium simulation, a picture of which is shown in figure 6, with eleven billion of particles.

The higher is the required resolution, the stronger are the accelerations, the smaller the time interval over which discretizing the orbits, the higher is the computational cost that includes computing time and memory requirements.
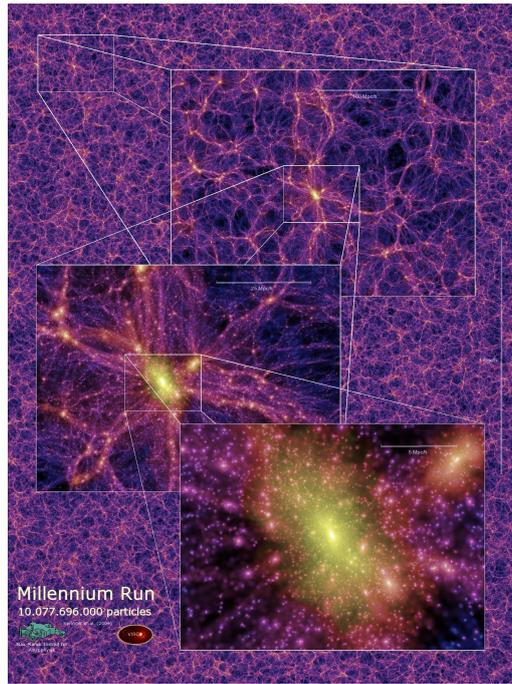
Figure 6: The Millennium Run used more than 10 billion particles to trace the evolution of the matter distribution in a cubic region of the Universe over 2 billion light-years on a side. Here is reported a projected density field for a 15Mpc/h thick slice of the redshift $z = 0$ output. (http://www.mpa garching.mpg.de).

This means that a compromise between accuracy and manageable computer requirements has to be chosen.

Special care must be employed to ensure both accuracy and efficiency, fundamental parameters to be chosen are:

- Size of the periodic box;

- Mass resolution: mass of each dark matter particle;

- Force resolution: the smallest scale down to which the gravitational force is correctly computed.

A number of numerical techniques are used to reproduce the dark matter distribution.

They differ, for the most part, only in the way the force of each particle is calculated.

The most common methods for N-body simulations are the direct summation (PP), the particle-mesh techniques (PM), its variant, the particle- particle-particle mesh techniques (P^{3}M) and the tree code.

The direct summation is the easiest technique and consists in computing the force acting on each particle, summing up the force from all other particles. In fact, it is also called particle- particle (PP) technique. The particles cannot be treated as point-mass because they will approach infinitesimally close one to the other. Because of this, it is common to use the so called softening parameters. This technique is quite accurate but the numerical solution of a N-body system,

for a fixed number of timesteps, scales as N^2. This because the evaluation of the force on each particle requires to take into account contributions from all other members of the system.

The particle mesh technique discretizes the force field, calculating it on a mesh with periodic boundary conditions, using an approach of average field. In practice after the grid is built, the value of the density field and of the gravitational potential are calculated on each grid point of it. The corresponding values on the particles positions are calculated by interpolation. Forces are calculated computing the gravitational potential on the grid, via the Fourier transform technique, to solve the Poisson equation (1.20). The advantage of the PM technique resides in the reduced number of operations that, thanks to the Fast Fourier Transform, are of the order of $\sim NlogN$ rather than N^2. The problem is in this case the resolution, that is limited by the mesh size. A variant of the PM is given by the particle-particle-particle mesh (P^3M): short range forces are evaluated by the PP and are added to the long range ones, evaluated by the PM. The accuracy is improved at the expenses of computational cost, which is anyway still lower than the one of the PP technique.

The tree code follows a completely different approach. It treats particles with a hierarchical system of partition. To calculate the force acting on the i-th particle, close particles, with short range forces, are treated in a PP-like method, while distant groups, with long range forces, are approximated by their lower multipole momenta, i.e they are treated as a single particles with mass equal to the total mass of the group, centered in its barycenter. The box of the simulation is iteratively divided until each cell, with more than one particle, contains at the end only one particle. This procedure permits to construct a tree structure where the nodes (leaves of the tree) are represented by the different partitions of the box. The communication between the nodes are the branches of the tree. The Tree code is usually faster than the PM or PM^3 for a comparable resolution. However one needs to store the whole structure of the tree, to run it efficiently, which means to increase the RAM memory requirement.

The arbitrary large dynamic range in the unsoftened dynamics and the expensive evaluation of the force have led to the development of a wide number of numerical and analytical-numerical techniques, aimed to obtain a reliable numerical solution with the minimum amount of computational resources.

In this context Pinocchio code has been thought and realized. The result is a powerful and innovative instrument for the study and the analysis of the properties of dark matter haloes, with cheap computational cost.

2.1.1 *Initial conditions*

The methods described until now calculate the evolution of the system at each timestep, but they all start from an initial condition that has to be specified. As presented in chapter 1, nowadays the standard cosmological model assumes that the Universe was initially homogeneous and isotropic with small over-densities that slowly are grown, giving rise to the structures we observe today. Let us consider the contrast density δ given by the equation (1.21). Let be $V = L^3$

the volume of the simulation. it is possible to write the contrast density in the Fourier series, in the limit $V \to \infty$ according to the equation (1.31) with periodic boundary conditions:

$$\begin{cases} \delta(0,y,z) = \delta(L,y,x) \\ \dot{\delta} = \dot{\delta}(L,y,z) \end{cases} \tag{2.1}$$

Let's consider the definition of the power spectrum given by equation (1.32). Let's assume that the growth of the cosmic structures starts from density perturbations with a gaussian distribution, as the CMB and large scale observations suggest. Then the power spectrum completely defines the fluctuations density field. Its form depends on cosmological parameters and on the nature of the dark matter. Once the power spectrum is known at the recombination, it is possible to generate the initial condition: to this purpose one generate a set of complex numbers with an equally distributed random phase and an amplitude with a variance given by the power spectrum desired (see equation (1.35)), i.e.:

$$\hat{\delta}_k = \sqrt{-2P(k)\ln(\alpha)}\exp(i2\pi\beta) \tag{2.2}$$

where α and β are two random numbers with uniform distribution $\in (0,1)$.

Now it is necessary to create the potential $\Phi^{(1)}(\mathbf{q})$ in a spatial grid, given by the Lagrangian coordinate $\mathbf{q}$, through the Fourier transform, to obtain the perturbation field given by the previous equation:

$$\Phi^{(1)}(\mathbf{q}) = \sum \frac{\hat{\delta}_k}{k^2}\exp(i\mathbf{q}\cdot\mathbf{k}) \tag{2.3}$$

Using this potential field initial positions and velocities are calculated with the Zel'dovich approximation according to the equations (1.54) and (1.55)

2.2 PINOCCHIO

PINOCCHIO, acronym for PINpointing Orbit Crossing Collapsed HIerarchical Objects, is a code developed first by (Monaco et al., 2002b,a; Taffoni, G. and Monaco, P. and Theuns, T., 2002).

It is able, with very limited computing resources, to generate catalogues of cosmological dark matter haloes with known mass, position, velocity, merger history from a realization of Gaussian initial (linear) density perturbation field, given on a 3D grid, reproducing with very good accuracy the hierarchical formation of dark matter distribution.

2.2.1 *The algorithm*

Pinocchio works in two fundamental steps:

1. ESTIMATES OF COLLAPSE TIMES. The instant at which a mass element undergoes collapse is identified, i.e. the orbit crossing(OC). This is computed numerically by applying the local ellipsoidal collapse approximation (see 1.3.2.2) to the full Lagrangian perturbation (Bond and Myers, 1996b,c; Monaco, 1995, 1997a,b).

2. FRAGMENTING THE COLLAPSED MEDIUM. The collapsed particles
   are grouped into disjoint haloes to mimic the hierarchical build up of halos
   trough accretion and merging.

The first step determines when a simulation particle enters in a high density
region whereas the second step identifies the haloes and the merger history.
Positions and velocities are calculated by the Lagrangian perturbation theory at
the first order.

This approach can be efficiently applied for generating input for galaxy for-
mation models. Moreover, it can be applied for generating simulated catalogues.

2.2.1.1 *Computing the Collapse time*

A Gaussian density field $\rho(\mathbf{q})$ is generated on a cubic grid of N^3 particles, in a
way much similar to that used to generate the initial conditions for an N-body
simulation, described in the section 2.1.1.

The density field represents the linear density contrast, i.e. the density contrast
at very early time t_i, linearly extrapolated to the present time, it is given by:

$$\delta_l = \frac{\delta(\mathbf{q}, t)}{D(t_i)} \tag{2.4}$$

where $\mathbf{q}$ denote Lagrangian (initial) coordinate of the particle and that corre-
sponds to the grid vertex, while $D^{(1)}(t)$ is the linear growing mode of perturba-
tion, normalized at $z = 0$.

Let $\phi_{(1)}(\mathbf{q})$ be the corresponding peculiar potential due to the density field.
Following the EPS approach both density and potential fields are smoothed by
convolving them with a Gaussian window of size R, that is in k-space:

$$\hat{W}(kR) = \exp\left(\frac{-k^2 R^2}{2}\right) \tag{2.5}$$

Smoothing radii are typically $\sim$ 20 logarithmically equi spaced, except for the
smallest smoothing radius, which is set to zero in order to recover the variance
at grid scale.

For each smoothing radius R, the first derivative of the potential $\nabla_q \phi^{(1)}(\mathbf{q})$
i.e. $p^{(1)}(\mathbf{q})$) and the second derivative, for the calculation of the deformation
tensor $\frac{dp_j^{(1)}(\mathbf{q})}{dq_i}$, are calculated by the Fast Fourier transform (FFT). They are
used to compute, respectively, the motion of the particle in the Zel'dovich (1970)
approximation and the evolution of the mass element based on ellipsoidal
collapse that we have treated in section 1.3.2, respetively.

The expected collapse time is calculated at each grid point, i.e. the time
at which a particle in the $\mathbf{q}$ position undergoes orbit crossing, t_{coll}. This is
done using the 3th order LPT truncation (see section 1.3.2.2) with an empirical
correction, made by Monaco (1997b) and Monaco (1997a), for the quasi spherical
case.

The growing mode $D^{(1)}(t)$ is used as time coordinate. To calculate the inverse
of the collapse time of each mass element $\mathbf{q}$ and for each smoothing radius R:

$$F(\mathbf{q}, R) \equiv 1/D(t_{coll}(\mathbf{q}, R)). \tag{2.6}$$

This way the dynamics of the gravitational collapse is almost independent of the background. The value of the F field at a single point $\mathbf{q}$ corresponds, in the excursion set approach, to the trajectories in the $F - R$ plane (or equivalently the $F - \sigma^2(R)$ plane) used to compute the mass function. If one considers the spherical collapse we would have $F = \frac{\delta}{\delta_c}$, having the absorbing barrier at δ_c. The mass function is a quantity obtained from the absorption rate of the $F(R)$ trajectories by a barrier put at a level F_c, given by the inverse of the time $(D^{(1)}(t))$ of the collapse.

Then the following quantity is recorded:

$$F_{max}(\mathbf{q}) \equiv max_R[F(\mathbf{q}; R)]. \tag{2.7}$$

that is the highest value of F along the trajectory and it is interpreted as the inverse of the time $(D^{(1)}(t))$ at which the particle is expected to go into orbit crossing.

Together with each point, the smoothing radius R_{max} at which $F = F_{max}$ is stored, and the corresponding Zel'dovich velocity v_{max} computed at the time $D(t) = 1/F_{R_{max}}$ appropriate for the smoothing radius R_{max}, necessary for building of DM haloes.

The collapsed mass element will not necessarily accrete onto any halo, but may instead become part of a filament or sheet (collectively referred to as filaments, since these have undergone OC as well, these structures trace the moderate over densities that connect collapsed haloes in simulations (Monaco et al., 2002b,a; Taffoni, G. and Monaco, P. and Theuns, T., 2002).

The next sections describe how the OC collapsed medium is divided into collapsed haloes and filaments according to the criteria of merging and accretion.

2.2.1.2 *Fragmentation*

The hierarchical formation of objects is followed through the grouping of collapsed particles into haloes by tracing the merging and accretion process for each particle individually. First the particles are sorted according to decreasing collapse redshift z_{coll}, and, starting from the highest z_{coll} it is decided the fate of the collapsed particle, working forward in time to the last particle to collapse.

For each particle on the initial grid position $\mathbf{q}$ the six nearest particles are called Lagrangian neighbors. Two are the main processes in the hierarchical formation of objects: the accretion of a particle onto a halo and the merging of two haloes. In figure 7 are shown the cases of the fragmentation process.

The following criteria decide the fate of particles. Let $R_N = N^{1/3}$ be defined as the Lagrangian radius of a halo of N particles, where the length are in units of the grid spacing.

- Seed haloes: the local maxima of the collapsed redshift, z_c, are seeds for a new halo: all six neighburs have not yet collapsed, the particle is then a peak of F_{max} and is treated as a new halo with one particle;

- Accretion: if a collapsing particle is not a local density maximum at the collapse redshift, it is checked whether any of its neighbors has already

been assigned to a halo. If the point is the neighbour of a single halo, then the code computs from its Zel'dovich velocity, the position of the halo, the mass at the collapse redshift, and their mutual distance. If the mass element has a distance d from the halo center of mass at the collapse time, such as :

$$d < f_a \times R_N + f_{ra} + \delta d \qquad (2.8)$$

then the accretion criteria is satisfied and the particle accretes onto that halo. f_a, f_{ra} and δd are parameters that we discuss later. If the particle is able to accrete onto two (or more) haloes, we assign it to the one for which d/R_N is the smallest. This criterion selects haloes with roughly fixed overdensity.

- Merging: if some of the neighbors of a collapsing particle belong to different haloes, then these haloes merge if their mutual distance d, again at the particle collapsed time, is smaller than a fraction of the largest of their typical radii R_N, i.e

$$d < f_m \times \max(R_{N1}, R_{N2}) + f_{rm}. \qquad (2.9)$$

Since PINOCCHIO considers only six Lagrangian neighbors, up to six haloes may merge at a given time, although binary and ternary mergers are more frequent. Moreover, the previous condition is used to check whether the particle is to be accreted on one of the groups. f_m and f_{rm} are other two parameters that we discuss later.

- Filaments: collapsed mass elements which are not local maxima and do not accrete onto a halo are assigned to a filament group, these particles tend in fact to occur in the mildly overdense regions that connect the haloes (visible as filamentary network between haloes in simulations). Note that filaments have undergone orbit-crossing, although they do not belong to relaxed haloes. In N-body simulations, some particles accrete into a halo directly from this filamentary network. PINOCCHIO accretes all filament neighbours of an accreting particle.

In figure 8 it is reported a summary scheme with the main tree block of the algorithm.

We have seen in eqs. (2.8), and (2.9) that there are five free parameters.

f_a is a parameter of order unity, analogous to the linking length parameter used to identify (friend-of-friend) haloes, an algorithm used in N-body simulation, where the linking length is 0.2 times the mean interparticle distance at $z = 0$. It determines the normalization of the mass function.

The f_m has similar function of f_a but controls the merging condition and therefore the overall slope of the mass function.

The f_{ra} and f_{rm}, are introduced in the accretion condition and in the merger one respectively, to avoid the production of few small haloes at high redshifts, a resolution effect seen because of the limited accuracy of the Zel'dovich approximation, when groups are very small and R_N is of the order of grid spacing. The
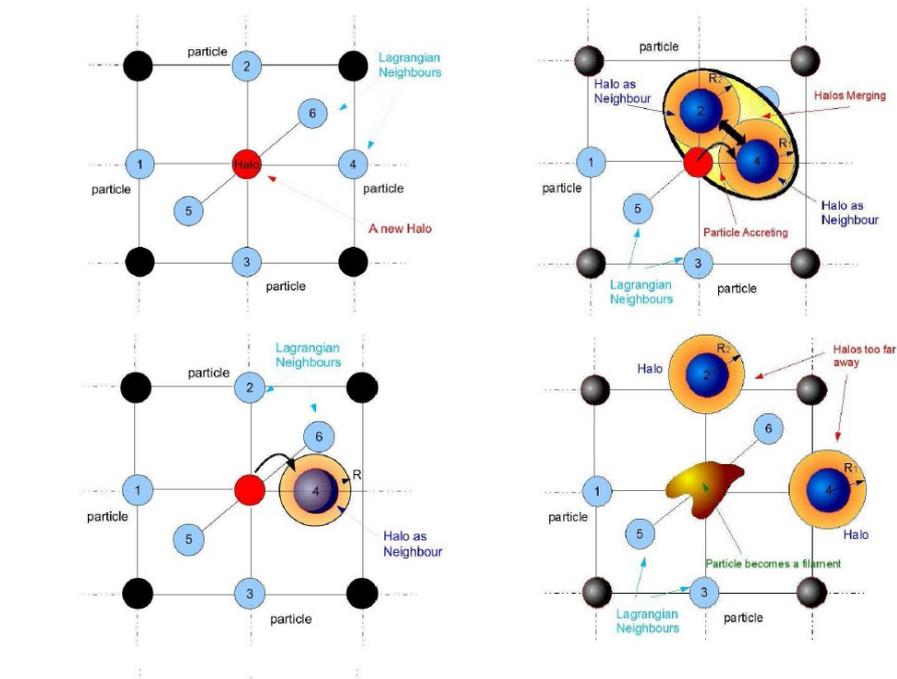
Figure 7: Cases of fragmentation process: The top panel on the left shows the six Lagrangian neighburs of a given particle, the bottom panel on the left illustrates how this particle accres onto a neighbouring halo, the top panel on the right depicts the merging of two haloes and a succesive accretion, if there is no accretion the particle is marked as belonging to a filament in the bottom panel on the right. (Heisenberg et al., 2011).

f_{rm} therefore controls the slope at low masses while f_{ra} the abundance of low mass haloes.

The f_s is a free parameter used to avoid the tendency of objects to accrete mass at higher redshifts. It controls the abundance of low mass haloes at low redshifts. It appears in the δd term given by:

$$\delta d = f_s (\sigma(R_N))^{1.7} R_N. \tag{2.10}$$

where R_N is given in grid units. This quantity is the error in reconstructing positions that gives the uncertainty on d, i.e. the distance between a collapsing particle and the center of mass of a group, $\sigma(R_N)$ is the variance calculated on the Lagrangian radius of the group. These corrections make all the parameters sufficiently generous so that the particle falls within d at late time, although the accuracy of LPT (Zel'dovich approximation) in estimating the velocities depends on the degree of non-linearity reached and it becomes worst at later time.

These parameters are explained in detail in the papers of Monaco et al. (2002b) and Monaco et al. (2002a).

The fragmentation code permits to have information of the merger and accretion history at each time. Haloes are in fact always updated when a collapsing particle touches them.

At each merger the largest halo retains its identification number (ID) that is the ID of the expired halo. The mass of each halo involved in the merging event is recorded together with the redshift at which the merger takes place.

Even though accretion is rigorously defined as the entrance of a single particle into the object, the merger of a halo with another one with less than 10 particles is always considered as an accretion event.

The merger trees extracted from PINOCCHIO provide a more complete description of the merging histories of haloes then the extended Press and Shechter (EPS) model. They not only follow the time evolution of the mass and number distribution of the progenitors, but also their distribution in space and their velocities.

The full catalogue of dark matter haloes, at each time, is provided and includes for each halo, mass, center of mass in Lagrangian space, displacements and peculiar velocities (Zel'dovich approximation).

2.2.2 *The code*

The first version of PINOCCHIO was written by Monaco et al. (2002a) in Fortran 77 and it was a serial code. It was created to carry out a run on a simple personal computer. Its great innovation lies in its low computer requirements. The code runs very fast compared to simulations. It is able to run a realization of 256^3 particles on a 512 Mbyte machine in slightly more than 6 hours. The first stage for the collapse time calculation needs about 6 hours while the fragmentation only few minutes. This means that it is possible to obtain reasonable merger histories without the need of a supercomputer. Because the RAM is usually the limiting factor, the PINOCCHIO code has been built to keep a minimum amount of information in the RAM and swap the rest on the disk.
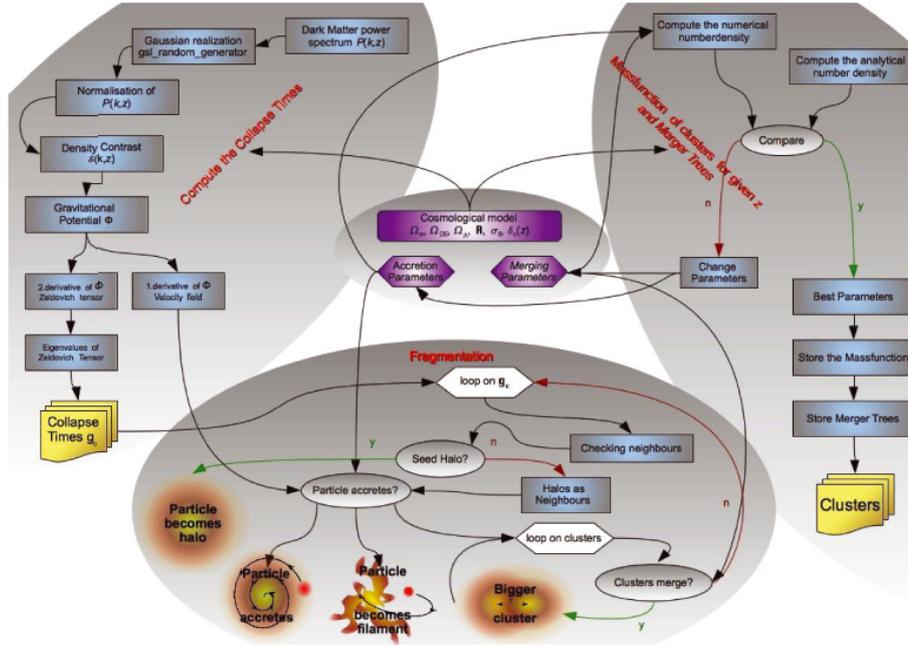
Figure 8: The main tree blocks of the algorithm are shown. For a given set of parameters (center) Pinocchio computes the collapse time (top left) and for each collapsing particle the fragmentation procedure is applied (bottom). Finally the statistical properties of the fragmented objects are analysed (top right) (Heisenberg et al., 2011).

The second version of PINOCCHIO was later developed. This was written in parallel in Fortran 90. An innovation of the code has been the use of the FFTW package for the computation of the Fast Fourier Transform. The FFTW takes care of most communications for the parallelization of the F_{max} computation. The fragmentation is so quick that between the serial and parallel version of it there are only small differences. Even in this version the memory is swapped to the local disk.

The most recent version, the third one, permits to run PINOCCHIO on a large number of cores and with limited disc access. In this version the memory requirement increases by a factor of three, since it is all kept in the RAM. The collapsing time is calculated as in the previous versions, while some changes have been made to the fragmentation part of the code. The box is divided in sub-volumes, and each of them is assigned to a task. The aim is to reduce the number of communications among the tasks during the process. Moreover to proceed in this way, each sub-volume needs to extend the calculation to a boundary layer. This method permits a stable reconstruction of the largest objects with a layer of 30 Mpc, that requires few of 10s per cent for large boxes but for small boxes with high resolution the layer requires significant overhead.

As we reported, FFTW allocates memory to tasks in planes, while the fragmentation code operates with sub-boxes. The communication between the two parts of the code, for the redistribution of the F_{max} and velocities follows a scheme for which tasks communicate in pairs.

The code for the generation of the initial linear density field in the Fourier space is obtained by linking the PINOCCHIO code with the N GenIC code by V.Springel (http://www.mpa garching.mpg.de/gadget/).

The relevant improvement therefore resides in the possibility of faithfully reproducing a simulation, once one has the knowledge of the cosmology and the random number seed.

The code considers different cosmologies: open universe, flat universe and Einstein De Sitter Universe.

Displacements and the final positions of the groups are calculated using the Zel'dovich approximation, as described in section 1.3.1.1

2.2.3 *PINOCCHIO vs N-body simulations*

To prove the validiy of PINOCCHIO we report in this section the results obtained by the comparison of PINOCCHIO data with the ones of N-body simulations (Monaco et al., 2002a).

We report here two of these simulations with different cosmology, the SCDM simulation ($\Omega_m = 1.0$, $\Omega_\Lambda = 0.0$, $\sigma_8 = 1$, $h = 0.5$) run with the PKDGRAV code, with box size of 500 h^{-1} Mpc and with 360^3 particles (Governato et al., 1999) and a ΛCDM simulation ($\Omega_\Lambda = 0.7$, $\Omega_m = 0.3$, $\sigma_8 = 0.9$, $h = 0.65$) with a smaller box of 100 h^{-1} Mpc and with 256^3 particles, evolved with the P^3M HYDRA code (Couchman, 1991).

2.2.3.1 *Mass function*

We describe here the comparison of PINOCCHIO and FOF (used in the numerical simulations) mass functions for the SCDM and the ΛCDM simulations. The results are shown in figure 9: on the left it is reported the comparison of the mass function obtained by PINOCCHIO, using three of the five parameters (see 2.2.1.2 and for more details see Monaco et al. (2002b) vs the one obtained by the N-body simulation for SCDM, while on the right it is reported mass function obtained by PINOCCHIO vs the one obtained the ΛCDM simulation, using the same combination of parameters; the results with the full five parameters are very similar. For reference, it has been also plotted the PS and Sheth & Tormen (ST) mass functions. The choice of the five parameters explained in section 2.2.1.2 produces a PINOCCHIO mass function which falls to within ~5 per cent of the simulated one from $z = 5$ to $z = 0$, for all mass bins with more than 30-50 particles per halo and for which the Poisson error bars are small. The only residual systematic is a modest ~10-20 per cent underestimate at the highest-mass bins and highest redshift. Because PINOCCHIO is calculated for the same initial conditions as the simulation, Poisson error bars are not the correct errors to use for this comparison (notice that the Poisson error bars of the PINOCCHIO mass function are obviously very similar to those of the numerical one). They are shown both for comparison with PS and ST and to understand which mass bins are affected by small number statistics.
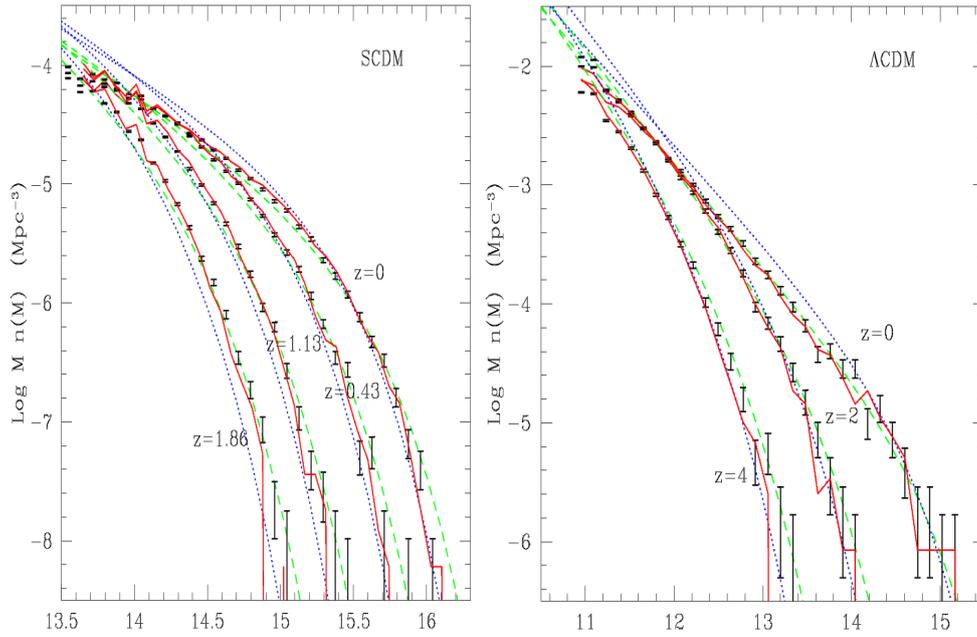
Figure 9: On the left: the top panel have simulated mass function for FOF selected haloes (Full lines with Poissonian error bars), PINOCCHIO mass function (filled circles), the fit by Sheth and Tormen (short-dashed lines) and PS function (long-dashed lines), at redshifts $z = 0$, 0.43, 1.13 and 1.86 (higher redshift curves are off-set by 0.1 dex both vertically and horizontally for improved clarity). Vertical lines show limits corresponding to simulation haloes with 10, 50, 100, 500 and 1000 particles (256^3 re-sampling); bottom panel have difference between simulated mass function and PINOCCHIO (filled dots), Sheth and Tormen fit (short-dashed line) and PS (long-dashed line) at $z = 0$ (Monaco et al., 2002b). On the right the mass function for the ΛCDM model at different redshift at different redshifts $z = 0$, 2, 4. continuous lines are the PINOCCHIO predictions, and dotted and dashed lines are the PS and ST predictions, respectively (Monaco et al., 2002a)
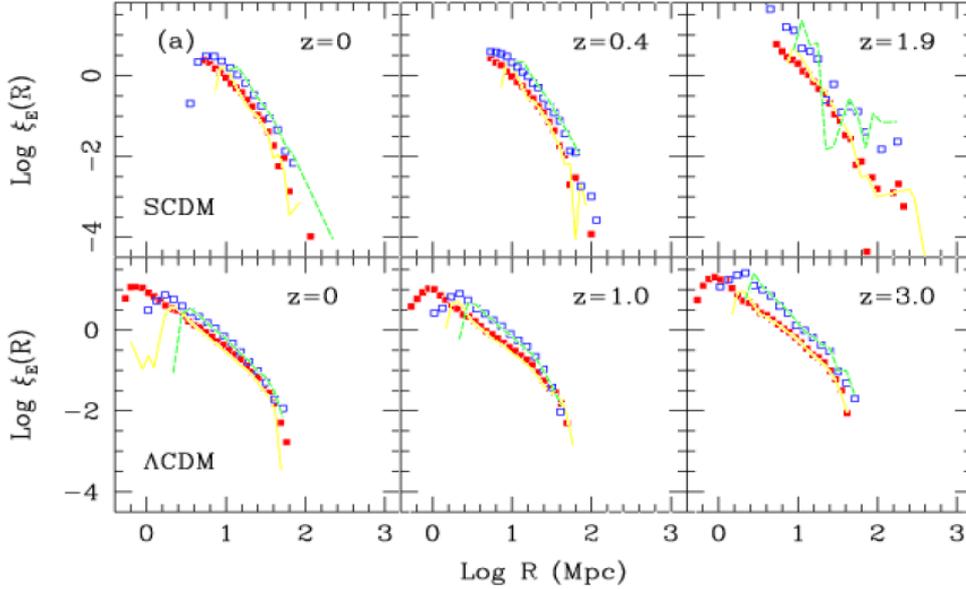
Figure 10: Eulerian correlation function for the SCDM and ΛCDM models at three redshifts indicated in the panel, and for two mass ranges. Symbols refer to simulation results, lines to PINOCCHIO predictions. Filled squares and continuous lines: correlation function for low mass haloes (mass M from 6.3×10^{11} to 3×10^{12} M$_\odot$), open squares and dashed lines: correlation function for massive haloes (M $> 3 \times 10^{12}$ M$_\odot$) (Monaco et al., 2002a).

2.2.3.2 *The 2-point correlation function*

We show in figure 10 the correlation function of haloes as a function of mass, in Eulerian space. The correlation function was computed using a standard pair counting algorithm. The agreement between PINOCCHIO and the simulations is very good down to ~1-2 comoving Mpc/h. The differences are of the order of ~10-20 per cent in amplitude and $\lesssim$10 per cent in terms of scale at which a fixed amplitude is reached. This means that both the correlation length r_0, at which $\xi(r_0) = 1$, and the length at which $\xi = 0$ are reproduced with an accuracy of better than 10 per cent. More important, the trends of increased correlation for the more massive haloes, or for haloes of a given mass with increasing redshift, are both well reproduced.

2.2.3.3 *Progenitors*

PINOCCHIO follows the merging of haloes in real time, and links each progenitor to its parent after each merging event. In N-body simulations haloes are identified at the end of the run, so it is necessary to analyze and cross-correlate a large number of outputs to follow the merger histories. In other words, the generation of the merger trees is by far less expensive (in term of CPU time and disk space) in PINOCCHIO than in a simulation; in fact PINOCCHIO automatically computes the merging history of haloes and does not need any further analysis.
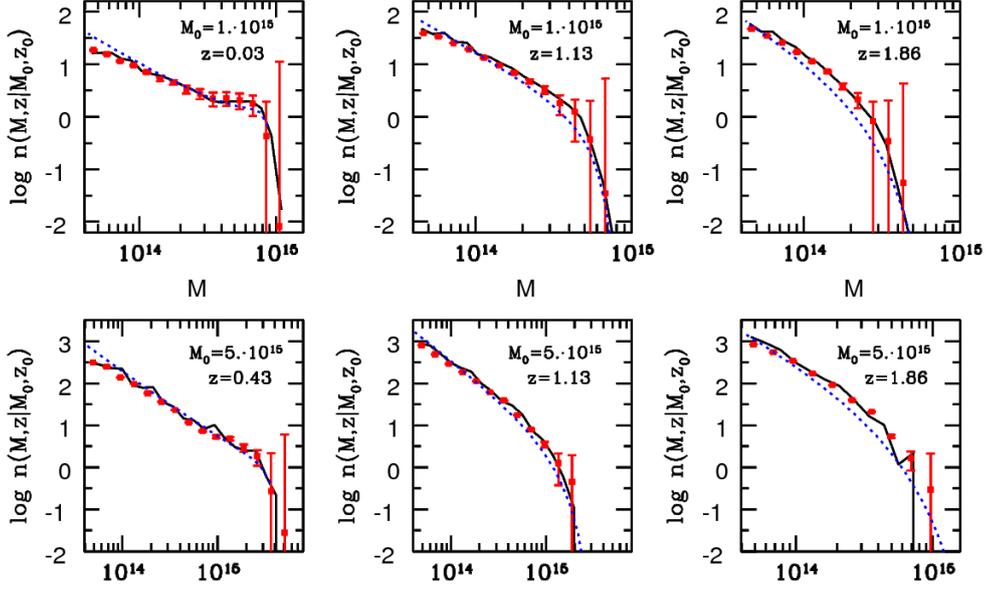
Figure 11: Conditional mass functions in the ΛCDM case for parent haloes identified at $z = 0$. The mass threshold is fixed at $M_{th} = 7.6 \times 10^{10}$ $M_\odot$ (10 particles), the increases from left to right and covers the values: $z = 0.43$ 1.13, 1.86. The mass of the parent halo increases from top to bottom, the adopted values are: $M_0 = 1. \times 10^{15} M_\odot$ and $5. \times 10^{15} M_\odot$. The points represent the simulation data while the solid lines are the prediction of PINOCCHIO; the dashed lines are the analytical predictions of the EPS formalism. (Taffoni, G. and Monaco, P. and Theuns, T., 2002).

We report in figure 11 the comparison of the mass function of the progenitors, i.e. the number density of progenitors of mass M at redshift z that merge to form the parent M_0 at redshift z_0 given by PINOCCHIO and by the simulations. Generally, at least 30 particles are necessary to identify reliable haloes both in the simulation and in PINOCCHIO, so this is the threshold mass considered for the statistical analysis.

For the build up of merger trees of the N-body simulations the following two rules have been used:

- if a parent halo contains less than 90 per cent of the mass of all its progenitors at redshift z, then it is excluded from the analysis (this happens in a few percent of cases);

- we assign to the progenitor the mass of all its particles that will flow in to the parent at z_0.

The PINOCCHIO conditional mass function and that obtained from the simulations are computed by averaging over a mass interval around $\log M_0$ of 0.01 dex.

The conditional mass function predicted using PINOCCHIO shows a very good agreement when compared with the simulations. In figure 11 we show that the PINOCCHIO prediction fits the simulations data with similar accuracy for all the considered parent mass and redshifts, with a discrepancy between

the two distribution which in general is less than 25 per cent. This means that PINOCCHIO reproduces the conditional mass function with better accuracy that the EPS prediction, that is no able to fit numerical simulations.

2.2.3.4 *Object-by-object comparison*

Agreement at the point-by-point level requires that each particle is predicted to reside in the correct halo with the correct mass. The improvement of PINOC-CHIO in the point-by-point comparison with respect to other methods is not primarily due to the more accurate dynamical description of collapse, rather it is due to the much more accurate description of the shape of the collapsing region.

To make this analysis for each object of one catalogue, the objects of the other catalogue that overlap for at least 30 per cent of the Lagrangian volume have been considered. Two haloes from different catalogues are cleanly assigned to each other, when one overlaps wih the other more than with any other halo. The fraction of haloes cleanly assigned is f_{cl}, which the fraction of halos not assigned is f_{split}, the remainder $1 - f_{cl} - f_{split}$ is the fraction of objects of one catalogue that do not overlap with any halo in the other catalogue. These fractions quantify the level to which two catalogues describe the same set of haloes. f_{ov} is defined as the average fraction that haloes overlap when they are cleanly assigned. All these estimators depend on whether PINOCCHIO is compared with simulations or vice-versa, but in general that difference is small as long as the agreement is good.

To prove the agreement of PINOCCHIO with the simulations at the object-by-object level we show in figure 12 the the comparison of PINOCCHIO with the SCDM simulation (left panel) and with the ΛCDM one (right panel).

In both cases the agreement is very good at higher redshift with ~80-90 per cent of objects cleanly assigned when the haloes have at least 50 particles. The degree of splitting is only $\lesssim 5$ per cent, while the average overlap of cleanly-assigned objects f_{ov} ranges from 60 per cent to 70 per cent nearly independent of mass and encouragingly larger than the 30 per cent lower limit. The agreement is slightly worse at lower redshift, with $f_{cl} \gtrsim 70$ per cent for haloes with at least 100 particles, and f_{split} ~5–10 per cent. Within LPT there is obviously no advantage in going to higher resolution, as the accuracy of LPT worsens with the degree of non-linearity and with it all the results. Anyway, the agreement is still very significant for the last output, with a high fraction of cleanly assigned objects and a modest degree of splitting. In any case the results always improve with increasing number of particles.

2.2.3.5 *Accuracy of reconstruction*

PINOCCHIO is able to accurately recontruct mass, positions and velocities of haloes. In figure 13 we show the accuracy with which PINOCCHIO is able to estimate mass, Eulerian position and velocity of the cleanly assigned objects. In particular, we show for ΛCDM the scatter plots of the masses, and of velocity and position along one coordinate axis. For comparison, the scatter plot of the displacements of FOF haloes from the initial to the final positions are
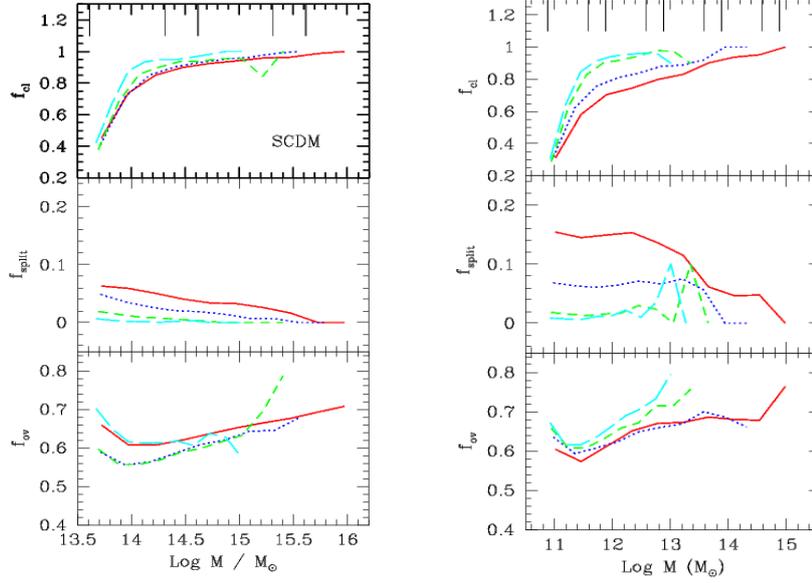
Figure 12: On the left comparison on an object-by-object level of haloes identified by PINOCCHIO and found in the SCDM simulation, using a variety of statistics. Continuous, dotted, short-dashed and long-dashed lines refer respectively to redshifts $z = 0$, 0.43, 1.13 and 1.86. Upper panel:fraction f_{cl} of cleanly assigned halo pairs between the two catalogs, as a function of mass; middle panel: fraction f_{split} of FOF haloes that are split in two PINOCCHIO haloes; lower panel: average overlap in Lagrangian space, f_{ov}, for cleanly paired-up haloes. Vertical lines show limits corresponding to simulation haloes with 10, 50, 100, 500 and 1000 particles (256^3 re-sampling). On the right comparison on an object-by-object level of haloes identified by PINOCCHIO and found in the ΛCDM simulation. Continuous, dotted, short-dashed and long-dashed lines refer respectively to redshifts $z = 0$, 1, 2 and 4. Top panel: fraction f_{cl} of cleanly assigned objects; middle panel: fraction f_{split} of non-cleanly assigned objects; bottom panel: average overlap f_{ov} for cleanly assigned objects. The vertical lines in the top panel indicate haloes with 10, 100, 10^3, 10^4 and 10^5 particles (heavy lines) or 50, 500, 5×10^3 and 5×10^4 particles (light lines). (Monaco et al., 2002b,a).
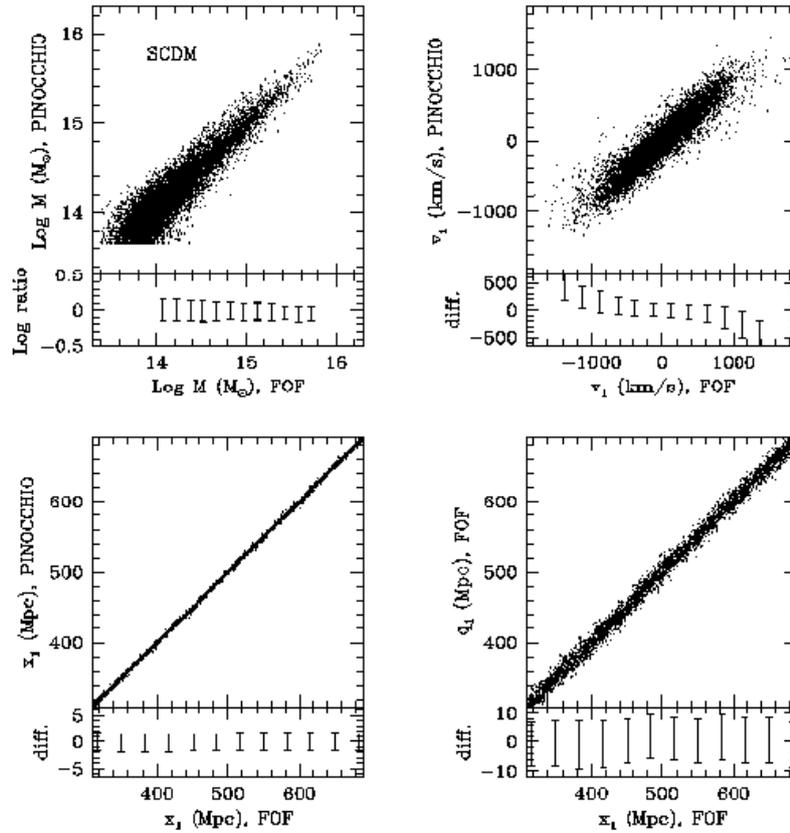
Figure 13: Difference in mass, position and velocity, $\log M$, x and v respectively, as estimated by PINOCCHIO and found from the simulation of ΛCDM, for cleanly assigned haloes. The scatter around the mean is plotted below each panel. The lower right panels show for comparison the displacement of haloes according to the simulation (Monaco et al., 2002a).

shown as well. Masses are recovered with an accuracy of ~40 per cent ΛCDM, nearly independent of mass. The average value is slightly biased. Positions are recovered with a 1D accuracy of ~1 Mpc, slightly depending on the box size and much smaller than the typical displacements, while velocities are recovered with a 1D accuracy of ~150 or 100 km/s. In general, the velocities of the fastest moving haloes are underestimated. This could be fixed by extending the calculation of velocities to second or third order LPT, that will be part of this work.

2.2.3.6 *Performance and scaling*

The hard and the soft scaling properties of the third version of the code, have been tested running the code on the PLX machine at the Centro Interuniversitario del Nord Est per il CALcolo (CINECA). In the hard test the number of particles for each processor are fixed and the number of processor is increased while in
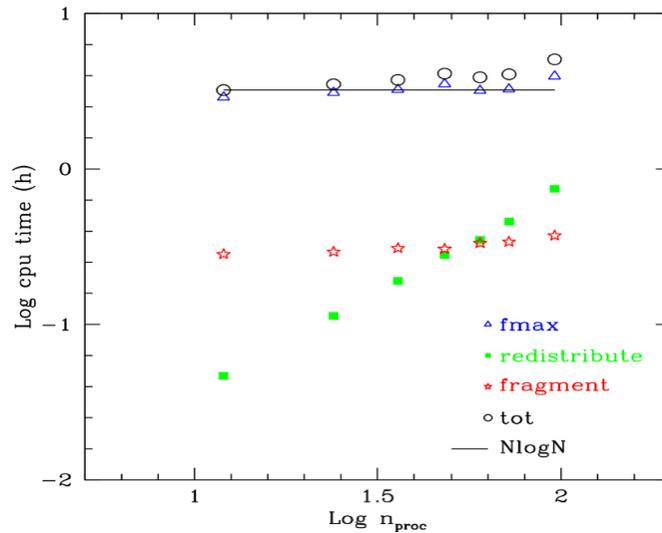
Figure 14: The picture shows the time required for PINOCCHIO for each stage of the run in function of the number of processors used: the triangle represents the time required fot the collapse time calculation, the squares for the particles sorting, the stars for fragmentation and finally the circles represent the total time required (Monaco workshop Trieste (2012) http://adlibitum.oats.inaf.it). The data comes from a run made with a 720^3 grid of size, 720Mpc/h box size, on PLX@cineca

the soft test, the number of processor is costant while the number of particles is proportional to the number of the processors.

We report in figure 14 a hard scaling test, i.e. the time required for each step of PINOCCHIO as function of the number of processor. The realization is obtained considering 720^3 particles, for a commoving 720Mpc/h box size length on 1 to 8 nodes, with each node including 12 cores with 48 Gb of RAM. Here the time needed to write the results on the disk is not considered. The triangle represents the time required for the collapse time calculation, the squares for the particles sorting, the stars for the fragmentation and finally the circles represent the total time required to complete the run. The horizontal line gives the ideal behavior of the expected scaling, which is constant because the number of grid points are fixed. The computation of collapsing time and of the fragmentation, despite the presence of the boundary belt, follow very closed the behaviour of the ideal case, thanks to the FFTW libraries.

In figure 15 we show the same quantities, but here the number of particles is increased, at fixed mass resolution, proportionally to the number of cores used, up to 1440^3 on 8 nodes. The solid line, as before, is the $Nlog_2N$ ideal total time required. Even in this case, the computation of collapsing time and of the fragmentation, although the boundary belt, follow very closed the behaviour of the ideal case.

To give another exsample of the speed of PINOCCHIO, a realization of 2048^3 particles run on 300 cores of the PLX machine requires only ~ 30 minutes: the collapse time calculation employs 65 per cent of time (40 per cent required for
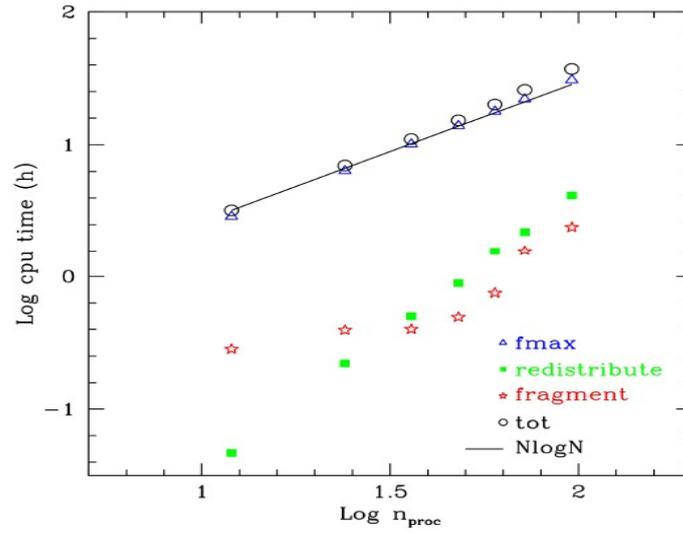
Figure 15: The picture shows the time required for PINOCCHIO for each stage of the same run described in figure 14 so the symbol represent the same quantities, but here the number of particles is scaled with the number processors (Monaco workshop Trieste (2012) http://adlibitum.oats.inaf.it).

the FFT), while the redistribution takes 20 per cent and fragmentation 15 per cent.

IMPROVEMENT ON THE CODE

3

The first part of this work is an approach to the numerical calculation. It is focus on the improvement of the PINOCCHIO code. The aim is to translate each algorithm that calculates the analytic solution of a specific cosmological quantity, with the correspondent algorithm that performs the calculation numerically. In fact some cosmological equations, such as the equation of evolution of the linear growing mode at first and second order, have no analytical solutions except for particular cases, instead numerical solutions provide the informations for the general cases. There are several numerical algorithms studied for different requests of the problem. It's a good practice to test the accuracy of numerical solutions by reproducing the analytical ones. Once the requested accuracy is achieved, the numerical calculation of the cosmological quantities allows the code to be more elastic and complete.

We have implemented the algorithm for the numerical solution of the first order equations of the growing mode and of the cosmic time that PINOCCHIO calculated using the analytical solutions. Later, we have used the same algorithm to implement the solution of the growing mode at the second order that previously was not implemented in PINOCCHIO. Then we have performed all the necessary comparisons between the analytic and the numerical results to check if the requested accuracy is achieved.

In this way we can evaluate the evolution of perturbations of the dark matter component with the same accuracy given by the analytical results, and in addition we can perform it for more general cosmological models even including the contribution of the dark energy equation of state. Moreover with the second order of the growth it is possible to improve the accuracy in the evaluation of the positions and velocities of dark matter halos, calculated by PINOCCHIO with the Zeldovich approximation (first order LPT).

In section § 3.1 we present the routines that PINOCCHIO uses for the calculation of the cosmological quantities, included the first order of the growth. In section § 3.2 we describe the method followed for the numerical calculations and in section § 3.3 we show the result of the comparison between the numerical.

3.1 COSMOLOGICAL ROUTINES OF PINOCCHIO

PINOCCHIO contains one file for the calculation of the cosmological quantities, included the growing mode, *cosmo.F90*, and a module, *pinocchio-common.F90* that contains the parameters needed to run the code, including the cosmological parameters, the basic properties of the (highest-resolution) box, the run name and so on.

More in detail, *cosmo.F90* calculates the following cosmological quantities we are interested in:

- the growth first order as a function of z, $D^{(1)}$, and its inverse function, $D^{(1)-1}$;

- the scale factor as a function of time, $a(t)$;

- the paramenter density as a function of the redshift, $\Omega_i(z)$;

- the hubble parameter as a function of redshift, $H(z)$;

- the cosmic time as a function of redshift, $t(z)$;

- Peebles' $f(\Omega)$ function, $(f(\Omega) = dlog(D^{(1)})/dloga)$.

We will first focus our attention on $D^{(1)}$, $D^{(2)}$ and $t(z)$. We have already presented in the first chapter the equations that describe the time evolution of the growth at the first order (equation (1.53)), at the second order (equation (1.62)) and the evolution of the cosmic time (equation (1.3)). We write them as a function of the scale factor, introducing the following change of variables. Considering $H = \frac{\dot{a}}{a} = \frac{1}{a}\frac{da}{dt}$, it follows:

$$\begin{cases} \frac{d}{dt} & = aH\frac{d}{da} \\ & = \dot{a}\frac{d}{da} \\ \frac{d^2}{dt^2} & = \frac{d}{dt}\left(\dot{a}\frac{d}{da}\right) \\ & = \ddot{a}\frac{d}{da} + \dot{a}^2\frac{d^2}{da^2} \end{cases} \tag{3.1}$$

therefore we have:

$$\begin{cases} \ddot{D}^{(1)} + (\frac{3}{a} + \frac{1}{E(a)}\frac{dE(a)}{da})\dot{D}^{(1)} - \frac{3\Omega_m}{2E(a)^2a^5}D^{(1)} = 0 \\ \ddot{D}^{(2)} + (\frac{3}{a} + \frac{1}{E(a)}\frac{dE(a)}{da})\dot{D}^{(2)} - \frac{3\Omega_m}{2E(a)^2a^5}D^{(2)} = -\frac{3\Omega_m}{2E(a)^2a^5}(D^{(1)})^2 \\ \dot{t} = 1/(E(a)H_0 a) \end{cases} \tag{3.2}$$

We write as well the quantities we calculated in order to the above equations, i.e. equation (1.11) for the evaluation of $E^2(z)$ and the integral of the equation of state of for the dark energy expressed in equation (1.12), respectively:

$$\begin{cases} H^2 = E^2(a)H_0^2 = \Omega_{ro}\left(\frac{a_o}{a}\right)^4 + \Omega_{mo}\left(\frac{a_o}{a}\right)^3 + \Omega_{ko}\left(\frac{a_o}{a}\right)^2 + \Omega_{DE}\left(\frac{a_o}{a}\right)^3 \exp\left(-3\int_1^a \frac{dP}{d\rho}\frac{da'}{a'}\right) \\ \frac{dP}{d\rho} = w + w_1(1-a) \end{cases}$$

$$\tag{3.3}$$

where we consider $\Omega_{ro} = 0$ because it is negligible in our case.

The previous version of *cosmo.F90* calculated the first order of the growth and of the cosmic time with the following equations, that are the exact solutions for particular cosmological models. From (1.53) and (1.15):

- Einstein De Sitter model ($\Omega_{mo} = 1$, $\Omega_{\Lambda o} = 0$):

$$D^{(1)}(a) = a(t) = D^{(1)}(z) = 1/(1+z), \tag{3.4}$$

$$t(z) = 2/3 \frac{t(z_o)}{(1+z)^{3/2}} \tag{3.5}$$

- Flat universe ($\Omega_{mo} < 1$, $\Omega_{\Lambda o} \neq 0$):

$$D_{(1)}(z) = h \int_h^\infty \frac{1}{x^2(x^2-1)^{1/3}} dx \tag{3.6}$$

where $h = \coth(\frac{3}{2}t(z)\sqrt{\Omega_{\Lambda o}})$;

$$t(z) = \frac{2}{3} \frac{t(z_0)}{\sqrt{\Omega_{\Lambda o}}} \sinh^{-1}\left(\frac{\sqrt{\frac{\Omega_{\Lambda o}}{\Omega_{mo}}}}{(1+z)^{3/2}}\right) \tag{3.7}$$

- Open Universe ($\Omega_{mo} < 1$, $\Omega_{\Lambda o} = 0$):

$$D^{(1)}(z) = \frac{5}{2(\Omega_0^{-1}-1)}\left(1 + (1+3(\tau^2-1))\left(1+\tau/2\log\left(\frac{\tau-1}{\tau+1}\right)\right)\right) \tag{3.8}$$

where τ is $\sqrt{\frac{(1+z)}{(\Omega_0^{-1}-1)}+1}$,

$$t(z) = \frac{1}{2t(z_o)}\Omega_{mo}(1-\Omega_{mo})^{-3/2}(\cosh^{-1}(\frac{z\Omega_{mo}-z\Omega_{mo}+2}{z\Omega_{mo}+\Omega_{mo}})$$
$$+2\frac{\sqrt{(1-\Omega_{mo})(z\Omega_{mo}+1)}}{z\Omega_{mo}/(1+z)}. \tag{3.9}$$

For each cosmological model treated above $H^2(z)$ (or $H^2(a)$) and the dark energy equation of state have the following expressions:

-
$$\begin{cases} H^2 = E^2(a)H_0^2 = \left(\frac{a_o}{a}\right)^3 \\ \frac{dP}{d\rho} = w_0 + w_1(1-a) = -1 . \end{cases} \tag{3.10}$$

-
$$\begin{cases} H^2 = E^2(a)H_0^2 = \Omega_{mo}\left(\frac{a_o}{a}\right)^3 + \Omega_{ko}\left(\frac{a_o}{a}\right)^2 + \Omega_{\Lambda o} \\ \frac{dP}{d\rho} = w + w_1(1-a) = -1; \end{cases} \tag{3.11}$$

-

$$\begin{cases} H^2 = E^2(a)H_0^2 = \Omega_{mo}\left(\frac{a_o}{a}\right)^3 \\ \frac{dP}{d\rho} = w + w_1(1-a) = -1; \end{cases} \tag{3.12}$$

The growth is normalized to 1, i.e at the value it assumes at $z = 0$, a_0 is taken equal to unit.

We have studied the analytical solution for the growth at the second order for the only cosmological model that admits it, i.e the EDS model. Considering that in this cosmology $D^{(1)}(a) = a(t) = t^{2/3}$ (see equation (1.15)), $E(a)^2 = 1/a^3 = 1/t^2$, it is easy to demonstrate that the solution of the equation (1.62) is:

$$D^{(2)}(a) = -3/7a^2 . \tag{3.13}$$

We consider also the approximation explained in section 1.3.1.2 provided by Bouchet et al.(1995-1996) (Bouchet (1996)) for a ΛCDM universe:

$$D^{(2)}(t) \simeq -3/7(D^{(1)}(t))^2\Omega_m^{1/143} . \tag{3.14}$$

.

3.2 NUMERICAL CALCULATION

In order to find the numerical solutions for our quantities, we have to solve a problem involving ordinary differential equations (ODEs) and numerical integration of functions.

3.2.1 *Differential equations*

We have differential equations of the second order for both the growth, and the first order equation of the cosmic time.

Problems involving ordinary differential equations (ODEs) can always be reduced to the study of sets of first-order differential equations, with a change of variable. Usually one defines the new variables to be one the derivative of the original variables. The generic problem in ordinary differential equations is thus reduced to a set of N coupled first-order differential equations with the functions having the general form:

$$\frac{dy_i(x)}{dx} = f_i(x, y_1, y_2, \dots, y_n) \tag{3.15}$$

where the functions f_i on the right-hand side are known.

A problem involving ODEs is not completely specified by its equations but also by the nature of the problem's boundary conditions, i.e, algebraic conditions on the values of the functions y_i. Usually, the nature of the boundary conditions determines which numerical methods will be feasible. In our case (the initial value problem) all the y_i are given at some starting value x_s, and it is desired to find the y_i's at some final point x_f , or at some discrete list of points (for example, at tabulated intervals).

There are different types of practical numerical methods for solving initial value problems for ODEs, all of them follow the Euler's method: such methods are based on the same idea, i.e. add small increments to the functions corresponding to derivatives, multiplied by step sizes.

3.2.2 *Integrals of functions*

The integrals of functions we have to solve, for the estimate of the $E(z)$ values for different cosmologies and for the equation of dark energy, are definite integrals.

The basic problem considered, consists into compute an approximate solution of a definite integral:

$$\int_a^b f(x)dx. \tag{3.16}$$

If $f(x)$ is a smooth well-behaved function, integrated over a small number of dimensions and the limits of integration are bounded, we can use methods for approximating the integral with arbitrary precision, called quadrature methods. The valuation of the integral $I = \int_a^b f(x)dx$ is equivalent to solve, for the values $I = y(b)$, the differential equation:

$$\frac{dy}{dx} = f(x) \tag{3.17}$$

with $y(a) = 0$ as boundary condition.

A large class of quadrature rules can be derived by constructing interpolating functions which are easy to integrate. Typically these interpolating functions are polynomials.

3.2.3 *Numerical tools: Runge-Kutta and Qromb*

As for the numerical analysis, we have chosen the Runge–Kutta method to solve our ordinary differential equations and the Romberg's method to compute the definite integrals of functions S. et al. (1992).

3.2.3.1 *Runge-Kutta*

The Runge Kutta method is known to be very accurate and well-behaved for a wide range of problems. It is a reasonably simple and a robust method when combined with an intelligent adaptive step-size routine. It propagates a solution over an interval by combining the information from several steps (each involving one evaluation of the right-hand f's), and then using the information obtained to match a Taylor series expansion up to some higher order.

We start considering the Eulerian formula:

$$y_{n+1} = y_n + hf(x_n, y_n) \tag{3.18}$$

that advances a solution from x_n to $x_{n+1} = x_n + h$. The formula is asymmetrical. It advances the solution through an interval h, but it uses derivative information

only at the beginning of that interval. This leads to a large truncation error per step. Runge-Kutta constructs a more symmetric integration method by making an Euler-like trial step to the midpoint of the interval, and then using the values of both x and y at the midpoint to make the real step across the interval. To be more exact,

$$\begin{cases} k_1 & = & hf(x_n, y_n) \\ k_2 & = & hf(x_n + h/2, y_n + k_1/2) \\ y_{n+1} & = & y_n + k_2 + O(h^3) \end{cases} \tag{3.19}$$

As indicated in the error term, this symmetrization cancels out the first-order error, making the method accurated at second-order. In fact, the above method is generally known as a second-order Runge-Kutta method. Of course, there is no need to stop at a second-order method. By using two trial steps per interval, it is possible to cancel out both the first and second-order error terms, and, thereby, construct a third-order Runge-Kutta method. Likewise, three trial steps per interval yield a fourth-order method, and so on. The general expression for the total error, ϵ, associated with integrating our ODE over an x-interval of order unity using an nth-order Runge-Kutta method is approximately:

$$\epsilon \sim \frac{\eta}{h} + h^n. \tag{3.20}$$

where η is the error due to the intial step of integration. Here, the first term corresponds to round-off error, whereas the second term represents truncation error. The minimum practical step-length, h_0, and the minimum error, ϵ_0, take the values:

$$\begin{cases} h_0 & \propto & \eta^{1/(n+1)} \\ \epsilon_0 & \propto & \eta^{n/(n+1)} \end{cases} \tag{3.21}$$

respectively. h_0 increases and ϵ_0 decreases as n gets larger. However, the relative change in these quantities becomes progressively less dramatic as n increases. It is possible to reasonably choose these parameters for the accuracy of the integration required.

Routine for carrying out one classical Runge Kutta's step on a set on N differential equation has as inputs the values of the indipendent variables, as outputs the values of the right hand side of equations (3.19) by a step h. For the calculations it requires to supply function derivatives and values of derivatives at starting point.

3.2.3.2   *Qromb*

The Romberg method, (QROMB in Numerical Recipes), applys Richardson extrapolation, a sequence acceleration method, used to improve the rate of convergence of a sequence, repeatedly on the trapezium rule or the rectangle rule (midpoint rule).

The midpoint rule or rectangle rule is the simplest method of quadrature that assumes the interpolating function as a constant function (a polynomial of degree zero) which passes through the point $((a+b)/2, f((a+b)/2))$. :

$$\int_a^b f(x)dx \approx (b-a)f\left(\frac{a+b}{2}\right).$$

(3.22)

The trapezoidal rule instead works with the interpolating function that is an affine function (a polynomial of degree 1) which passes through the points $(a, f(a))$ and $(b, f(b))$.

$$\int_a^b f(x)dx \approx (b-a)\frac{f(a)+f(b)}{2}.$$

(3.23)

Romberg's method evaluates the integrand at equally-spaced points. The integrand must have continuous derivatives, though fairly good results may be obtained if only a few derivatives exist. The Romberg method is adequate when the integral must be evaluated many times and the integrand is an analytical function, since the function may be evaluated at arbitrary argument values, and the assumptions behind Romberg integration hold.

The routine for carrying out values of the integral of the function has as input the function together with the extremes of integration. A parameter ϵ sets the accuracy required.

3.2.4 *Implementation*

To find numerically the solutions of the equations (3.2) we need to express them in a suitable form for the application of the Runge-Kutta algorithm, that accepts linear differential equation at the first order. The first equation of the system (3.2) has therefore the form:

$$\ddot{D}^{(1)} = a_1 \dot{D}^{(1)} + a_2 D^{(1)}$$

(3.24)

where we have called with a_1 and a_2 the coeffients respectively of $\dot{D}^{(1)}$ and $\ddot{D}^{(1)}$:

$$\begin{cases} a_1 &= -\left(\frac{3}{a} + \frac{1}{E(a)}\frac{dE(a)}{da}\right) \\ a_2 &= \frac{3\Omega_m}{2E(a)^2 a^5} \end{cases}$$

(3.25)

Let also consider the following change of variables:

$$\begin{cases} y_1 &= \dot{D}^{(1)} \\ y_2 &= D^{(1)} \end{cases}$$

(3.26)

it's possible to write a system of two first order differential equations:

$$\begin{cases} \dot{y}_1 = & \ddot{D}^{(1)} = a_1 y_1 + a_2 y_2 \\ \dot{y}_2 = & y[1] \end{cases}$$

(3.27)

The equation of evolution of cosmic time in the system (3.2) is already a linear differential equation of first order, we express it in this way:

$$\begin{cases} y_3 & = tH_0 = (2/3)a^3 \\ \dot{y}_3 & = (1/E(a))x \end{cases}$$
(3.28)

We have followed the same modus operandi for the growth at the second order, becuase equations (1.53) and (1.62)) differ only for the right side term, making the following substitutions:

$$\begin{cases} y_4 & = \dot{D}^{(2)} \\ y_5 & = D^{(2)} \end{cases}$$
(3.29)

we obtain

$$\begin{cases} \dot{y}_4 = & \ddot{D}^{(2)} = a_1 y_4 + a_2 y_5 + a_2 (y_2)^2 \\ \dot{y}_5 = & y_4 \end{cases}$$
(3.30)

where a_1 and a_2 are the same coefficents of (3.25).

Runge Kutta routine (RK) requires the intial value of the function and of its derivatives, we have considered the ones for an Eds model. This choise is not tying because, as we described in section 1.1.1, the Universe is well approximated by the EDS model when $a(t)$ is small.

We decided to adopt the successive initial values:

$$\begin{cases} a = 10^{-6} \\ D^{(1)}(a) = a \\ D^{(2)} = -3/7a^2 \\ t(a)H_0 = 2/3a^3 \end{cases}$$
(3.31)

In the RK routine, we have given a much smaller number for the initial value of the scale factor a, that's because in this way, RK "has the time to adjust" the behaviour of the function it has to calculate.

We have implemented these informations in a new program *numerical-grow.F90*, using the same program language of PINOCCHIO, i.e. fortran 90. *numerical-grow.F90* comunicates with *cosmo.F90*, for the parameters of the cosmology and for the range of time considered. Both extend from $z = 0$ and $z = 20$ even if the binning in redshift it's different:*cosmo.F90* has 200 logaritmic spaced binning while *numerical-grow.F90* has variable binning in redshift given by the adaptive step control of Runge Kutta.

3.3 COMPARISON OF NUMERICAL INTEGRATIONS WITH ANALYTICAL SOLUTIONS

Our aim is to reach an accuracy of 10^{-4} between the analytical solutions and the numerical ones, in the models where it's possible to do this comparison. Once the numerical solutions are tested to be valid in these particular cases, we extend the validity of our calculations for all general cases.

We have done a preliminary test varying the two free parameters of the Runge Kutta routine ϵ (error associated at the Runge Kutta method) and h (initial step for the integration), to find the best combination of them minimizing as much as possible the discrepacy between the analytical and the numerical result, for open, closed and flat Universe.

We have choosen the growth at the first order to conduct this evaluation. The accuracy requested was reached for different values of ϵ and h. The results carry us to the conclusion that it was enough to consider ϵ and h equal to 10^{-6} as a good compromise between a good accuracy and computational time, push to lower values would have been therefore uselessly.

We have tested a large number of cosmological models for $D^{(1)}$, $D^{(2)}$ and $t(a)$. For each of these models we have considered the contribution of the dark energy constant, i.e we have choosen $w = -1$ and $w_1 = 0$:

- EDS Universe:

    - $(\Omega_{mo} = 1, \Omega_{\Lambda o} = 0)$

- Flat Universe:

    - $(\Omega_{mo} = 0.1, \Omega_{\Lambda o} = 0.9)$
    - $(\Omega_{mo} = 0.2, \Omega_{\Lambda o} = 0.8)$
    - $(\Omega_{mo} = 0.3, \Omega_{\Lambda o} = 0.7)$
    - $(\Omega_{mo} = 0.4, \Omega_{\Lambda o} = 0.6)$
    - $(\Omega_{mo} = 0.6, \Omega_{\Lambda o} = 0.4)$
    - $(\Omega_{mo} = 0.2, \Omega_{\Lambda o} = 0.8)$

- Open Universe:

    - $(\Omega_{mo} = 0.1, \Omega_{\Lambda o} = 0)$
    - $(\Omega_{mo} = 0.2, \Omega_{\Lambda o} = 0)$
    - $(\Omega_{mo} = 0.3, \Omega_{\Lambda o} = 0)$;

For each of them we have analyzed the logarithmic fit of the numerical quantities with the analytical ones and the residuals, both as function of the scale factor.

In figure 16, 17, 18 we show six different flat models. On the top panels it's reported the logarithmic behaviour of the growth in function of the scale factor, while in the bottom panel we show the relative error between the the analytic and numerical results $(D^{(1)}_{num} - D^{(1)}_{an})/D^{(1)}_{an}$. The behaviour of the relative error for all the six cases is of the order of 10^{-4} and from $a = 0.5$ to $a = 1$ it decreases, this is because both the numerical and the analytical result are normalized to $a = 1$. Morover with the increasing of the value of Ω_{mo} (and subsequently of the decreasing of $\Omega_{\Lambda o}$) there is a progressive improvement of the accuracy, that because we get closer to the ESD model.

The required accuracy is reached also for the numerical cosmic time. For the EDS model the difference between the numerical values of the cosmic time versus the analytical one is null. We show in figure 19 the open models. The numerical
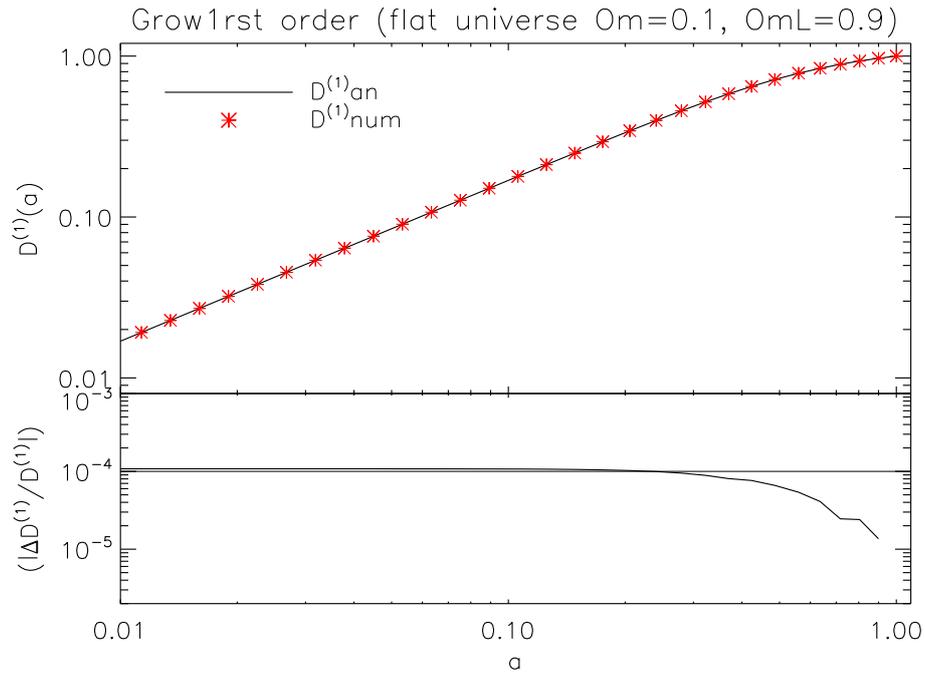
(a)
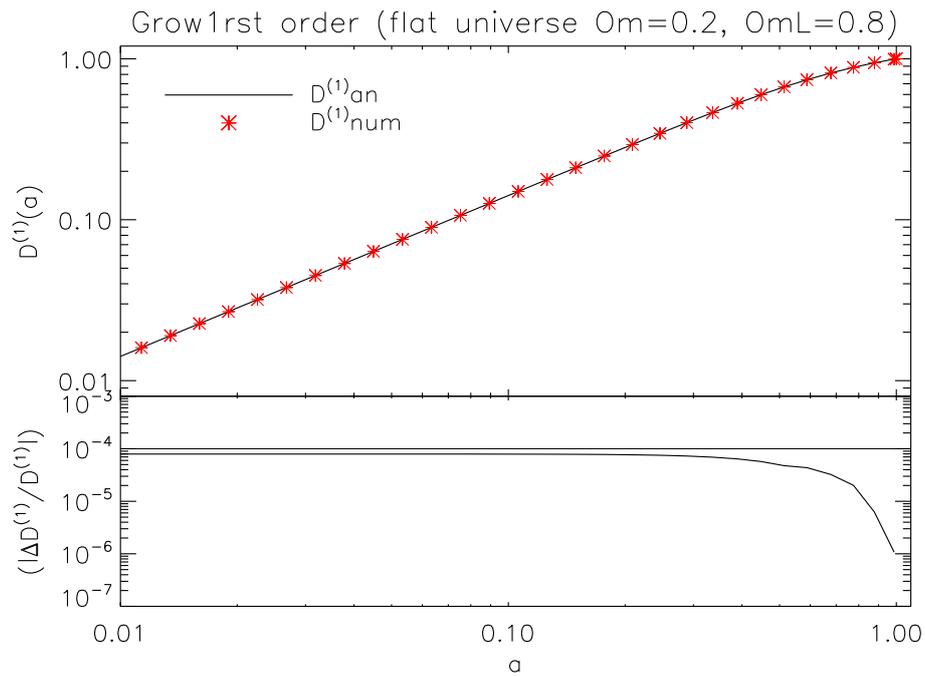
(b)

Figure 16: The figure shows the fit between the numerical and the analytical growth at the first order, for different cosmological flat models, respectively in a)$\Omega_{mo} = 0.1$, $\Omega_{\Lambda o} = 0.9$ and in b)$\Omega_{mo} = 0.2$, $\Omega_{\Lambda o} = 0.8$. For each cases we present on the top of the graph the logaritmic behaviour of the growth in function of the scale factor, while in the bottom panel we show the relative error between the the analytic and numerical results.

(a)

(b)

Figure 17: Same as figure 16 but with cosmological parameters in a)$\Omega_{mo} = 0.3$, $\Omega_{\Lambda o} = 0.7$ and in b)$\Omega_{mo} = 0.4$, $\Omega_{\Lambda o} = 0.6$
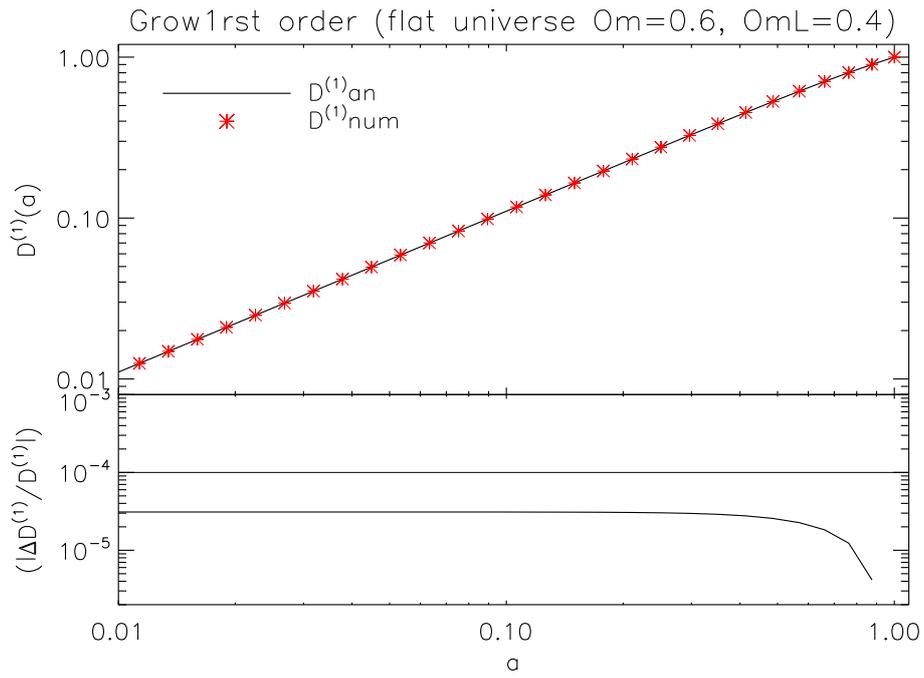
(a)
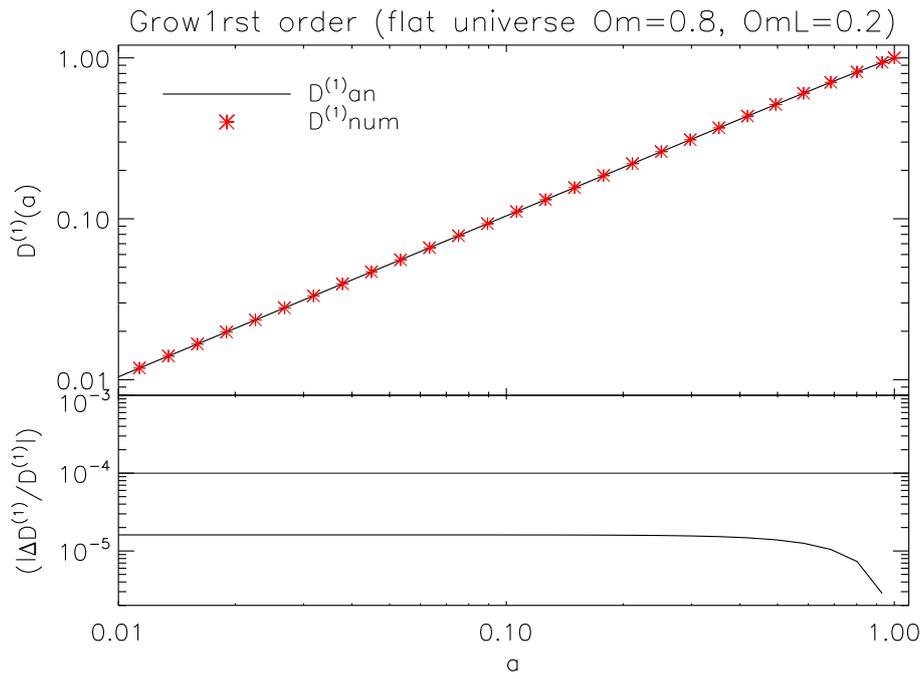
(b)

Figure 18: Same as figure 16 but with cosmological parameters in a)$\Omega_{mo} = 0.6$, $\Omega_{\Lambda o} = 0.4$ and in b)$\Omega_{mo} = 0.8$, $\Omega_{\Lambda o} = 0.2$

cosmic time reproduces perfectly the analytical one in all the cases considered with an accuracy, this time, of the order of 10^{-6}. For range of $0.1 \lesssim a \lesssim 1$ we reach an accuracy of 10^{-8} and nearly 10^{-9} for the open model with $\Omega_{mo} = 0.8$.

Even for the cosmic time the integrations seems to be sensitive to the $\Omega_{\Lambda o}$ parameter. The accuracy increases once we get closer to the EDS model.

For the growth at the second order, we have proceeded considering first of all the EDS model, because it's the only case that admits exact analytical solution. It is illustrated in figure 20. The analytic and numerical points are perfectly superimposed, their relative error is costant and of the order of 10^{-7} for all the values.

Then we have considered different flat models, shown in figures 21 and 22. We want to analyze this time the discrepacy between the numerical results and the approximation given by (3.14). The residuals differ by about 10^{-7} in both the models shown, until $z \sim 0.2$ (top panel), until $z \sim 0.3$ (bottom panel), to grow after exceeding slightly the 10^{-4} value around $z \sim 1$. figure 21)

Open models are considered to check that our numerical solutions are more accurate than the analytical approximation of Bouchet. The discrepacy is of the order of 10^{-2}.

It is possible to conclude that the results pointed out until now confirm that the accuracy we verified is the one we aspected for all the quantities and the cosmological models considered.

We summarize our result in this last graph 24 where we reported the numerical growing mode at first and second order for all the different cosmological models we considered.

Once we have verified that the routines Runge-Kutta and Qromb have correctly the proprierties for our requests, we have linked the *cosmo.F90* to *numerical-growth.F90* so as Runge-Kutta evaluates its adaptive step control of the ODEs, in a range given by each bin of *cosmo.F90*. For each redshift there is therefore now the correspondance between analytical and numerical points.

Morover thanks to the implementation with Qromb of the general form of $E(z)$, that includes the equation of state of dark energy, it has been possible to translate the analytical calculation, in particular cases, of the other quantities we have listed in section 3.1, of *cosmo.F90* in the correspondent numerical one.
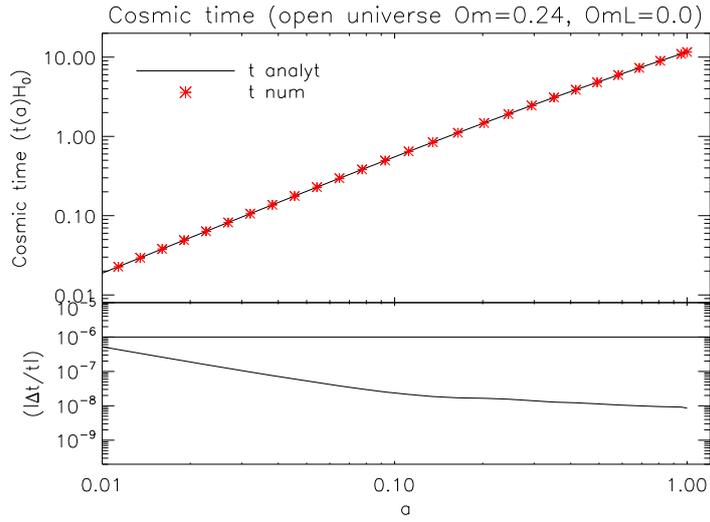
(a)
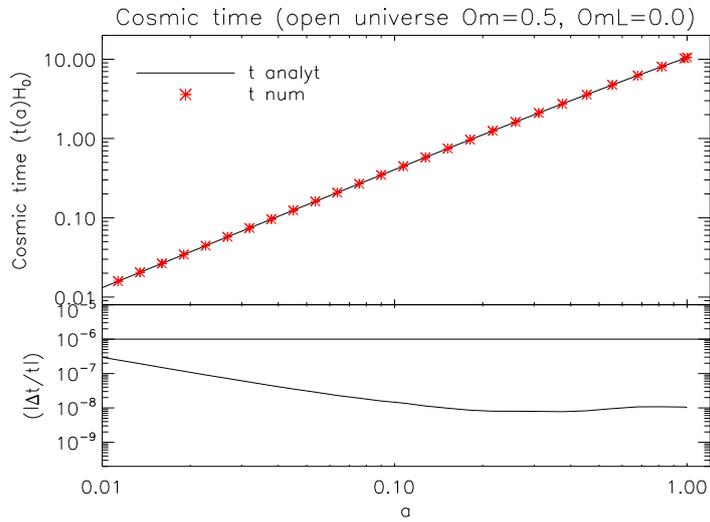
(b)
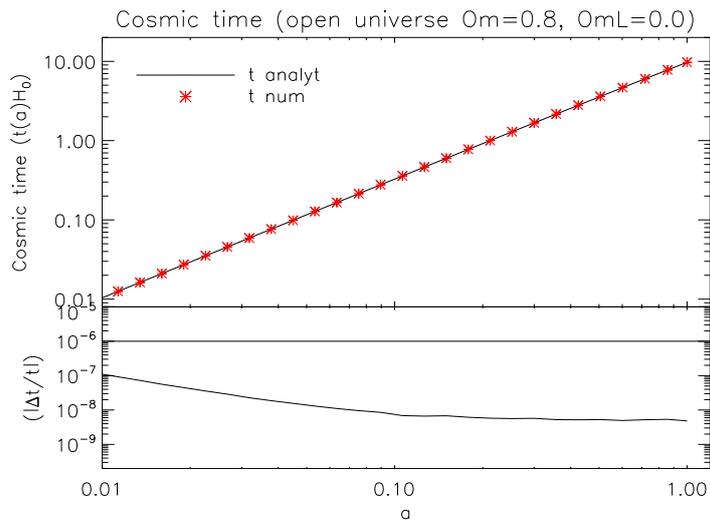
(c)

Figure 19: The figure shows the behaviour of the numerical and analytical cosmic time, for different cosmological open models, respectively in a)$\Omega_{mo} = 0.24$, $\Omega_{\Lambda o} = 0$, b)$\Omega_{mo} = 0.5$, $\Omega_{\Lambda o} = 0$, in c)$\Omega_{mo} = 0.8$, $\Omega_{\Lambda o} = 0$. For each case we present on the top of the graph the logarithmic trend of the growth in function of the scale factor, while in the bottom panel we show the relative error between the the analytic and numerical results.
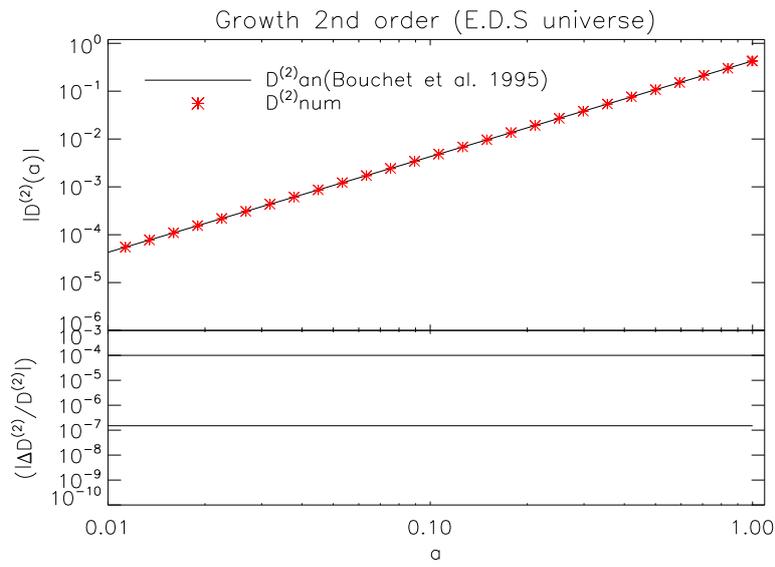
Figure 20: The figure shows the behaviour of the numerical and analytical growth at the second order, for the EDS model: $\Omega_{mo} = 1$, $\Omega_{\Lambda o} = 0$. On the top of the graph it's reported the logaritmic trend of the growth (2nd order) in function of the scale factor, while in the bottom panel we show the relative error between the the analytic and numerical results.
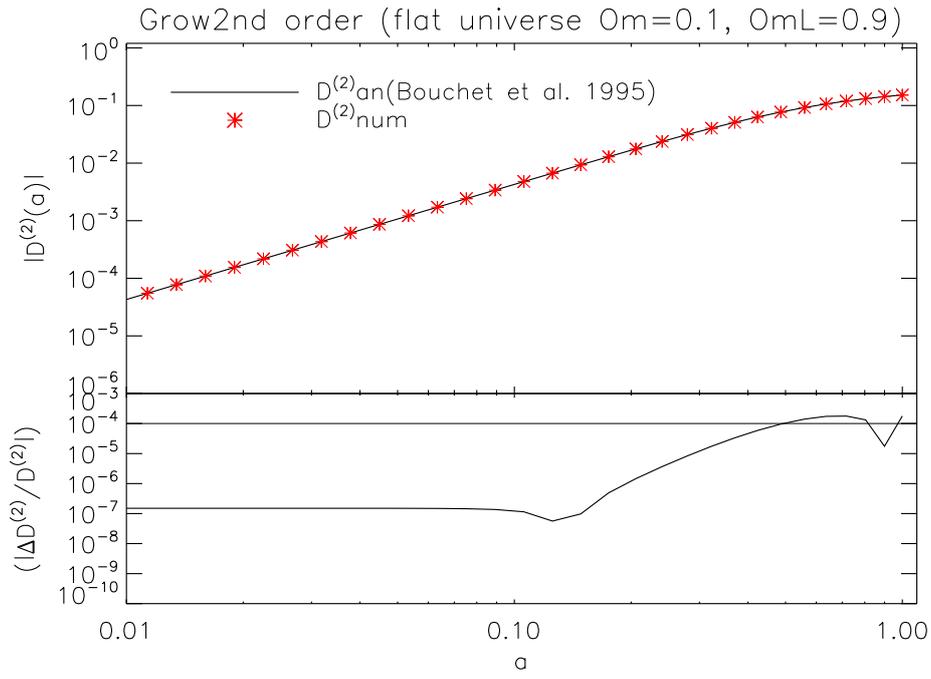
(a)
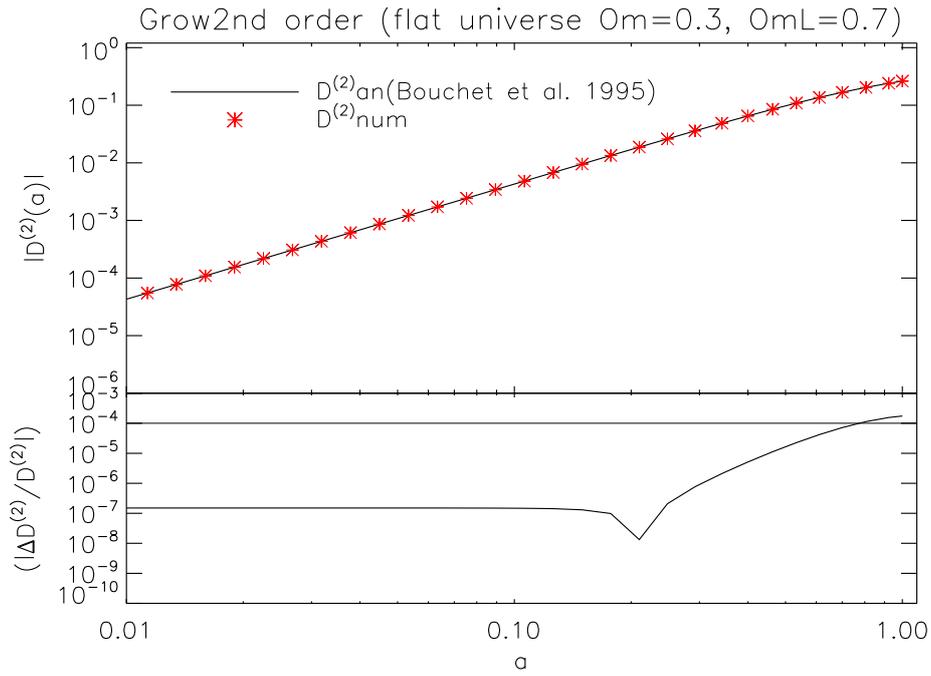
(b)

Figure 21: The figure shows the behaviour of the numerical and analytical growth at the second order, for different cosmological flat models, respectively in a)$\Omega_{m0} = 0.1$, $\Omega_{\Lambda 0} = 0.9$ and b)$\Omega_{m0} = 0.3$, $\Omega_{\Lambda 0} = 0.7$. For each case we present on the top of the graph the logaritmic trend of the growth (2nd order) in function of the scale factor, while in the bottom panel we show the relative error between the the analytic and numerical results.

(a)

(b)

Figure 22: The same of figure 21 but with cosmological parameters in a)$\Omega_{mo} = 0.4$, $\Omega_{\Lambda o} = 0.6$, and in b)$\Omega_{mo} = 0.8$, $\Omega_{\Lambda o} = 0.2$
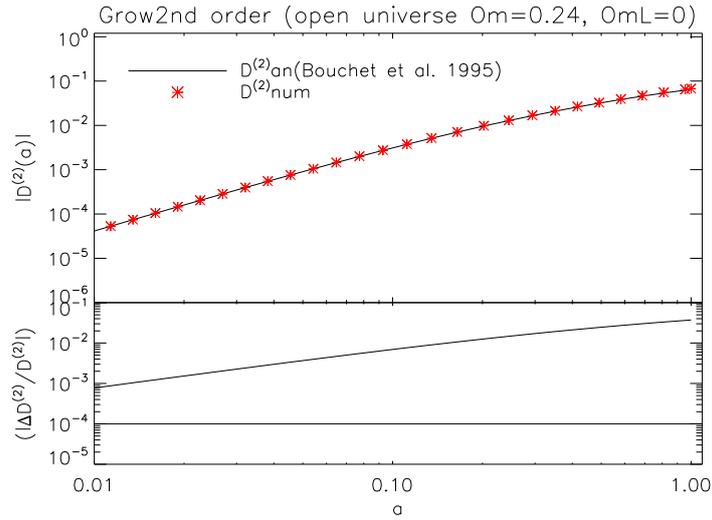
(a)
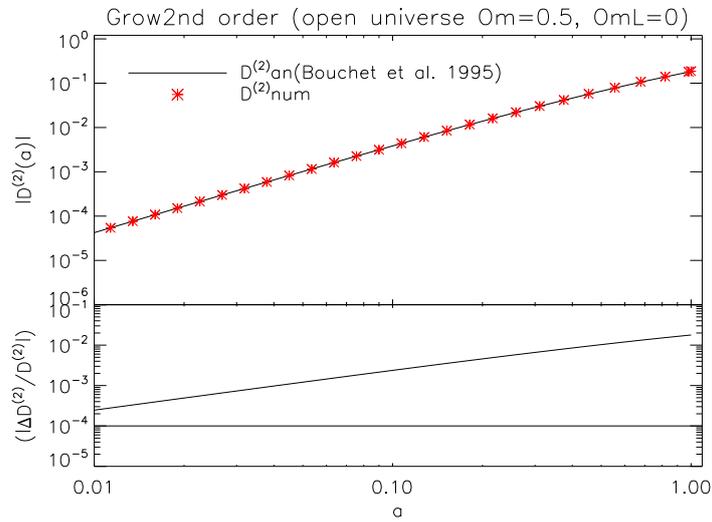
(b)
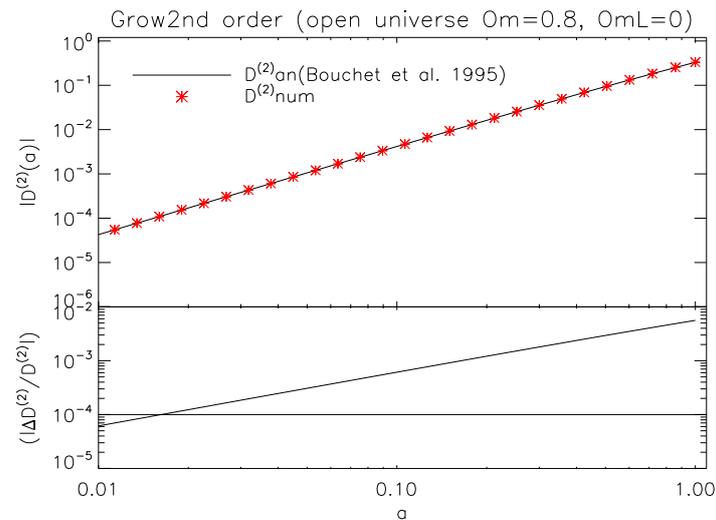
(c)

Figure 23: The figure shows the behaviour of the numerical and analytical growth at the second order, for different cosmological open models, respectively in a)$\Omega_{mo} = 0.24$, $\Omega_{\Lambda o} = 0$, b)$\Omega_{mo} = 0.5$, $\Omega_{\Lambda o} = 0$, in c)$\Omega_{mo} = 0.8$, $\Omega_{\Lambda o} = 0$. For each case we present on the top of the graph the logaritmic trend of the growth (2nd order) in function of the scale factor, while in the bottom panel we show the relative error between the the analytic and numerical results.

Figure 24: Numerical growing mode at first (on the top) and second order (on the bottom) for different cosmological models.

4

SIMULATED PAST-LIGHT-CONE CATALOGUES WITH PINOCCHIO

The second part of this work describe the realization of simulated past-light-cone catalogues of dark matter halos with PINOCCHIO.

Modern galaxy surveys such as the Sloan Digital Sky Survey (a tridimentional map of which is reported in figure 25) and the 2-degree Field Galaxy Redshift Survey have revolutionised our view of the galaxy distribution and have played a key role in shaping the constraints on our cosmological model.

Many ongoing and future obsevational campaigns as EUCLID, BOSS (Baryon Oscillation Spectroscopic Survey), VIPERS (VIMOS Public Extragalactic Redshift Survey), DES (Dark Energy Survey) will survey large parts of the sky to a depth level sufficient to reach $z \sim 1$ to probe the origin of the accelerating universe and help uncover the nature of dark energy by measures of high precision.

The uncertainies in the estimate of physical parameters from observable quantities will be but dominated by sistematics connected to sample variance and to the bias with which galaxies trace mass.

An accurate assessments of these systematics requires the use of past-light-cone-catalogues. Past-light-cone allow us to cast the predictions of theoretical models of dark matter and structures formation in a form that can be directly compared with observables. Moreover because they are free of a variety of effects, such as non-uniform coverage of the sky and a selection function that varies strongly with radial distance from the observer, the comparison of their synthetic data results with the observational ones, can help to adjust and tune the performance of a real survey to reduce any systematic effects.

For these reasons they find different fields of applications in the astrophysical contest. We have focused on halos with mass larger than $\sim 10^{14}$ M$\odot$. Many works have in fact already proven the ability of clusters number counts to constrain cosmological parameters. We have therefore realized a pipeline for estimating the sample variance of our number counts, identified by the past-light cone catalogues of PINOCCHIO.
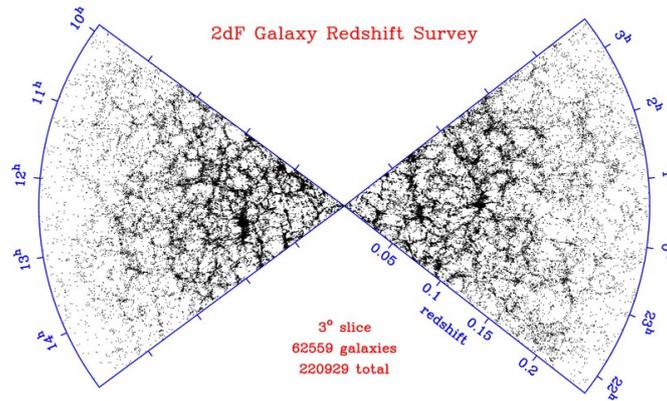
Figure 25: Proiection on the plane of 3° of the tridimentional map produced by 2dFGRS. It contains 63000 galaxies and it shows the structure on large scale of the galaxies distribution (http://msowww.anu.edu.au/2dFGRS/ http://msowww.anu.edu.au/2dFGRS/).

In section § 4.1 we present a brief description of the past-light-cone, in section§ 4.2 we present the procedure we used to build it within PINOCCHIO and in section § 4.3 we describe the analysis done with it.

4.1 LIGHT-CONE CATALOGUES

A first step to understand the clustering of galaxies is to consider the clustering in terms of dark matter halos. The standard picture of structure formation, that we described in the first chapter, predicts that the luminous objects form in the gravitational potential of dark matter halos. Therefore, a detailed description of halo clustering is the most basic step toward understanding the clustering of these objects (Hamana et al., 2001).

A powerful tool to reach this aim are light-cone simulated catalogues of dark matter halos. Light-cone simulated catalogues of dark matter are catalogues in which the halos are located at the position at which the light emitted from them has time to reach the observer. They are able to mimic real surveys because they give the expected redshift distributions of the cosmic structures.

The catalogues that we present and we have considered are *past-light-cone* catalogues of dark matter halos, in which dark matter halos are placed according to the epoch at which they first cross the observer's past lightcone.

Eventually the halo model that produces the dark matter halo population and distribution can be combined with model that gives the relation between the halos and luminous objects creating light cone catalogue of galaxies or of cluster of galaxies. This gives rise to a lot of different applications for the study of all the properties of these objects and of the processes involved in the hierarchical structure formation. Cole et al. (1998) constructed mock redshift catalogues for a wide range of cosmologies, focusing on the study of the statistical properties that are used to quantified large scale structures as correlation function $\xi(r)$ on scale of $1 - 10 \mathrm{Mpch}^{-1}$ and power spectrum $P(k)$ to facilitate this tasks for the
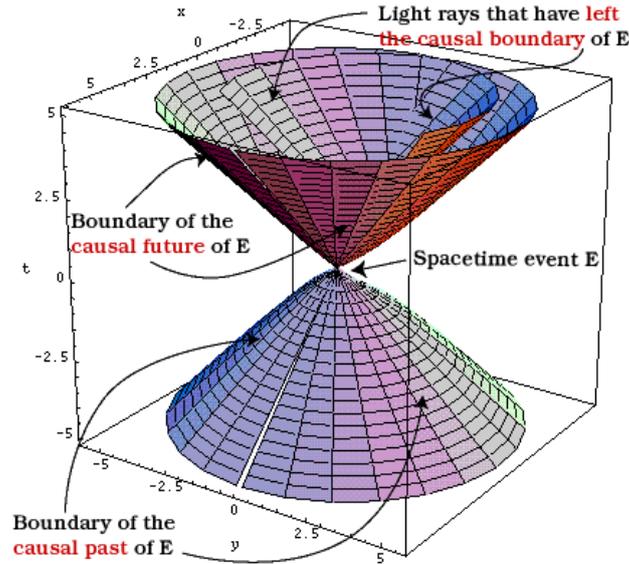
Figure 26: Light cone rapresentation in 3D: two spatial coordinates and one time coordinate. The upper cone is the future light cone, the bottom cone is the past one (http://themaclellans.com/timetravel.html http://themaclellans.com/timetravel.html).

SSDS surveys and the 2dF one. Jouvel et al. (2009) developed a "COSMOS Mock software" to help design future Dark Energy missions in terms of the number of galaxies available for the weak-lensing analysis as a function of the PSF size and depth of the survey. Manera et al. (2012) created mock catalogues to be used to compute covariance matrices of large scale clustering measurements, and to test the method of analysis. Merson et al. (2012) considered light-cone catalogues for the study of the effectiveness of the BzK color selection which is a tecnique designed to isolate galaxies in the redshift interval $1.4 \leqslant z \leqslant 2.5$. Wang et al. (2013) used a realistic mock SDSS DR7, obtained by populating dark matter halos in the Millenium simulation with galaxies, to test their Hamiltonian Markov Chain Montecarlo(HMC) method for the reconstruction of the initial density field of the local Universe.

4.1.1 *Light-cone*

According to Einstein's relativity, a light-cone is a representation of the path of light through spacetime, emanating from a single event E and traveling in all directions.

Strictly speaking, we consider here a locally flat spacetime, the use of the Minkowski space to describe physical systems over finite distances, applies only in the Newtonian limit of systems without significant gravitation. In the case of significant gravitation, spacetime becomes curved and one must abandon special relativity in favor of the full theory of general relativity.

In figure 26 we show the 3-dimentional rapresentation of the light-cone.

In this representation the speed of light is set 1 and a light ray follows a path at 45 degrees with with respect to the axes. The vertical axis represent the time line, while the horizonal one the space one. This means that each observer drawing a spacetime diagram in which he is at rest must have the worldlines of light at the same angle of 45 degrees from his worldline (his time axis), and 45 degrees from his plane of simultaneity (his space axis).

The light from an event spreads out in a circle after the event E occurs. When this growing circle is plotted with the vertical axis of the graph, representing time, the result is a cone, known as the future light cone. It represents the future history of a light emitted at that event. The past-light-cone represents all directions from which light can be received at that event, it behaves like the future light cone in reverse, a circle that contracts in radius at the speed of light until it converges to a point at the exact position and time of the event E. In practice the light would form an expanding or contracting sphere, rather than a circle in 3D, and the light cone would actually be a four-dimensional shape.

In flat spacetime, the future light cone of an event is the boundary of its causal future and its past-light-cone is the boundary of its causal past.

In a Friedmann–Lemaître–Robertson–Walker (FLRW) metric, that we recall, is the metric for an homogenus (costant space curvature) and isotropic Universe (see 1.1.1) for an observer located at redshift $z = 0$, the proper radial distance between the observer and a light event ($ds^2 = 0$) at redshift z is:

$$r_c(z) = a(t) \int_0^z \frac{c \, dz'}{H(z)} = a(t) \frac{c}{H_0} \int_0^z d\prime z / E(z), \tag{4.1}$$

where $H(z)$ is $H_0 E(z)$ and $E(z)$ is given by equation (1.11) expressed as a function of the redshift z. The proper radial distance characterizes all the events on the past-light-cone for an observer at $z = 0$, i.e. it is the proper distance from which, we can receive light signals. The equation (4.1) is valid for the concordance model of Universe. If we consider a Universe with different curvature we have:

$$r_c(z) = a(t) \frac{c}{H_0} \frac{1}{\Omega_{k2}} \sin\left(\Omega_{k2} \int_0^z d\prime z / E(z)\right) \tag{4.2}$$

for a positive curvature of the Universe,

$$r_c(z) = a(t) \frac{c}{H_0} \frac{1}{\Omega_{k2}} \sinh\left(\Omega_{k2} \int_0^z d\prime z / E(z)\right) \tag{4.3}$$

for negative curvature, where in both equation the term $\Omega_{k2} = \sqrt{|\Omega_k|}$.

4.2 CONSTRUCTION OF THE PAST-LIGHT-CONE

The simulated catalogues we have constructed are past-light-cone-catalogues. In this section we provide a description of the procedure we have followed, for constructing past-light-cone simulated catalogues with PINOCCHIO.

PINOCCHIO, as we have described in chapter 2, is able, with very limited computing resources, to generate catalogues of cosmological dark matter halos with known mass, position, velocity, merger history on a 3D grid.

The 3D grid represents a sample of the volume of the entire Universe, the past-light-cone represents the part of this sample, that is visible for an observer located at the vertex of the Universe. Therefore we have selected the halos that are on the observer's past-light-cone.

At the operative level, we have operated in the third version of PINOCCHIO (see section 2.2.2), in the following steps:

- choose and realize the cosmological volume to investigate;

- generate a population dark matter distribution;

- apply the condition of belonging to the past-light-cone to the dark matter halos;

4.2.1 *Generation of the cosmological volume*

Current and future surveys are designed to probe ever larger cosmological volumes. As a result there is a growing demand for simulations with boxes of sufficient size to match the volumes of these surveys. Unfortunately, current computing power means that there's to make a compromise between the volume of the simulation box and the resolution at which the simulation is carried out.

It's possible therefore to construct a sufficiently large cosmological volume replicating smaller box simulation.

For very shallow surveys (e.g. with a median redshift $z \lesssim 0.05$), the lookback time is sufficiently small that typical structures properties will not have undergone significant evolution across the redshift interval covered by the survey. Instead for very deep surveys which cover a significant lookback time, the structures proprerties have a substantial evolution and so the growth of large-scale structure.

The simulation box side-length, L_{box}, corresponds to the co-moving distance out to the corrisponding z. Therefore, in order to generate a cosmological volume that is of sufficient size to fully contain any survey that extends to a certain redshift z_{max} and so to a certain co-moving radial distance r_{max} (see equation (4.1) with $r_c(z) = r_{max}(z_{max})$), it is necessary to tile replications of the simulation box, as we have said before.

Let be n_{rep} the number of replications per axis, that need to be stacked around a box, to reach the maximum co-moving radial distance r_{max} we have:

$$n_{rep} = \left\lfloor \frac{r_{max}}{L_{box}} \right\rfloor + 1, \tag{4.4}$$

with $(2n_{rep} + 1)^3$ repetitions, that creates a superbox. To give an idea, if the observer is located at the center of the box with $L_{box} = 3 Gpc/h$, for a maximum redshift $z_{max} = 2$, we need almost 5 repetions of the box size, reaching a maximum co-moving radial distance of $r_{max} = 7.5 Gpc/h$. We have constructed a full sky past-light-cone in order to have the possibility of cosidering each desired geometry of the survey.

Our original box have a side-lengh of the box of 3Gpc/h for a grid size of 1200 particles. We have considered 5^3 repetitions. The past-light-cone is generated to have the observer in a random position within the superbox.

4.2.2 *Generation of the dark matter halos within the lightcone*

The population of dark matter halos is reproduced according to the algorithm of the PINOCCHIO code, described in section 2.2.1.

The light cone selection of halos consists in identifying those halos whose light has time to reach the observer at $z = 0$. These objects are those that are on the past light cone surface (see section 4.1.1).

Taking into account what we explained in section 4.1.1, a halo will be "visible" to the observer when the following condition is satisfied:

$$\|\mathbf{x}_{halo}(z) - \mathbf{x}_{plc}(z)\| - r_c(z) = 0, \tag{4.5}$$

where $\mathbf{x}_{halo}(z)$ is the comoving coordinate of the location of the halos, z is the redshift at which the halo crosses the past-light-cone, $\mathbf{x}_{plc}(z)$ is the comoving coordinate of the position of the observer ($\mathbf{x}_{plc}(z)$ takes into account of the replications of the box) and r_c is the distance that light could have traveled from the redshift z up to the observer, i.e. the comoving radial distance that is visible to the observer at the present day.

We made some modifications in the part of the code that works on the fragmentation of the collapsed medium. Each time a particle collapsed, we checked the the neighbor halos. Each halo is updated from the the last evaluated redshift, z_{last}. If the halo at z_{last} is inside the past-light-cone, but at the instant z_{check} we check the condition, it is outside, we can calculate the instant the halo crosses the past light cone, $z_c$ross. If both at z_{last} and z_{check} the halo is out of the past-light-cone, we update z_{last}. Then We redo the check with the coordinates of the position the halo assumes in the repeated boxes. We perform the light-cone selection in the interval of time in which the masses of the neighbor halos remain constant. In fact positions and velocities are calculated via the Zel'dovich approximation. As a matter of fact, halo masses change depending on the accretion and the merging they are involved in. If the mass changes, the position and the velocity of the halo also varies because they are calculated from the barycenter of the groups of particles that form the halo. For this reason we verified the condition at different times of the halo formation process, as computed by the PINOCCHIO code.

From what we have discussed above, it's clear that we need to estimate in a given istant z_{check} the proper distance to verify the condition (4.5). Then we want to estimate the istant at which the halo crosses the past-light-cone, z_{cross}.

We start integrating in PINOCCHIO code, in the new version of *cosmo.F90*, the equation for the proper distance for the cosmological parameters of the simulation and taking into account the dark energy contribution, described by the dark energy equation of state.

The calculation of the integrals above is performed again with qromb (see section 3.2.3.2).

The z_{cross} is a solution of a root-finding one-dimensional problem. The redshift is the independent variable.

There are different types of practical numerical methods for solving this kind of problem. All are based on the Intermediate Value Theorem. In essence, this theorem says that if f is a continous function on $[a, b]$ and the sign of $f(a)$ is different from the sign of $f(b)$, then there must be some point, c, in the interval such that $f(c) = 0$, and is thus a root of the function. Numerical root-finding methods use iteration, producing a sequence of numbers that hopefully converge towards a limit (the so-called "fixed point") which is a root. The first values of this series are initial guesses. The method computes subsequent values based on the old ones and the function f. Questions include ability to separate close roots, robustness in achieving reliable answers despite inevitable numerical errors, and rate of convergence.

We decide to choice the Brent's method. It is a complicated but popular root-finding algorithm combining the bisection method, the secant method and inverse quadratic interpolation. It has the reliability of bisection but it can be as quick as some of the less reliable methods.

4.3 PIPELINE FOR NUMBER COUNTS OF VIRIALIZED OBJECTS

We present in this section a pipeline for estimating the sample variance in the number counts of virialized objects, identified by the past-light cone catalogues of PINOCCHIO over a range of redshifts. We address considerations mainly to cluster surveys. In fact we set a thereshold mass of $M_{\text{th}} = 1.61 \cdot 10^{14}$ solar mass (we consider halo with more than 100 particles), below which halos are not taken into account in our catalogues.

Among different probes of large scale structures, many works have already proven the ability of clusters number counts in constrainig cosmological parameters Borgani and Guzzo (2001). The abundance of clusters above a certain mass in a given area of the sky as a function of redshift is very sensitive to the amplitude and growth rate of perturbations as well as the comoving volume per unit redshift and solid angle. Recently, the use of very large volume cosmological simulations has led to significant improvements in our knowledge of the theoretical cluster mass function (Sheth and Tormen, 1999; Jenkins et al., 2001; Evrard et al., 2002; Zheng et al., 2002; Warren et al., 2006). In particular, Jenkins et al. (2001) found that the mass function at cluster scales can be described by a universal analytic fit for all cosmologies and redshifts to an accuracy of 5-10 per cent in amplitude. Evrard et al. (2002) found that it is dominated by the uncertainties in conversion in mass definitions and sample variance.

The sample variance is in fact generally comparable to or greater than shot noise and thus cannot be neglected in deriving precision cosmological constraints (Hu and Kravtsov, 2003).

4.3.1 *Sample Variance*

The number density of simulated cluster population of the past-light-cone is affected by uncertainty due to sample variance.

When we consider the simulated superbox, we consider just a sample of the entire volume of the Universe.

Our aim is to evaluate the uncertainity, called sample variance, associated to the single realization for the number counts from the simulated past-light-cone catalogues. Because the generation is a stocastic process, we have evaluated the espected values and the variance of the number counts, creating different ten realizations of the same volume of the Universe.

We proceed in this way. Each set is produced on a box of 3 Gpc/h lenght size, of $N_{par} = 1200^3$ particles. We have considered $n_{rep} = 5$. We have changed for each realization the random seeds that generates the density field on the superbox, but we fix the following cosmological parameters: matter density $\Omega_{mo} = 0.26$, dark energy density $\Omega_{\Lambda} = 0.74$, baryon density $\Omega_{bo} = 0.044$, $\sigma_8 = 0.8$, spectral index $n = 0.96$, $w_0 = -1$ and $w_1 = 0$ for the equation of state and the Hubble parameter $h = 0.7$. We have realized in this way 10 past-light-cone catalogues from which we have obtained the number counts, i.e. the number of dark matter halos in a redshift bin. We have worked on a range of redshift of $0 \leqslant z \leqslant 2$, divided by 20 logarithmic spaced bins.

Let's call the number of halos of the i-th bin in redshift and of the j-th realization, divided by the amplitude of the redshift bin $(dN/dz)_{i,j}$. This quantity is linked to the mass function by:

$$(dN/dz)_{i,j} = \Delta\Omega \int_{z_i}^{z_{i+1}} \frac{dV}{dz d\Omega} \int_{M_{th}}^{\infty} n(M, z) dM \qquad (4.6)$$

where $\Delta\Omega$ in the fraction of the solid angle considered, $\frac{dV}{dz d\Omega}$ is the comoving volume for unit of redshift and solid angle, M_{th} is the thereshold mass and $n(M, z)$ is the mass function. The cosmology therefore enters both on the definition of the mass function both in the volume considered.

In figure 27 on the top we show the behaviour of the number counts of each realization, $(dn/dz)_{i,j}$, as function of the average redshift of the halos of each bin ($z_{aver,i} = \frac{\sum_{k=1}^{N_i} z_{k,i}}{N_i}$) where N_i are the counted halos in the i-th bin in redshift). They are compared with the predictions of number counts given by using the analytic fit of Warren et al. (2006) for the mass function (dotted-dashed red line). The cyan lines are number counts of the ten different realisations. They all confirm the peak at redshift $z = 0.5$ predicted by the analytical curve, given in red. The number of halos in the redshift bin centered in 0.5 reach the value of $\sim 5 \cdot 10^5$. They separate themself from it from a minum of ~ 6 per cent in the range of $0.8 \lesssim z \lesssim 2$ to a maximum of ~ 25 per cent around $z = 2$. The larger difference at $z = 2$, is a noted characteristic of PINOCCHIO that understimates at high redshift the mass function of Warren et al. (2006). We are not worried about about this discrepance because there is not consensus between simulations that are still not able to find a universal analytical fit of the mass function at

high redshift. The difference of the realisations, in the mass function tail is the consequence of the variance of the sample.

In figure 27, where we have also reported together with the behaviour of the $(dN/dz)_{i,j}$, the sample variance (blue lines) and the Poisson noise (red bars) in function of the redshift bin. Given a probability distribution $p((dN/dz)_{i,j})$ for the number counts with mean value $\langle (dN/dz) \rangle_i$ and variance $\langle (dN/dz) \rangle_i$, we define the sample variance of the counts $(dN/dz)_{i,j}$ as:

$$\sigma_{sv} = \sqrt{\left\langle \left(\frac{dN}{dz} \right)^2 \right\rangle_i - \left\langle \left(\frac{dN}{dz} \right) \right\rangle_i^2} \tag{4.7}$$

where $\langle \left(\frac{dN}{dz} \right) \rangle_i$ is the mean value for each bin, of the $(dN/dz)_{i,j}$, over the $N_{re} = 10$ realizations:

$$\left\langle \left(\frac{dN}{dz} \right) \right\rangle_i = \sum_{j=1}^{N_{re}} \left(\frac{dN}{dz} \right)_{i,j} / N_{re} \tag{4.8}$$

The sample variance is the uncertainty in eccess to the Poissonian noise, this last one is calculated taking into account a single realization of reference (random seed 23489):

$$\sigma_{Pois,i} = \left(\frac{\sqrt{dN}}{dz} \right)_{i,ref} = \left\langle \frac{dN}{dz} \right\rangle_i / \sqrt{< N_{i,ref} >} \tag{4.9}$$

The $(dN/dz)_{i,j}$ are included, for all the redshift range considered, in the curves traced by $\langle (dN/dz)_i \rangle \pm \sigma_{sv}$

This effect is more visible and quantifiable in the study of the nomalized profiles of the number counts The $(dN/dz)_{i,j}$ of each realisation, for each bin, are normalized to the mean value of that bin $\langle \frac{dN}{dz} \rangle_i$:

$$\left(\frac{dN}{dz} \right)_{norm,i} = \left(\frac{dN}{dz} \right)_{i,j} / \left\langle \left(\frac{dN}{dz} \right) \right\rangle_i \tag{4.10}$$

The probability distribution now is $p((dN/dz)_{norm,i})$ for number counts with mean $\langle (dN/dz) \rangle_{norm,i}$ and variance $\langle (dN/dz)^2 \rangle_{norm,i}$. We define the relative error of the counts $(dN/dz)_{norm,i}$, i.e the relative error due to the sample variance, as:

$$\epsilon_{sv} = \sqrt{\left\langle \left(\frac{dN}{dz} \right)^2 \right\rangle_{norm,i} - \left\langle \left(\frac{dN}{dz} \right) \right\rangle_{norm,i}^2} \tag{4.11}$$

calculated for each z bin.

We have calculated again the Poisson noise of the normalized quantities taking into account the number counts of one realization of reference (random seed 23489). The Poisson relative error, for each bin in redshift, is given by:

$$\epsilon_{Pois,i} = \left(\frac{\sqrt{dN}}{dz} \right)_i / \left(\frac{dz}{dN} \right)_i = 1/\sqrt{N_i} \tag{4.12}$$

The results are reported on the bottom of figure 27 where we plot, with cyan lines, the $\left(\frac{dN}{dz}\right)_{norm,i}$ of each realisation in function of the bin in redshift.

The normalized profiles of number counts are distributed around the unit, the smallest error is around $z = 0.5$. The contribution of the sample variance uncertainity is weakly larger than the Poisson noise, for all the redshift range. Between $0.3 \lesssim z \lesssim 1.1$ their uncertainities are both of the order of 3 per cent. For z bigger than $z = 1.1$ their uncertainities both grow reaching nearly the 20 per cent at $z = 2$.

We show moreover that uncertainties due to sample variance are comparable to or greater than the Poisson noise. Therefore, the variance should not be neglected in analyses aiming at deriving precision cosmological constraints.

4.3.1.1 *Survey window*

PINOCCHIO has now the potentialities to mimic data for future surveys as Euclid. This wide survey will cover 15000 deg^2 of the extragalactic sky. With the intention of emulating the volume that the Euclid's survey will cover, we select all the halos of our past-light-cone within a solid angle with aperture of ~ 1.29 rad, pointing along one random direction of the sky. Our aim is to evaluate the uncertainity associated to the single realization for the number counts of this geometrical configuration. We have considered 5 different random directions, specified by the coordinates $(\theta_{dir}, \phi_{dir})$ for each of the 10 realizations, we did before.

The criteria of selection of the halos we were interested in, uses the spherical trigonometry laws. We have considered all the halos that satisfy the following:

$$\begin{aligned} \alpha \leqslant \arccos(&\cos(\phi_{dir} - \phi_{halo})\sin(\pi/2 - \theta_{halo})\sin(\pi/2 - \theta_{dir}) + \\ &+ \cos(\pi/2 - \theta_{halo})\cos(\pi/2 - \theta_{dir})) \end{aligned} \tag{4.13}$$

We have proceeded evaluating the Poisson noise and the sample variance over the 50 realisations. In figure 28 we report the behaviour of the $(dN/dz)_i$ (on the top) and the the normalized profiles of $(\frac{dN}{dz})_{norm,i}$ (on the bottom), both in cyan lines, the error Poisson noise with red bars and sample variance with blue lines as function of the bin in redshift.

The $(dN/dz)_i$ have a peak around $z = 0.5$, reaching the value of about $2 \cdot 10^5$, that is smaller than the one reached around the same redshift by the $(dN/dz)_i$ of the ten realisation. This is due to the fact that we are considering a smaller volume. On the tail the sample variance presents the same proprierties respect the one illustrated on the bottom of figure 27, but it is larger in the range $0.3 \lesssim z \lesssim 1.1$ as . Its contribution is weakly larger than the Poisson noise, and comparable to Poisson noise in the range of $0.8 \lesssim z \lesssim 2$. With with respect to previous case at low redshift, and mainly to $z = 0$, the sample variance is significant, it is around 15 per cent. This is precisely due to the fact that we are looking at a smaller part and then a smaller volume. The different distribution of the large scale structures causes the increasing of the sample variance.

Because number counts are sensitive to cosmological parameters (mainly to Ω_{mo} and σ_8), we want to apply the same geometrical criteria of selection for the halos but considering different cosmological models.
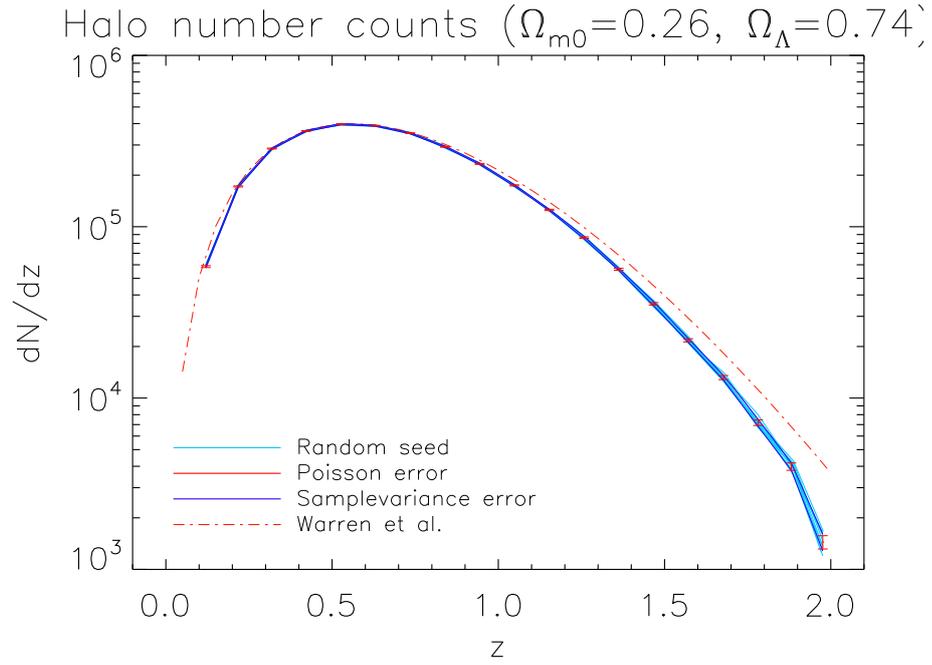
(a)
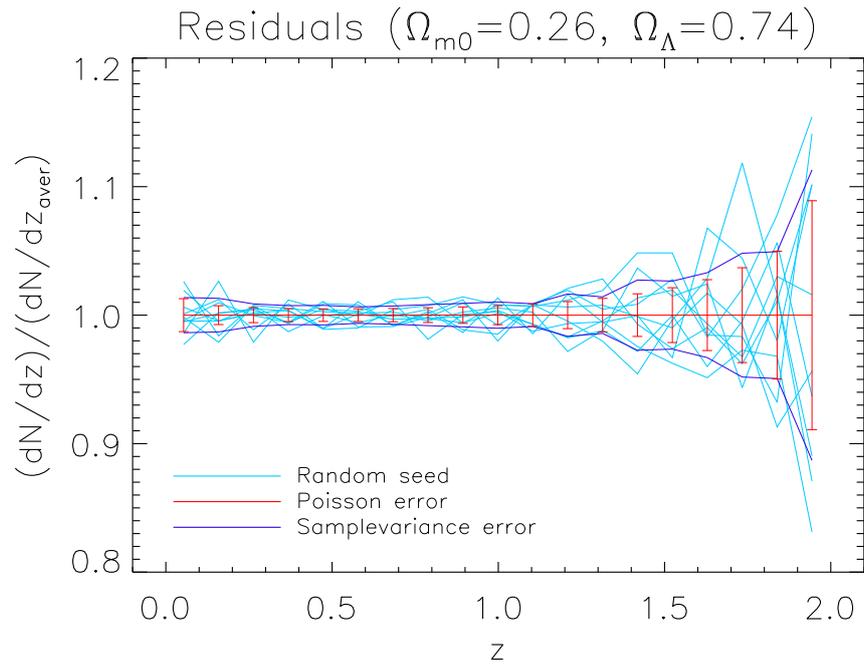
(b)

Figure 27: On the top we have reported the behaviour of the $(dN/dz)_{i,j}$, sample variance (blue lines) and Poisson noise (red bars), as function of the redshift bin. On the bottom we have shown the normalized profiles of $(\frac{dN}{dz})_{norm,i}$ for the different realisations (cyan lines), error Poisson bars (red bars) and variance uncertainity (blue lines) as function of the bin in redshift.
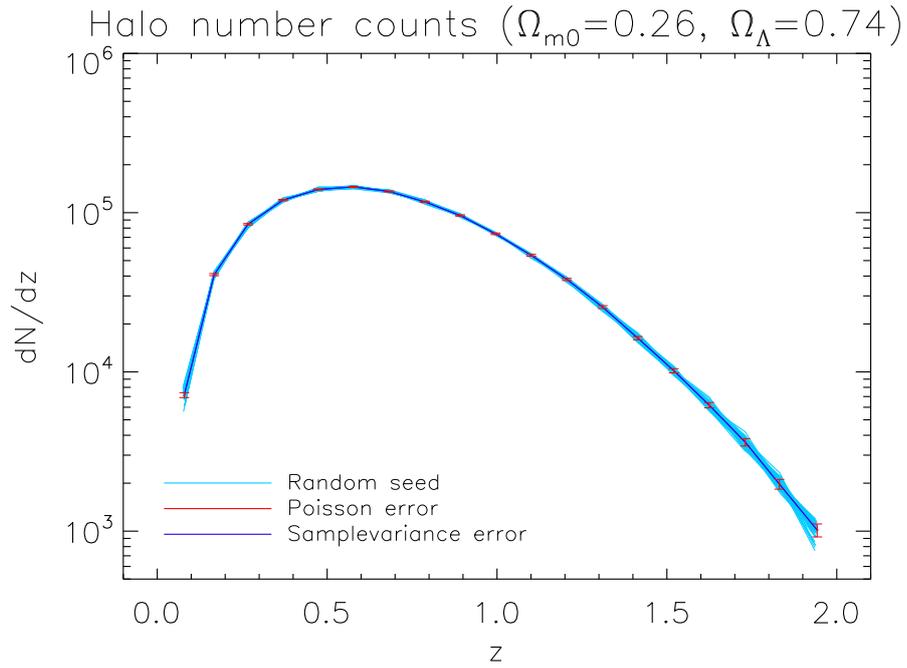
(a)
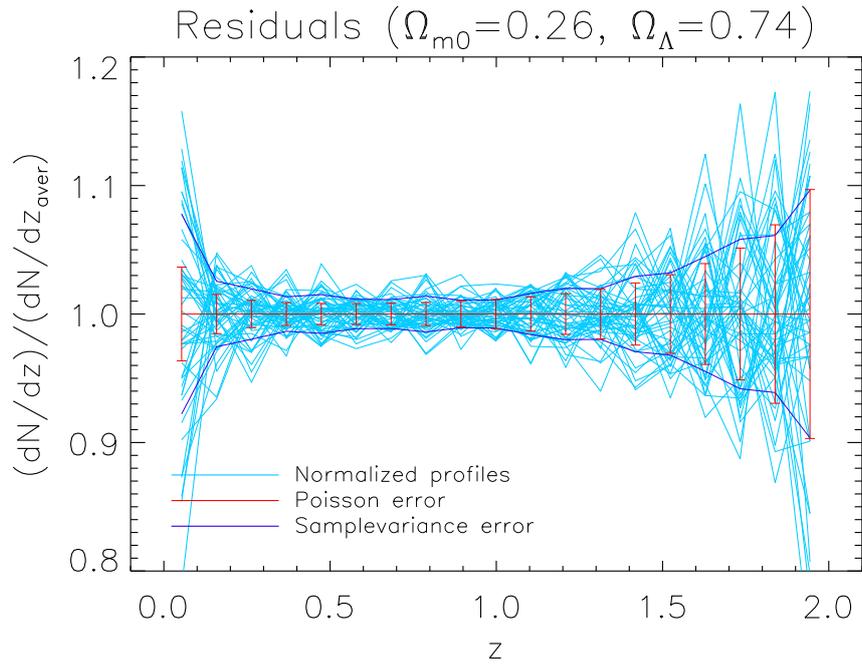
(b)

Figure 28: Behaviour of the $(dN/dz)_i$ (on the top), normalized profiles of $(\frac{dn}{dz})_{norm,i}$ for different 50 realisation (on the bottom) as a function of redshift bin. The sample variance is in in blu lines and the Poisson noise in red bars.

Taking the lambda cold dark model ($\Omega_{mo} = 0.26$, $\Omega_{\Lambda o} = 0.74$, $\sigma_8 = 0.8$, random seed 23489) as model of reference we have studied other cosmological models, for which we have fixed $\Omega_b = 0.044$, $h = 0.7$, while we have choosen Ω_{mo} and σ_8 values refering to the relation given by Vikhlinin et al. (2009):

$$\sigma_8 = 0.813 \left(\frac{\Omega_{mo}}{0.25} \right)^{0.47} \tag{4.14}$$

We have studied the following models:

- $\Omega_{mo} = 0.26$, $\Omega_{\Lambda o} = 0.0$, $\sigma_8 = 0.8$ $(0.26 - 0.0 - 0.8)$;

- $\Omega_{mo} = 0.26$, $\Omega_{\Lambda o} = 0.6$, $\sigma_8 = 0.8$ $(0.26 - 0.6 - 0.8)$;

- $\Omega_{mo} = 0.26$, $\Omega_{\Lambda o} = 0.9$, $\sigma_8 = 0.8$ $(0.26 - 0.9 - 0.8)$;

- $\Omega_{mo} = 0.2$, $\Omega_{\Lambda o} = 0.74$, $\sigma_8 = 0.9$ $(0.2 - 0.74 - 0.9)$;

- $\Omega_{mo} = 0.3$, $\Omega_{\Lambda o} = 0.74$, $\sigma_8 = 0.74$ $(0.3 - 0.74 - 0.74)$;

- $\Omega_{mo} = 0.3$, $\Omega_{\Lambda o} = 0.7$, $\sigma_8 = 0.74(0.3 - 0.7 - 0.74)$;

- $\Omega_{mo} = 0.2$, $\Omega_{\Lambda o} = 0.8$, $\sigma_8 = 0.9$ $(0.2 - 0.8 - 0.9)$;

- $\Omega_{mo} = 0.5$, $\Omega_{\Lambda o} = 0.5$, $\sigma_8 = 0.58$ $(0.5 - 0.5 - 0.58)$;

We have consider for each model only one realization with the random seed of reference (23489) for the generation of the density distribution in the superbox. For the valuation of the $(dN/dz)_i$ we have choosen one single direction, the same for each of them, the $(\frac{dN}{dz})_{norm,i}$ are normalized to the mean value calculated over the 50 realizations of the cosmological standard models we have considered before. Poisson noise and the sample variance are evaluated over the same 50 realisations with the same analysis described by (4.9), (4.7) equations for the $(dN/dz)_i$ and by (4.12), (4.11) equations for $(\frac{dN}{dz})_{norm,i}$.

In figure 29 we show on the top the $(dN/dz)_i$, on the bottom the normalized profiles of each flat model as a function of the redshift bin. The dotted blue lines represent the standard cold dark matter model, dotted-dotted-dotted-dashed lines represent the $0.2 - 0.8 - 0.9$ model, dotted-dashed lines represent the $0.3 - 0.7 - 0.74$ model and dashed lines represent the $0.5 - 0.5 - 0.58$ model. With cyan lines we have reported the sample variance and with green lines the 3σ.

Looking at the top of the panel, we see how the peaks of $(dN/dz)_i$ are shifted toward lower redshift at the increasing of Ω_{mo}.

Looking at the bottom panel we note how all the curves spread out significantly from the 3σ at about redshift $0.2 - 0.3$. This means that this statistics permits to distinguish the slight variation of cosmological parameters when considering the same geometry of the Universe.

Similar considerations can be reached for the closed and open models. In figure 30 we show the results for closed models. On the top we have reported the behaviour of the $(dN/dz)_i$, and on the bottom the normalized profiles of $(\frac{dn}{dz})_{norm,i}$ as a function of redshift bin. Dotted blue lines represent the
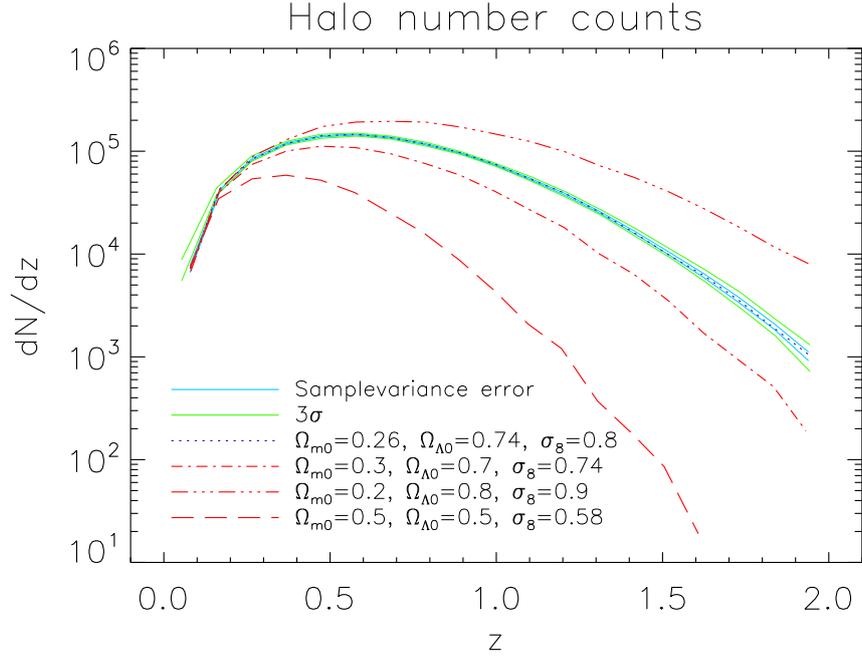
(a)
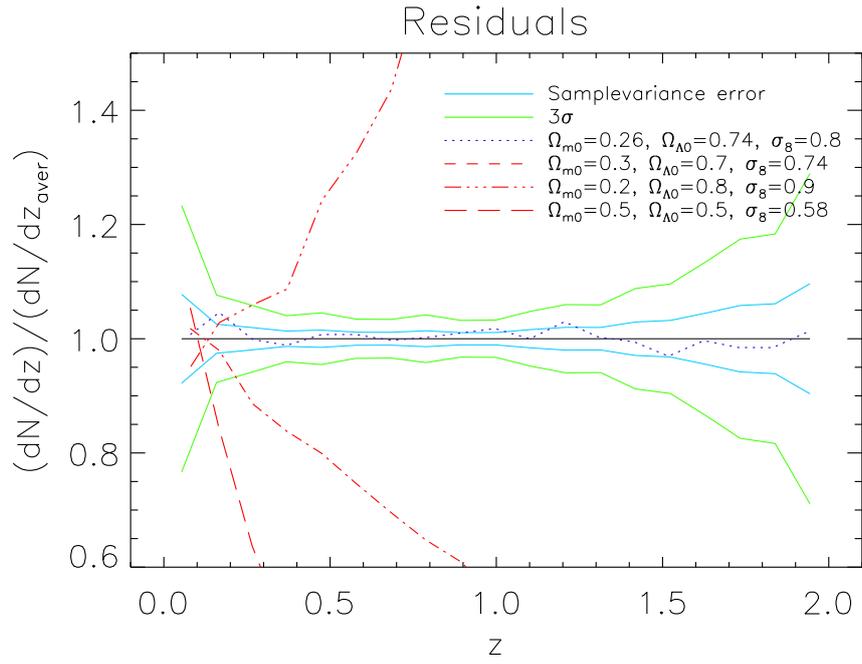
(b)

Figure 29: Behaviour of the $(dN/dz)_i$ (on the top), normalized profiles of $(\frac{dn}{dz})_{norm,i}$ for different flat models (on the bottom) as a function of redshift bin. Dotted blue lines represent the standard cold dark matter model, with cosmological parameters $\Omega_{m0} = 26$, $\Omega_{\Lambda 0} = 0.74$, $\sigma_8 = 0.8$, dotted-dotted-dotted-dashed lines represent the $\Omega_{m0} = 0.2$, $\Omega_{\Lambda 0} = 0.8$, $\sigma_8 = 0.9$ model, dotted-dashed lines represent the $\Omega_{m0} = 0.3$, $\Omega_{\Lambda 0} = 0.7$, $\sigma_8 = 0.74$ model and dashed lines represent the $\Omega_{m0} = 0.5$, $\Omega_{\Lambda 0} = 0.5$, $\sigma_8 = 0.58$ model. With cyan lines we have reported the sample variance and with green lines the 3σ.

reference cold dark matter model, dashed lines represent the $0.26 - 0.9 - 0.8$ model, dotted-dashed lines represent the $0.3 - 0.74 - 0.74$ model. With cyan lines we have reported the sample variance and with green lines the 3σ.

Looking at the top of the panel, even this time, we see how the peaks of $(dN/dz)_i$ are shifted toward lower redshift at the increasing of Ω_{mo}. This effect is more evident for the model with $\Omega_{mo} = 0.26$ whose behavior distingueshes from the one of standard model, only at $z = 0.9$. The same result is confirmed looking at the normalized profiles, model with $0.3 - 0.74 - 0.74$ separates iteself from the 3σ at $z \sim 0.3$ while the model $0.26 - 0.9 - 0.8$ at $z \sim 0.8$.

For open models, reported in figure 31, we are able to distinguish variation of parameters at very low redshift for the $0.26 - 0.0 - 0.8$ model, that move away from the 3σ at redshift ~ 0.1 or for the $0.2 - 0.74 - 0.8$.

We have verified until now how number counts from PINOCCHIO past-light-cone catalogues respond to the variation of cosmological parameters.

To better investigate the sensitiviness of the number counts to cosmological parameters we have decided to groups models fixing one parameter at time and varying the others.

We have started fixing Ω_{mo}. Looking at the top panel of figure 32 we note that at the increasing of the $\Omega_{\Lambda o}$ parameter, the $(dN/dz)_i$ decrease for redshift greater than $z \sim 0.8$, and increase for redshift smaller than $z \sim 0.8$. The bottom figure of the same panel show more clearly the behaviour of the models. $z \sim 0.8$ is the redshift after which all the different models are definitely out of the 3σ.

Model for which we fix $\Omega_{\Lambda o}$ parameters, shown in figure 33 are sensitive for variation of the order of 1 per cent on the Ω_{mo} parameter. Looking at the bottom of the panel we note that models with $\Omega_{mo} = 0.2; 0.3$ have a significant difference in the $(\frac{dN}{dz})_{norm,i}$, of ~ 20 per cent at redshift ~ 0.4. At the same redshift, models with fix Ω_{mo}, has the same discrepancy but for a greater variation of the parameter $\Omega_{\Lambda o} = 0.0$, $\Omega_{\Lambda o} = 0.9$. That could mean that number counts are more sensitive to a variation of Ω_{mo}, rather than a variation of the $\Omega_{\Lambda o}$ parameter. That is in part what Vikhlinin et al. (2009) assumes. It is confirmed by the resuls shown in the next graphs in figure 34, where we have varied only the value of σ_8, considering, in addition to the stardard model, the following:

- $\Omega_{mo} = 0.26$, $\Omega_{\Lambda o} = 0.74$, $\sigma_8 = 0.7$;

- $\Omega_{mo} = 0.26$, $\Omega_{\Lambda o} = 0.74$, $\sigma_8 = 0.9$;

Despite the other models in which number counts are within the sample variance uncertainity at redshift closed to 0, differentiating later with with respect to the reference model, here number counts are shifted with respect the concordance model as we see at the top and more clearly at the bottom of the figure 34. In particular the model with greater value of σ_8 has, at $z \sim 0$, has normalized profile far from unit of about 40 per cent. The situation is specular at $z \sim 0$, for the normalized profile with lower value of σ_8. This is reasonable because increasing or decreasing σ_8 means to give more or less power to the spectrum of fluctuation to a given scale.

Halo number counts

(a)

Residuals
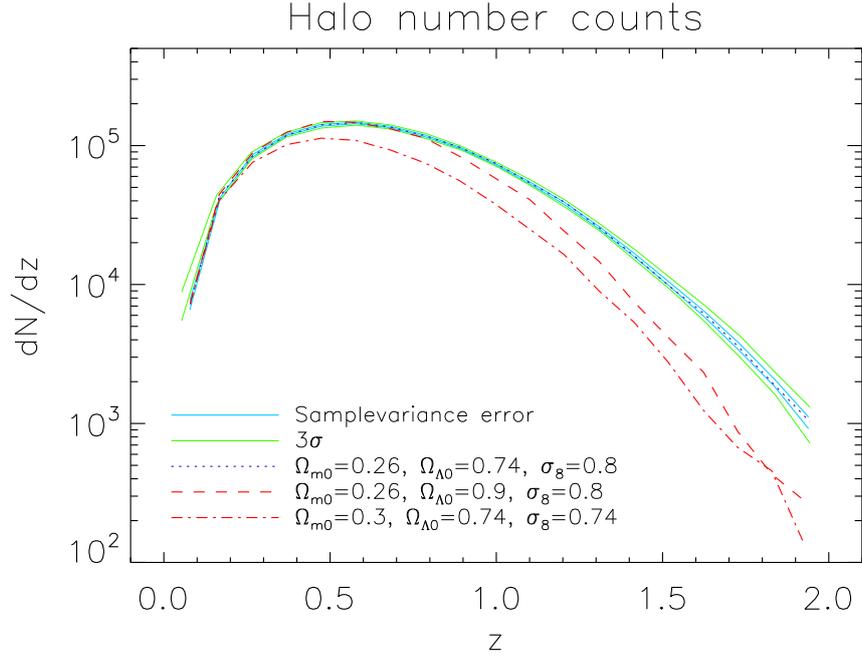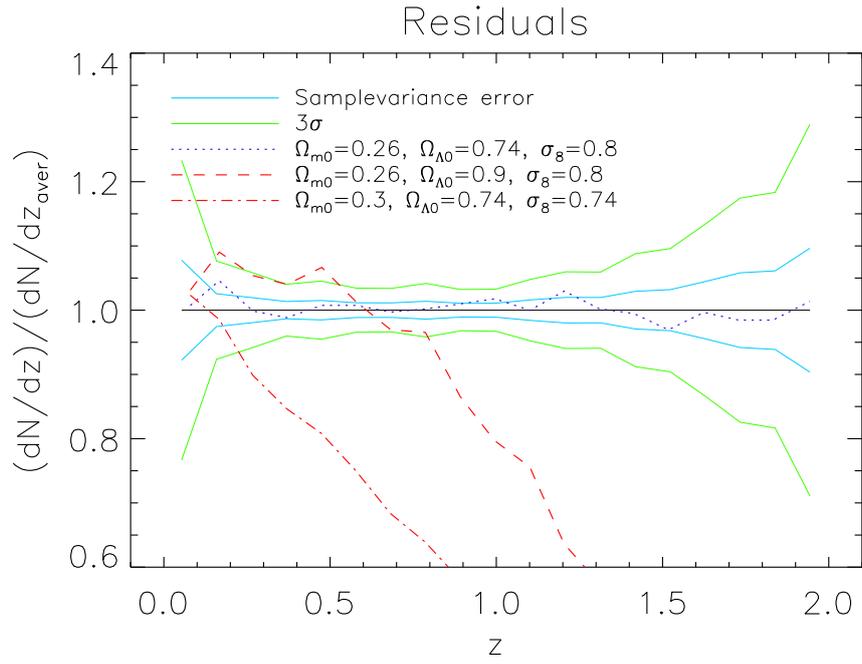
(b)

Figure 30: Behaviour of the $(dN/dz)_i$ (on the top), normalized profiles of $\left(\frac{dn}{dz}\right)_{norm,i}$ for different closed models (on the bottom) as a function of redshift bin. Dotted blue lines represent the standard cold dark matter model, with cosmological parameters $\Omega_{m0} = 26$, $\Omega_{\Lambda 0} = 0.74$, $\sigma_8 = 0.8$, dashed lines represent the $\Omega_{m0} = 0.26$, $\Omega_{\Lambda 0} = 0.9$, $\sigma_8 = 0.8$ model, dotted-dashed lines represent the $\Omega_{m0} = 0.3$, $\Omega_{\Lambda 0} = 0.74$, $\sigma_8 = 0.74$ model. With cyan lines we have reported the sample variance and with green lines the 3σ.
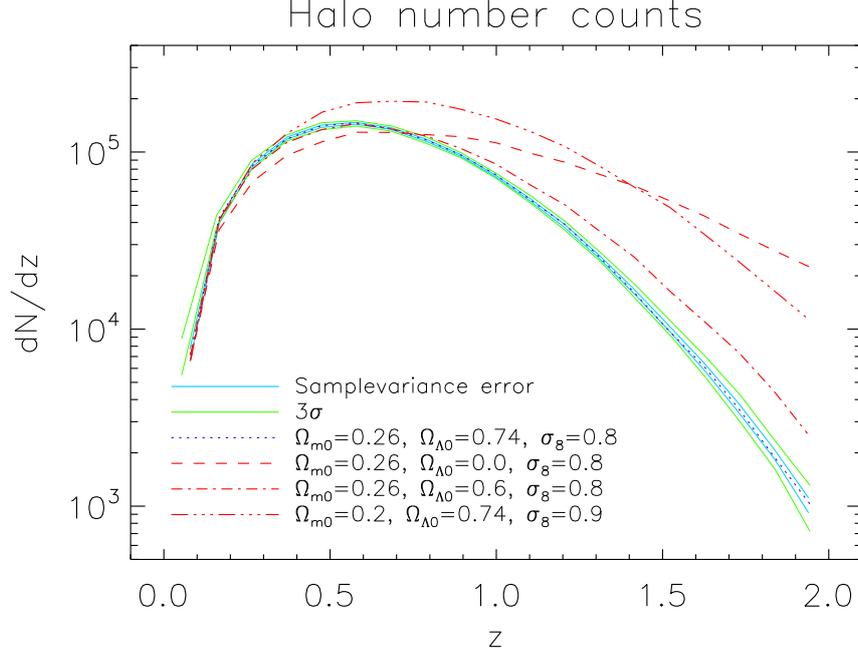
(a)
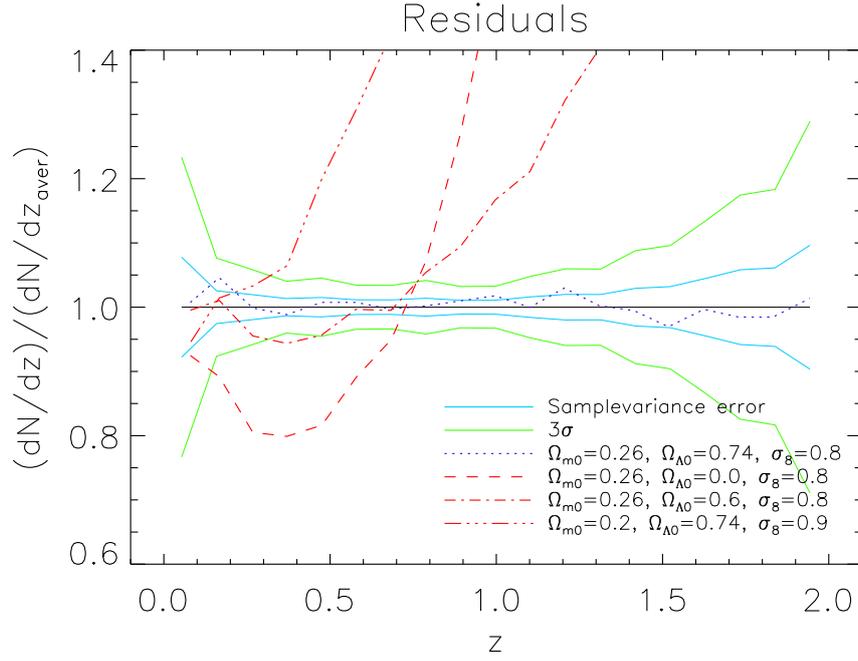
(b)

Figure 31: Behaviour of the $(dN/dz)_i$ (on the top), normalized profiles of $(\frac{dn}{dz})_{norm,i}$ for different open models (on the bottom) as a function of redshift bin. Dotted blue lines represent the standard cold dark matter model, with cosmological parameters $\Omega_{mo} = 26$, $\Omega_{\Lambda o} = 0.74$, $\sigma_8 = 0.8$, dashed lines represent the $\Omega_{mo} = 0.26$, $\Omega_{\Lambda o} = 0.0$, $\sigma_8 = 0.8$ model, dotted-dotted-dotted-dashed lines represent the $\Omega_{mo} = 0.2$, $\Omega_{\Lambda o} = 0.74$, $\sigma_8 = 0.8$ model, dotted-dashed lines represent the $\Omega_{mo} = 0.26$, $\Omega_{\Lambda o} = 0.6$, $\sigma_8 = 0.8$ model. With cyan lines we have reported the sample variance and with green lines the 3σ.

(a)

(b)

Figure 32: Behaviour of the $(dN/dz)_i$ (on the top), normalized profiles of $(\frac{dn}{dz})_{norm,i}$ for different models (on the bottom), with fix Ω_{m0}, as a function of redshift bin. Dotted blue lines represent the standard cold dark matter model, with cosmological parameters $\Omega_{m0} = 26$, $\Omega_{\Lambda 0} = 0.74$, $\sigma_8 = 0.8$, short dashed lines represent the $\Omega_{m0} = 0.26$, $\Omega_{\Lambda 0} = 0.0$, $\sigma_8 = 0.8$ model, long dashed lines represent the $\Omega_{m0} = 0.26$, $\Omega_{\Lambda 0} = 0.9$, $\sigma_8 = 0.8$ model, dotted-dashed lines represent the $\Omega_{m0} = 0.26$, $\Omega_{\Lambda 0} = 0.6$, $\sigma_8 = 0.8$ model. With cyan lines we have reported the sample variance and with green lines the 3σ
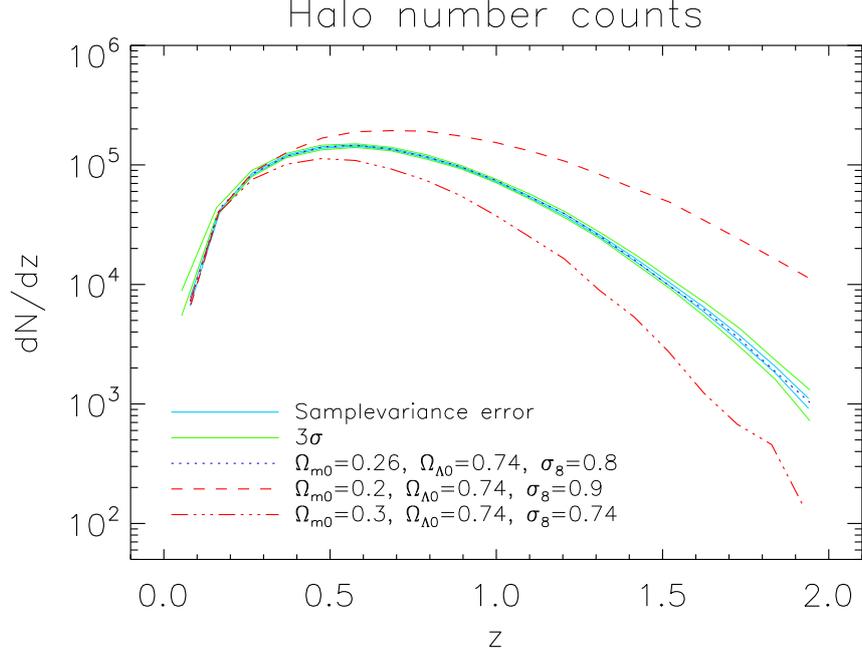
(a)
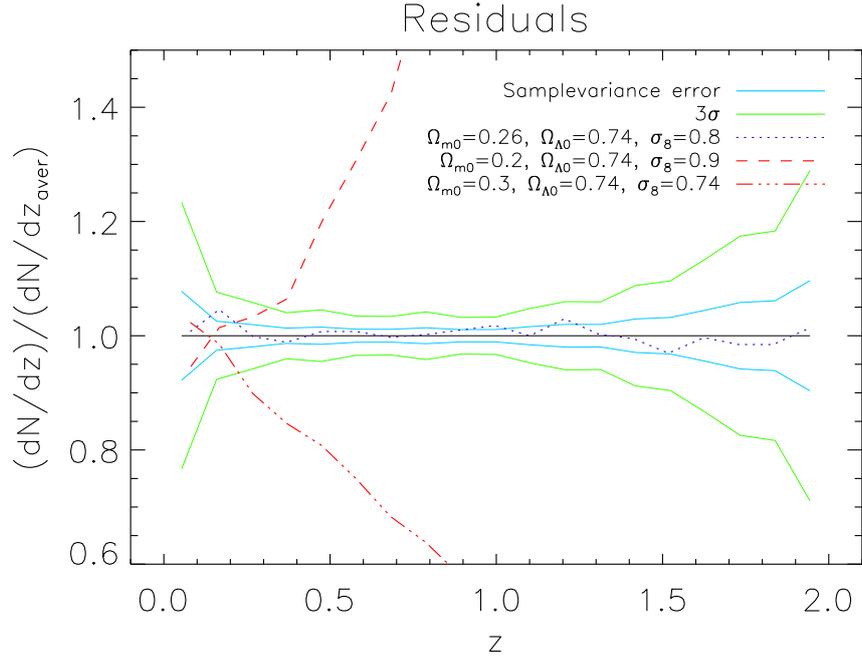
(b)

Figure 33: Behaviour of the $(dN/dz)_i$ (on the top), normalized profiles of $(\frac{dn}{dz})_{norm,i}$ for different models (on the bottom) with different value of σ_8, as a function of redshift bin. Dotted blue lines represent the standard cold dark matter model, with cosmological parameters $\Omega_{mo} = 26$, $\Omega_{\Lambda o} = 0.74$, $\sigma_8 = 0.8$, dashed lines represent the $\Omega_{mo} = 0.26$, $\Omega_{\Lambda o} = 0.74$, $\sigma_8 = 0.7$ model, dotted-dotted-dotted-dashed lines represent the $\Omega_{mo} = 0.26$, $\Omega_{\Lambda o} = 0.74$, $\sigma_8 = 0.9$ model. With cyan lines we have reported the sample variance and with green lines the 3σ

(a)

(b)

Figure 34: Behaviour of the $(dN/dz)_i$ (on the top), normalized profiles of $(\frac{dn}{dz})_{norm,i}$ for different values of σ_8 (on the bottom) as a function of redshift bin. Dotted blue lines represent the standard cold dark matter model, with cosmological parameters $\Omega_{m0} = 26$, $\Omega_{\Lambda 0} = 0.74$, $\sigma_8 = 0.8$, short dashed lines represent the $\Omega_{m0} = 0.26$, $\Omega_{\Lambda 0} = 0.74$, $\sigma_8 = 0.7$ model, long dashed lines represent the $\Omega_{m0} = 0.26$, $\Omega_{Lambda0} = 0.74$, $\sigma_8 = 0.9$ model. With cyan lines we have reported the sample variance and with green lines the 3σ.

The last analysis we have done, consists in appreciating the variation of number counts, changing the contribution of the dark energy for a model with $\Omega_{mo} = 26$, $\Omega_{DE} = 0.74$, $\sigma_8 = 0.8$. Considering the equation of state (1.12) we have evaluated two cases:

- $w_0 = -1$, $w_1 = 0.1$

- $w_0 = -0.9$, $w_1 = 0.1$

Number counts obtained by the first model are within the uncertainity of the 3σ and for low redshift (until $z \sim 1$) within the uncertainity of sample variance as it's possible to observe on the bottom of panel shown in figure 35. The normalized profile follows the same curve of the standard concordance model unless of 0.5 per cent. The sensitiviness to the parameters is more evident in the second model considered, whose normalised profile separetes iteself from 3σ at $z \sim 1$ to grow more rapidly.

This allows to forecast which cosmological models are distinguishable from the standard one, and eventually beyond which redshift, using the number counts technique.
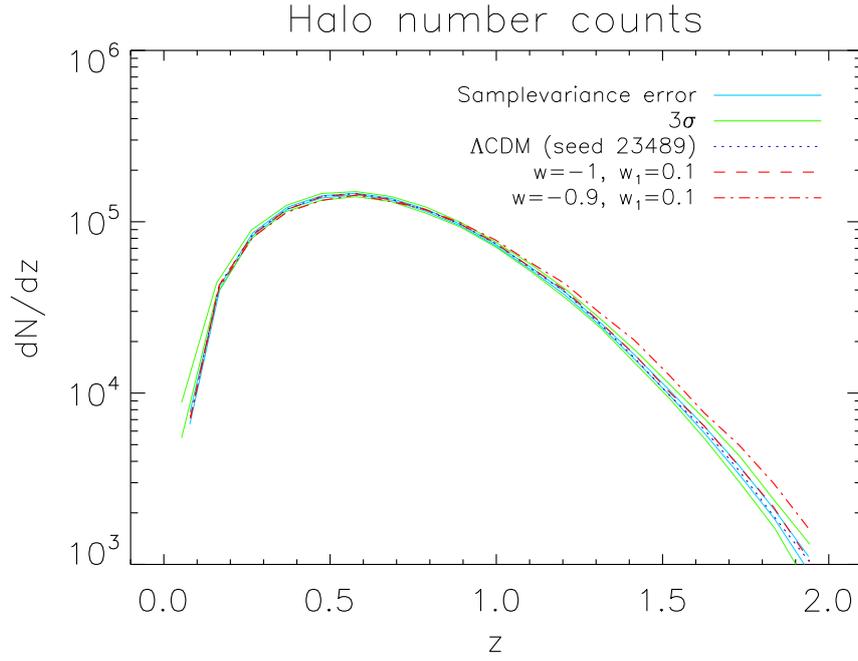
(a)
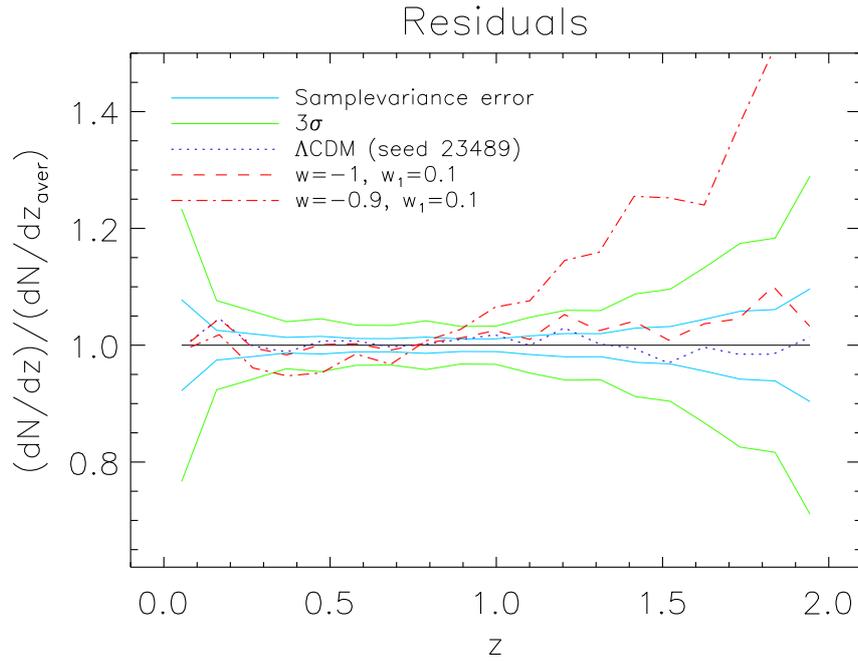
(b)

Figure 35: Behaviour of the $(dN/dz)_i$ (on the top), normalized profiles of $(\frac{dn}{dz})_{norm,i}$ for different open models (on the bottom) as a function of redshift bin. Dotted blue lines represent the standard cold dark matter model, with cosmological parameters $\Omega_{mo} = 26$, $\Omega_{\Lambda o} = 0.74$, $\sigma_8 = 0.8$, dashed lines represent the model with $w_0 = -1$, $w_1 = 0.1$, dotted-dashed lines represent the model with $w_0 = -0.9$, $w_1 = 0.1$. With cyan lines we have reported the sample variance and with green lines the 3σ.

CONCLUSIONS

The aim of this Thesis is the development of the PINOCCHIO code following two main directions of development. The first direction aims at fixing particular limitations of the code to contribute to an improved generality and versatility of the code. The development in the second direction is the creation of past-light-cone catalogues of dark matter halos. These catalogues allow to estimate the uncertainties of the observable quantities, mimicking the configuration of the real catalogues, that will be produced by future surveys like Euclid. In particular we want to understand to what extent the sample variance and poissonian noise affect the number counts of halos, which is aimed to determine the cosmological parameters. To generate these catalogues we used PINOCCHIO code, which has the advantage of generating, with small computational resources, large catalogues of dark matter halos. PINOCCHIO is an approximated approach based on the Lagrangian Perturbation Theory for the computation of the displacement of mass elements from their initial position, and is able to accurately reproduce the density field of matter on large scale.

The original version of PINOCCHIO used analytical solutions for the computation of cosmological quantities, that did not allow to apply the code to a generic cosmology. To overtake this limitation we implemented numerical solutions to cosmological quantities, namely the first order linear growth rate and the proper distance as a function of the scale factor.

We checked the accuracy of these numerical solutions to quantify their reliability by comparing them to analytical solutions in the case in which they are avaible. We built these solutions by requiring a minimum relative error of 10^{-4} between the numerical and the analytical solutions. The first order growth rate test results in an accuracy of $10^{-4} - 10^{-8}$ depending on the considered cosmological model. The same test has been performed on the cosmic time solution, resulting in an accuracy of $10^{-6} - 10^{-8}$.

We introduced, in the cosmological routines of PINOCCHIO, the numerical solution of the second order LPT growth rate in order to improve the accuracy with which PINOCCHIO is able to reproduce the distribution of dark matter halos. We tested the accuracy in the only case (EdS model) where the analytical solution is admitted. The accuracy results to be $\sim 10^{-7}$.

Given the good accuracy obtained with these numerical implementations, we consider them to be reliable for any cosmology. These modifications have improved the versatility of PINOCCHIO, making it an ideal tool for building

past light cones catalogues in any cosmological model. In these catalogues dark matter halos are stored in the catalogue according to the epoch at which they first cross the observer's past light cone. In order to generate a cosmological volume that is of sufficient size to fully contain any survey that extends to $z = 2$ and we tiled 5 replications of a simulation box of 3 Gpc/h, for a grid size of 1200 particle, reaching the comoving radial distance of 7.5 Gpc/h, for an observer placed in the center of the box. The past-light-cone is generated to have the observer in a random position within the superbox. We have constructed a full sky past-light-cone in order to have the possibility of cosidering catalogues for each desired geometry of the survey.

Different cosmological applications are possible with these kinds of catalogues. In this Thesis we focused on the estimate of the two main sources of uncertainties in the dark matter halo number counts on a single realization of a volume of the Universe, namely the sample variance and the Poissonian noise. We have first generated 10 realizations of the reference cosmological model varying the random seed for building the initial density field, and computed the sample variance and the poissonian error of the number counts. With the intention of emulating the volume of the Universe that Euclid will cover, we also computed the number counts within a solid angle of fixed aperture of $\sim 1.29 rad$, pointing along one random direction of the sky. We considered 5 different random directions for each of the 10 realizations. This results in a total sample of 50 realizations, for which we computed the sample variance and the poissonian error. In both cases the sample variance appeared to be greater that the poissonian error. This is important to assign a reliable uncertainty to the number count statistics and therefore to the cosmological information one can extract from it. In particular the sample variance defines the range of number counts within which a cosmological model is compatible with the standard one.

This allows to forecast which cosmological models are distinguishable from the standard one, and eventually beyond which redshift, using the number counts technique. Many different cosmological models have been compared to the standard one, both open, flat and closed.

This put in evidence one of the main advantages of PINOCCHIO. The speed of computation and low memory request allow to explore wide regions of the cosmological parameter space, as well as considering very large spatial volumes. Such a versatility is not within N-body simulations reach; these, on the other hand, are more precise in reproducing the details of the physical processes.

The work carried out so far is based on the analysis of the statistics of dark matter halos. In order to reliably compare with observations, some recipies for populating such halos with galaxies must be considered. Semi-analytical models appear to be a good tool for this purpuse, but they would require such a high mass resolution that they would not take advantage of the peculiarities of PINOCCHIO. Around 6000^3 particles would be required to use a semi-analytical model on the PINOCCHIO outputs, making the code highly resource consuming. Nevertheless other methods for populating dark matter halos that require less computing resources exist. E.g. the halo occupation distribution (HOD) would require around 2500^3 particles, making it feasible to be applied to

the PINOCCHIO outputs. We are planning to consider it as a development of the code.

Another direction of development deals with the specific reproduction of the Euclid observational configuration, as far as the geometry of the observed volume is concerned. Also, a more detailed halo mass selection function would allow to estimate the sample variance more accurately and this, in turn, would allow to distinguish the different cosmological models in an even more reliable way. In fact with this accurate selection function, a high number of cosmological parameters can be explored and reliable confidence regions in the parameter space can be estimated.

Once we have these more refined information, it will be possible to perform studies on the matter power spectrum as well as the correlation function, still preserving the advantages of using the light cones that mimick the observational configuration.

J. Binney and S. Tremaine. *Galactic dynamics*. 1987.

J. R. Bond and S. T. Myers. The Peak-Patch Picture of Cosmic Catalogs. I. Algorithms. *Astrophysical Journal Supplement*, 103:1, March 1996a. doi: 10.1086/192267.

J. R. Bond and S. T. Myers. The Peak-Patch Picture of Cosmic Catalogs. II. Validation. *Astrophysical Journal Supplement*, 103:41, March 1996b. doi: 10.1086/192268.

J. R. Bond and S. T. Myers. The Peak-Patch Picture of Cosmic Catalogs. III. Application to Clusters. *Astrophysical Journal Supplement*, 103:63, March 1996c. doi: 10.1086/192269.

J. R. Bond, S. Cole, G. Efstathiou, and N. Kaiser. Excursion set mass functions for hierarchical Gaussian fluctuations. *Astrophysical Journal*, 379:440–460, October 1991. doi: 10.1086/170520.

S. Borgani and L. Guzzo. X-ray clusters of galaxies as tracers of structure in the Universe. *Nature*, 409:39–45, January 2001. doi: 10.1038/35051000.

S. Borgani, P. Coles, and L. Moscardini. Cluster correlations in the Zel'dovich approximation. *Montly Notices of the Royal Astronomical Society*, 271:223, November 1994.

F. R. Bouchet. Introductory Overview of Eulerian and Lagrangian Perturbation Theories. In S. Bonometto, J. R. Primack, and A. Provenzale, editors, *Dark Matter in the Universe*, page 565, 1996.

T. Buchert and J. Ehlers. Lagrangian theory of gravitational instability of Friedman-Lemaitre cosmologies – second-order approach: an improved model for non-linear clustering. *Montly Notices of the Royal Astronomical Society*, 264:375–387, September 1993.

T. Buchert, A. L. Melott, and A. G. Weiss. Testing higher-order Lagrangian perturbation theory against numerical simulations I. Pancake models. *Astronomy and Astrophysics*, 288:349–364, August 1994.

P. Catelan. Lagrangian dynamics in non-flat universes and non-linear gravitational evolution. *Montly Notices of the Royal Astronomical Society*, 276:115–124, September 1995.

S. Cole, S. Hatton, D. H. Weinberg, and C. S. Frenk. Mock 2dF and SDSS galaxy redshift surveys. *Montly Notices of the Royal Astronomical Society*, 300:945–966, November 1998. doi: 10.1046/j.1365-8711.1998.01936.x.

H. M. P. Couchman. Mesh-refined P3M - A fast adaptive N-body algorithm. *Astrophysical Journal Letters*, 368:L23–L26, February 1991. doi: 10.1086/185939.

A. E. Evrard, T. J. MacFarland, H. M. P. Couchman, J. M. Colberg, N. Yoshida, S. D. M. White, A. Jenkins, C. S. Frenk, F. R. Pearce, J. A. Peacock, and P. A. Thomas. Galaxy Clusters in Hubble Volume Simulations: Cosmological Constraints from Sky Survey Populations. *Astrophysical Journal*, 573:7–36, July 2002. doi: 10.1086/340551.

F. Governato, A. Babul, T. Quinn, P. Tozzi, C. M. Baugh, N. Katz, and G. Lake. Properties of galaxy clusters: mass and correlation functions. *Montly Notices of the Royal Astronomical Society*, 307:949–966, August 1999. doi: 10.1046/j. 1365-8711.1999.02706.x.

T. Hamana, N. Yoshida, Y. Suto, and A. E. Evrard. Clustering of Dark Matter Halos on the Light Cone: Scale, Time, and Mass Dependence of the Halo Biasing in the Hubble Volume Simulations. *Astrophysical Journal Letters*, 561: L143–L146, November 2001. doi: 10.1086/324677.

L. Heisenberg, B. M. Schäfer, and M. Bartelmann. A study of relative velocity statistics in Lagrangian perturbation theory with PINOCCHIO. *Montly Notices of the Royal Astronomical Society*, 416:3057–3066, October 2011. doi: 10.1111/j. 1365-2966.2011.19252.x.

http://adlibitum.oats.inaf.it. Massive generation of mock catalogues with pinocchio.

http://msowww.anu.edu.au/2dFGRS/. The 2df galaxy redshift survey.

http://themaclellans.com/timetravel.html. Lightcone.

http://www.mpa garching.mpg.de. The millenium simulation.

http://www.mpa garching.mpg.de/gadget/. V. springel.

W. Hu and A. V. Kravtsov. Sample Variance Considerations for Cluster Surveys. *Astrophysical Journal*, 584:702–715, February 2003. doi: 10.1086/345846.

A. Jenkins, C. S. Frenk, S. D. M. White, J. M. Colberg, S. Cole, A. E. Evrard, H. M. P. Couchman, and N. Yoshida. The mass function of dark matter haloes. *Montly Notices of the Royal Astronomical Society*, 321:372–384, February 2001. doi: 10.1046/j.1365-8711.2001.04029.x.

S. Jouvel, J.-P. Kneib, O. Ilbert, G. Bernstein, S. Arnouts, T. Dahlen, A. Ealet, B. Milliard, H. Aussel, P. Capak, A. Koekemoer, V. Le Brun, H. McCracken, M. Salvato, and N. Scoville. Designing future dark energy space missions. I. Building realistic galaxy spectro-photometric catalogs and their first applications. *Astronomy and Astrophysics*, 504:359–371, September 2009. doi: 10.1051/0004-6361/200911798.

B. Lanzoni, G. A. Mamon, and B. Guiderdoni. Merging history trees for dark matter haloes: tests of the Merging Cell Model in a CDM cosmology. *Montly Notices of the Royal Astronomical Society*, 312:781–793, March 2000. doi: 10.1046/j.1365-8711.2000.03164.x.

M. Manera, R. Scoccimarro, W. J. Percival, L. Samushia, C. K. McBride, A. J. Ross, R. K. Sheth, M. White, B. A. Reid, A. G. Sánchez, R. de Putter, X. Xu, A. A. Berlind, J. Brinkmann, B. Nichol, F. Montesano, N. Padmanabhan, R. A. Skibba, R. Tojeiro, and B. A. Weaver. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: a large sample of mock galaxy catalogues. *ArXiv e-prints*, March 2012.

S. Matarrese, O. Pantano, and D. Saez. A Relativistic Approach to Gravitational Instability in the Expanding Universe - Second-Order Lagrangian Solutions. *Montly Notices of the Royal Astronomical Society*, 271:513, December 1994.

A. I. Merson, C. M. Baugh, J. C. Helly, V. Gonzalez-Perez, S. Cole, R. Bielby, P. Norberg, C. S. Frenk, A. J. Benson, R. G. Bower, C. G. Lacey, and C. d. P. Lagos. Lightcone mock catalogues from semi-analytic models of galaxy formation - I. Construction and application to the BzK colour selection. *ArXiv e-prints*, June 2012.

Houjun Mo, Frank Van den Bosch, and Simon White. *Galaxy Formation and Evolution*. Cambridge University Press, The Edinburgh Building, Cambridge CB2 8RU, UK, May 2010. ISBN 9780521857932.

P. Monaco. Non-spherical collapse and the mass function. page 407, 1994.

P. Monaco. The Mass Function of Cosmic Structures with Nonspherical Collapse. *Astrophysical Journal*, 447:23, July 1995. doi: 10.1086/175853.

P. Monaco. A Lagrangian Dynamical Theory for the Mass Function of Cosmic Structures - I. Dynamics. *Montly Notices of the Royal Astronomical Society*, 287:753–770, June 1997a.

P. Monaco. A Lagrangian dynamical theory for the mass function of cosmic structures - II. Statistics. *Montly Notices of the Royal Astronomical Society*, 290:439–455, September 1997b.

P. Monaco. The Cosmological Mass Function. *Fundamentals of Cosmic Physics*, 19:157–317, 1998.

P. Monaco and G. Murante. Cosmological mass function with 1D gravity. *Physical Review D*, 60(10):103502, November 1999. doi: 10.1103/PhysRevD.60.103502.

P. Monaco, T. Theuns, and G. Taffoni. The pinocchio algorithm: pinpointing orbit-crossing collapsed hierarchical objects in a linear density field. *Montly Notices of the Royal Astronomical Society*, 331:587–608, April 2002a. doi: 10.1046/j.1365-8711.2002.05162.x.

P. Monaco, T. Theuns, G. Taffoni, F. Governato, T. Quinn, and J. Stadel. Predicting the Number, Spatial Distribution, and Merging History of Dark Matter Halos. *Astrophysical Journal*, 564:8–14, January 2002b. doi: 10.1086/324182.

F. Moutarde, J.-M. Alimi, F. R. Bouchet, R. Pellat, and A. Ramani. Precollapse scale invariance in gravitational instability. *Astrophysical Journal*, 382:377–381, December 1991. doi: 10.1086/170728.

NASA/WMAP. The time-line of the universe.

J. L. Pauls and A. L. Melott. Hierarchical pancaking: why the Zel'dovich approximation describes coherent large-scale structure in N-body simulations of gravitational clustering. *Montly Notices of the Royal Astronomical Society*, 274: 99–109, May 1995.

F. Lucchin P.Coles. *The Origin and evolution of cosmic structures*. 2002.

J. A. Peacock. *Cosmological Physics*. January 1999.

P. J. E. Peebles. *The large-scale structure of the universe*. 1980.

P. J. E. Peebles. *Principles of Physical Cosmology*. 1993.

W. H. Press and P. Schechter. Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. *Astrophysical Journal*, 187:425–438, February 1974. doi: 10.1086/152650.

P. Rosati, S. Borgani, and C. Norman. The Evolution of X-ray Clusters of Galaxies. *Annual Review of Astronomy & Astrophysics*, 40:539–577, 2002. doi: 10.1146/annurev.astro.40.120401.150547.

W.Press S., Teukolsky, W. Vetterling, and B. Flannery. *Numerical Recipes in Fortran*. Cambridge University Press, The Pitt Building, Cambridge CB2 1RP, USA, 1992.

N. Seto. On the initial approximation in N-body simulations. In J. Paul, T. Montmerle, and E. Aubourg, editors, *19th Texas Symposium on Relativistic Astrophysics and Cosmology*, December 1998.

S. F. Shandarin and Y. B. Zeldovich. The large-scale structure of the universe: Turbulence, intermittency, structures in a self-gravitating medium. *Reviews of Modern Physics*, 61:185–220, April 1989. doi: 10.1103/RevModPhys.61.185.

R. K. Sheth and G. Tormen. Large-scale bias and the peak background split. *Montly Notices of the Royal Astronomical Society*, 308:119–126, September 1999. doi: 10.1046/j.1365-8711.1999.02692.x.

Taffoni, G. and Monaco, P. and Theuns, T. PINOCCHIO and the hierarchical build-up of dark matter haloes. *Montly Notices of the Royal Astronomical Society*, 333:623–632, July 2002. doi: 10.1046/j.1365-8711.2002.05441.x.

A. Vikhlinin, A. V. Kravtsov, R. A. Burenin, H. Ebeling, W. R. Forman, A. Hornstrup, C. Jones, S. S. Murray, D. Nagai, H. Quintana, and A. Voevodkin. Chandra Cluster Cosmology Project III: Cosmological Parameter Constraints. *Astrophysical Journal*, 692:1060–1074, February 2009. doi: 10.1088/0004-637X/692/2/1060.

J. V. Villumsen. A new hierachical particle-mesh code for very large scale cosmological N-body simulations. *Astrophysical Journal Supplement*, 71:407–431, November 1989. doi: 10.1086/191380.

H. Wang, H. J. Mo, X. Yang, and F. C. van den Bosch. Reconstructing the Initial Density Field of the Local Universe: Method and Test with Mock Catalogs. *ArXiv e-prints*, January 2013.

M. S. Warren, K. Abazajian, D. E. Holz, and L. Teodoro. Precision Determination of the Mass Function of Dark Matter Halos. *Astrophysical Journal*, 646:881–885, August 2006. doi: 10.1086/504962.

S. D. M. White and J. Silk. The growth of aspherical structure in the universe - Is the local supercluster an unusual system. *Astrophysical Journal*, 231:1–9, July 1979. doi: 10.1086/157156.

Y. B. Zel'dovich. Gravitational instability: An approximate theory for large density perturbations. *Astronomy and Astrophysics*, 5:84–89, March 1970.

Z. Zheng, J. L. Tinker, D. H. Weinberg, and A. A. Berlind. Do Distinct Cosmological Models Predict Degenerate Halo Populations? *Astrophysical Journal*, 575:617–633, August 2002. doi: 10.1086/341434.

We shall not cease from exploration,
and the end of all our exploring will be to arrive where we started and know the place
for the first time.
T. S. Eliot

Only those who will risk going too far can possibly find out how far one can go.
T. S. Eliot

Credete a chi cerca la verità, non a chi la trova.
Andrè Gide

La scienza non ha promesso la felicità, ma la verità.
La questione è sapere se con la verità si farà mai la felicità.
Emile Zolà

ACKNOWLEDGMENTS

devoted, going back along away friends, all important for me. Thank you, not in a particular order, Claudia, Quad, Omar, Evelina, Tanja, Raffa. Claudia and Omar to be my friends and my second parents!. Thank you Clà, you have always been with me, in the hard moments and in the funny momentes, thank you for your frankness and your example of determination. Thank you Omar for you great sympathy, for your advices, thank you to be a good hearted and..ok even for your grammatical corrections!. Thank you Eve to be a sunny person, for your smile and for the good energy you have always given to me, thank you for the amusing time given to me, even at work!. Thank you Tanja you are a great friend and a great compain during my studies, thank you for all the laughs and the funny experiences together everywhere. Thank you Q for our emphaty, to be in synk and connected with me, to be the strict and kindly mirror in which I always can see me, thank you for your thoughs. Thank you Raffa you have never adbandoned me, in spite of everything, you have listened to me, helped me, trusted in me, thank you for all the moments together I will never forget. Thanks to my recent friends but not less important Tullio, Valentina. Thank you Tullio, for your inifiny availability, for your unconditional affection, for your emotionality and for giving me a continuos comparison. Thank you Valentina for your goodness and generosity, thank you have been always kind with me. Thank you all the persons that have been part of this long universitarian experience. They all leave me a lot. Finally thank you Naima, Dexter Morgan, Giuseppe and the Muggia's Carnival, etnoblog, the white lights, the japan restaurant, Viale 39 (casa), the day after, the blasfemies, the rum, the vodka, the vodka redbull, the redbull, the green grass of the neighbours, the Mia Dolce Melassa Amara and . . . Thanks to Trieste where this chapter is started and concluded.

15/03/13