Understanding Galaxy Formation and Evolution

Vladimir Avila-Reese[1]

Instituto de Astronomía, Universidad Nacional Autónoma de México, A.P. 70-264, 04510, México,D.F. avila@astroscu.unam.mx

The old dream of integrating into one the study of micro and macrocosmos is now a reality. Cosmology, astrophysics, and particle physics intersect in a scenario (but still not a theory) of cosmic structure formation and evolution called Λ Cold Dark Matter (ΛCDM) model. This scenario emerged mainly to explain the origin of galaxies. In these lecture notes, I first present a review of the main galaxy properties, highlighting the questions that any theory of galaxy formation should explain. Then, the cosmological framework and the main aspects of primordial perturbation generation and evolution are pedagogically detached. Next, I focus on the "dark side" of galaxy formation, presenting a review on ΛCDM halo assembling and properties, and on the main candidates for non–baryonic dark matter. It is shown how the nature of elemental particles can influence on the features of galaxies and their systems. Finally, the complex processes of baryon dissipation inside the non–linearly evolving CDM halos, formation of disks and spheroids, and transformation of gas into stars are briefly described, remarking on the possibility of a few driving factors and parameters able to explain the main body of galaxy properties. A summary and a discussion of some of the issues and open problems of the ΛCDM paradigm are given in the final part of these notes.

1 Introduction

Our vision of the cosmic world and in particular of the whole Universe has been changing dramatically in the last century. As we will see, galaxies were repeatedly the main protagonist in the scene of these changes. It is about 80 years since E. Hubble established the nature of galaxies as gigantic self-bound stellar systems and used their kinematics to show that the Universe as a whole is expanding uniformly at the present time. Galaxies, as the building blocks of the Universe, are also tracers of its large–scale structure and of its evolution in the last 13 Gyrs or more. By looking inside galaxies we find that they are the arena where stars form, evolve and collapse in constant interaction with the interstellar medium (ISM), a complex mix of gas and plasma, dust, radiation, cosmic rays, and magnetics fields. The center of a significant fraction of galaxies harbor supermassive black holes. When these "monsters" are fed with infalling material, the accretion disks around them release, mainly through powerful plasma jets, the largest amounts of energy known in astronomical objects. This phenomenon of Active Galactic Nuclei (AGN) was much more frequent in the past than in the present, being the high–redshift quasars (QSO's) the most powerful incarnation of the AGN phenomenon. But the most astonishing surprise of galaxies comes from the fact that luminous matter (stars, gas, AGN's, etc.) is only a tiny fraction ($\sim 1-5\%$) of all the mass measured in galaxies and the giant halos around them. What this dark component of galaxies is made of? This is one of the most acute enigmas of modern science.

Thus, exploring and understanding galaxies is of paramount interest to cosmology, high–energy and particle physics, gravitation theories, and, of course, astronomy and astrophysics. As astronomical objects, among other questions, we would like to know how do they take shape and evolve, what is the origin of their diversity and scaling laws, why they cluster in space as observed, following a sponge–like structure, what is the dark component that predominates in their masses. By answering to these questions we would able also to use galaxies as a true link between the observed universe and the properties of the early universe, and as physical laboratories for testing fundamental theories.

The content of these notes is as follows. In §2 a review on main galaxy properties and correlations is given. By following an analogy with biology, the taxonomical, anatomical, ecological and genetical study of galaxies is presented. The observational inference of dark matter existence, and the baryon budget in galaxies and in the Universe is highlighted. Section 3 is dedicated to a pedagogical presentation of the basis of cosmic structure formation theory in the context of the Λ Cold Dark Matter (ΛCDM) paradigm. The main questions to be answered are: why CDM is invoked to explain the formation of galaxies? How is explained the origin of the seeds of present–day cosmic structures? How these seeds evolve?. In §4 an updated review of the main results on properties and evolution of CDM halos is given, with emphasis on the aspects that influence the propertied of the galaxies expected to be formed inside the halos. A short discussion on dark matter candidates is also presented (§§4.2). The main ingredients of disk and spheroid galaxy formation are reviewed and discussed in §5. An attempt to highlight the main drivers of the Hubble and color sequences of galaxies is given in §§5.3. Finally, some selected issues and open problems in the field are resumed and discussed in §6.

2 Galaxy properties and correlations

During several decades galaxies were considered basically as self–gravitating stellar systems so that the study of their physics was a domain of Galactic

Dynamics. Galaxies in the local Universe are indeed mainly conglomerates of hundreds of millions to trillions of stars *supported against gravity either by rotation or by random motions*. In the former case, the system has the shape of a *flattened disk*, where most of the material is on circular orbits at radii that are the minimal ones allowed by the specific angular momentum of the material. Besides, disks are dynamically fragile systems, unstable to perturbations. Thus, the mass distribution along the disks is the result of the specific angular momentum distribution of the material from which the disks form, and of the posterior dynamical (internal and external) processes. In the latter case, the shape of the galactic system is a concentrated *spheroid/ellipsoid*, with mostly (disordered) radial orbits. The spheroid is dynamically hot, stable to perturbations. Are the properties of the stellar populations in the disk and spheroid systems different?

Stellar populations

Already in the 40's, W. Baade discovered that according to the ages, metallicities, kinematics and spatial distribution of the stars in our Galaxy, they separate in two groups: 1) Population I stars, which populate the plane of the disk; their ages do not go beyond 10 Gyr –a fraction of them in fact are young ($\lesssim 10^6$ yr) luminous O,B stars mostly in the spiral arms, and their metallicites are close to the solar one, $Z \approx 2\%$; 2) Population II stars, which are located in the spheroidal component of the Galaxy (stellar halo and partially in the bulge), where velocity dispersion (random motion) is higher than rotation velocity (ordered motion); they are old stars (> 10 Gyr) with very low metallicities, on the average lower by two orders of magnitude than Population I stars. In between Pop's I and II there are several stellar subsystems. [1].

Stellar populations are true fossils of the galaxy assembling process. The differences between them evidence differences in the formation and evolution of the galaxy components. The Pop II stars, being old, of low metallicity, and dominated by random motions (dynamically hot), had to form early in the assembling history of galaxies and through violent processes. In the meantime, the large range of ages of Pop I stars, but on average younger than the Pop II stars, indicates a slow star formation process that continues even today in the disk plane. Thus, the common wisdom says that *spheroids form early in a violent collapse (monolithic or major merger), while disks assemble by continuous infall of gas rich in angular momentum, keeping a self–regulated SF process.*

[1] Astronomers suspect also the existence of non–observable Population III of pristine stars with zero metallicities, formed in the first molecular clouds $\sim 4\ 10^8$ yrs ($z \sim 20$) after the Big Bang. These stars are thought to be very massive, so that in scaletimes of 1Myr they exploded, injected a big amount of energy to the primordial gas and started to reionize it through expanding cosmological HII regions (see e.g., [20, 27] for recent reviews on the subject).

Interstellar Medium (ISM)

Galaxies are not only conglomerates of stars. The study of galaxies is incomplete if it does not take into account the ISM, which for late–type galaxies accounts for more mass than that of stars. Besides, it is expected that in the deep past, galaxies were gas–dominated and with the passing of time the cold gas was being transformed into stars. The ISM is a turbulent, non–isothermal, multi–phase flow. Most of the gas mass is contained in neutral instable HI clouds ($10^2 < T < 10^4$K) and in dense, cold molecular clouds ($T < 10^2$K), where stars form. Most of the volume of the ISM is occuppied by diffuse ($n \approx 0.1$cm^{-3}), warm–hot ($T \approx 10^4 - 10^5$K) turbulent gas that confines clouds by pressure. The complex structure of the ISM is related to (i) its peculiar thermodynamical properties (in particular the heating and cooling processes), (ii) its hydrodynamical and magnetic properties which imply development of turbulence, and (iii) the different energy input sources. The star formation unities (molecular clouds) appear to form during large–scale compression of the diffuse ISM driven by supernovae (SN), magnetorotational instability, or disk gravitational instability (e.g., [7]). At the same time, the energy input by stars influences the hydrodynamical conditions of the ISM: the star formation results self–regulated by a delicate energy (turbulent) balance.

Galaxies are true "ecosystems" where stars form, evolve and collapse in constant interaction with the complex ISM. Following a pedagogical analogy with biological sciences, we may say that the study of galaxies proceeded through taxonomical, anatomical, ecological and genetical approaches.

2.1 Taxonomy

As it happens in any science, as soon as galaxies were discovered, the next step was to attempt to classify these news objects. This endeavor was taken on by E. Hubble. The showiest characteristics of galaxies are the bright shapes produced by their stars, in particular those most luminous. Hubble noticed that by their external look (morphology), galaxies can be divided into three principal types: Ellipticals (E, from round to flattened elliptical shapes), Spirals (S, characterized by spiral arms emanating from their central regions where an spheroidal structure called bulge is present), and Irregulars (Irr, clumpy without any defined shape). In fact, the last two classes of galaxies are disk–dominated, rotating structures. Spirals are subdivided into Sa, Sb, Sc types according to the size of the bulge in relation to the disk, the openness of the winding of the spiral arms, and the degree of resolution of the arms into stars (in between the arms there are also stars but less luminous than in the arms). Roughly 40% of S galaxies present an extended rectangular structure (called bar) further from the bulge; these are the barred Spirals (SB), where the bar is evidence of disk gravitational instability.

From the physical point of view, the most remarkable aspect of the morphological Hubble sequence is the ratio of spheroid (bulge) to total luminosity.

This ratio decreases from 1 for the Es, to ~ 0.5 for the so–called lenticulars (S0), to $\sim 0.5 - 0.1$ for the Ss, to almost 0 for the Irrs. *What is the origin of this sequence? Is it given by nature or nurture? Can the morphological types change from one to another and how frequently they do it?* It is interesting enough that roughly half of the stars at present are in galaxy spheroids (Es and the bulges of S0s and Ss), while the other half is in disks (e.g., [11]), where some fraction of stars is still forming.

2.2 Anatomy

The morphological classification of galaxies is based on their external aspect and it implies somewhat subjective criteria. Besides, the "showy" features that characterize this classification may change with the color band: in blue bands, which trace young luminous stellar populations, the arms, bar and other features may look different to what it is seen in infrared bands, which trace less massive, older stellar populations. We would like to explore deeper the internal physical properties of galaxies and see whether these properties correlate along the Hubble sequence. Fortunately, this seems to be the case in general so that, in spite of the complexity of galaxies, some clear and sequential trends in their properties encourage us to think about regularity and the possibility to find driving parameters and factors beyond this complexity.

Figure 1 below resumes the main trends of the "anatomical" properties of galaxies along the Hubble sequence.

The advent of extremely large galaxy surveys made possible massive and uniform determinations of global galaxy properties. Among others, the Sloan Digital Sky Survey (SDSS[2]) and the Two–degree Field Galaxy Redshift Survey (2dFGRS[3]) currently provide uniform data already for around 10^5 galaxies in limited volumes. The numbers will continue growing in the coming years. The results from these surveys confirmed the well known trends shown in Fig. 1; moreover, it allowed to determine the distributions of different properties. Most of these properties present a *bimodal* distribution with two main sequences: the red, passive galaxies and the blue, active galaxies, with a fraction of intermediate types (see for recent results [68, 6, 114, 34, 127] and more references therein). The most distinct segregation in two peaks is for the specific star formation rate ($\dot{M}_s/M_s$); there is a narrow and high peak of passive galaxies, and a broad and low peak of star forming galaxies. The two sequences are also segregated in the luminosity function: the faint end is dominated by the blue, active sequence, while the bright end is dominated by the red, passive sequence. It seems that the transition from one sequence to the other happens at the galaxy stellar mass of $\sim 3 \times 10^{10} M_\odot$.

[2] *www.sdss.org/sdss.html*
[3] *www.aao.gov.au/2df/*

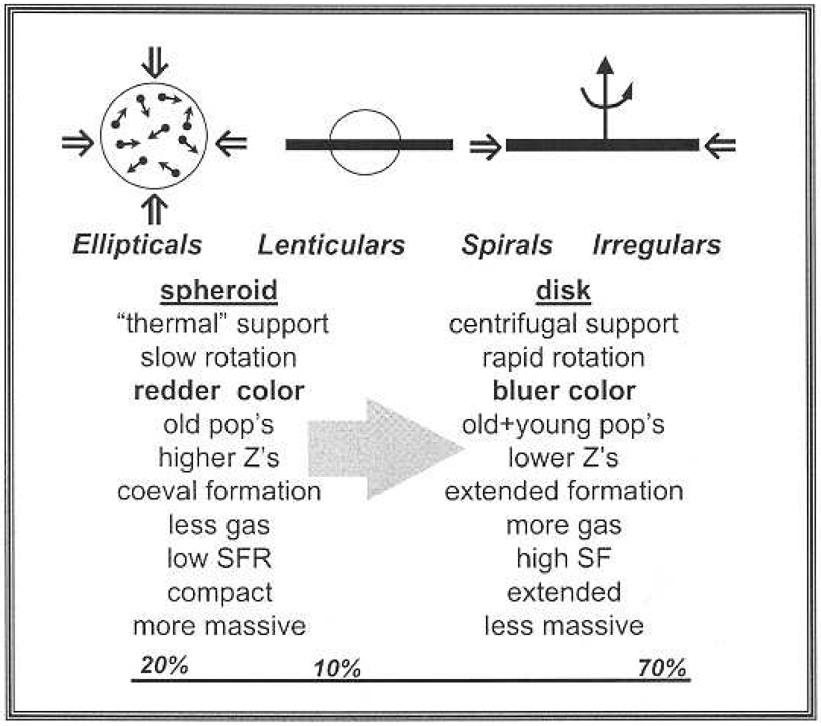

Fig. 1. Main trends of physical properties of galaxies along the Hubble morphological sequence. The latter is basically a sequence of change of the spheroid–to–disk ratio. Spheroids are supported against gravity by velocity dispersion, while disks by rotation.

The hidden component

Under the assumption of Newtonian gravity, the observed dynamics of galaxies points out to the presence of enormous amounts of mass not seen as stars or gas. Assuming that disks are in centrifugal equilibrium and that the orbits are circular (both are reasonable assumptions for non–central regions), the measured rotation curves are good tracers of the total (dynamical) mass distribution (Fig. 2). The mass distribution associated with the luminous galaxy (stars+gas) can be inferred directly from the surface brightness (density) profiles. For an exponential disk of scalelength R_d (=3 kpc for our Galaxy), the rotation curve beyond the optical radius ($R_{opt} \approx 3.2R_d$) decreases as in the Keplerian case. The observed HI rotation curves at radii around and beyond R_{opt} are far from the Keplerian fall–off, implying the existence of hidden mass called *dark matter (DM)* [99, 18]. The fraction of DM increases with radius.

It is important to remark the following observational facts:

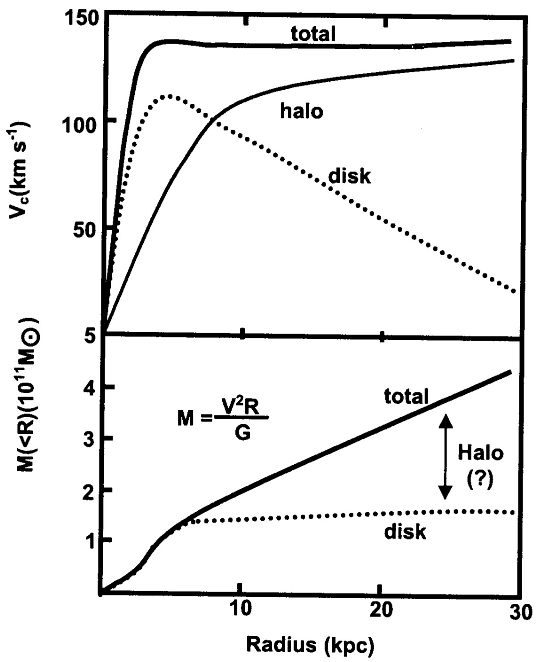

Fig. 2. Under the assumption of circular orbits, the observed rotation curve of disk galaxies traces the dynamical (total) mass distribution. The outer rotation curve of a nearly exponential disk decreases as in the Keplerian case. The observed rotation curves are nearly flat, suggesting the existence of massive dark halos.

• the *outer rotation curves are not universally flat* as it is assumed in hundreds of papers. Following, Salucci & Gentile [101], let us define the average value of the rotation curve logarithmic slope, $\bigtriangledown \equiv (dlogV/dlogR)$ between two and three R_d. A flat curve means $\bigtriangledown = 0$; for an exponential disk without DM, $\bigtriangledown = -0.27$ at $3R_s$. Observations show a large range of values for the slope: $-0.2 \le \bigtriangledown \le 1$

• the rotation curve shape ($\bigtriangledown$) correlates with the luminosity and surface brightness of galaxies [95, 123, 132]: it increases according the galaxy is fainter and of lower surface brightness

• at the optical radius R_{opt}, the DM–to–baryon ratio varies from $\approx$ 1 to 7 for luminous high–surface brightness to faint low–surface brightness galaxies, respectively

• the roughly smooth shape of the rotation curves implies a fine coupling between disk and DM halo mass distributions [24]

The HI rotation curves extend typically to $2 - 5R_{opt}$. The dynamics at larger radii can be traced with satellite galaxies if the satellite statistics allows for that. More recently, the technique of (statistical) *weak lensing* around galaxies began to emerge as the most direct way to trace the masses of galaxy halos. The results show that a typical L_* galaxy (early or late) with a stellar mass of $M_s \approx 6 \times 10^{10} \mathrm{M}_\odot$ is surrounded by a halo of $\approx 2 \times 10^{12} \mathrm{M}_\odot$ ([80] and more references therein). The extension of the halo is typically $\approx 200-250$kpc. These numbers are very close to the determinations for our own Galaxy.

The picture has been confirmed definitively: luminous galaxies are just the top of the iceberg (Fig. 3). The baryonic mass of (normal) galaxies is only $\sim 3-5\%$ of the DM mass in the halo! This fraction could be even lower for dwarf galaxies (because of feedback) and for very luminous galaxies (because the gas cooling time > Hubble time). On the other hand, the universal baryon–to–DM fraction ($\Omega_B/\Omega_{DM} \approx 0.04/0.022$, see below) is $f_{B,Un} \approx 18\%$. Thus, galaxies are not only dominated by DM, but are much more so than the average in the Universe! This begs the next question: if the majority of baryons is not in galaxies, where it is? Recent observations, based on highly ionized absorption lines towards low redshfit luminous AGNs, seem to have found a fraction of the missing baryons in the interfilamentary warm–hot intergalactic medium at $T \lesssim 10^5 - 10^7$ K [89].

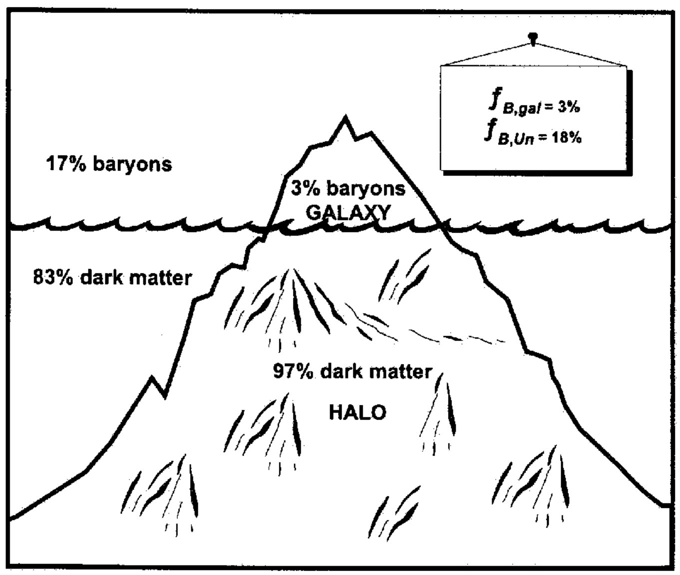

Fig. 3. Galaxies are just the top of the iceberg. They are surrounded by enormous DM halos extending 10–20 times their sizes, where baryon matter is only less than 5% of the total mass. Moreover, galaxies are much more DM–dominated than the average content of the Universe. The corresponding typical baryon–to–DM mass ratios are given in the inset.

Global baryon inventory: The different probes of baryon abundance in the Universe (primordial nucleosynthesis of light elements, the ratios of odd and even CMBR acoustic peaks heights, absorption lines in the Lyα forest) have been converging in the last years towards the same value of the baryon density: $\Omega_b \approx 0.042 \pm 0.005$. In Table 1 below, the densities (Ω's) of different baryon components at low redshfits and at $z > 2$ are given (from [48] and [89]).

Table 1. Abundances of the different baryon components ($h = 0.7$)

Component	Contribution to Ω
Low redshifts	
Galaxies: stars	0.0027 ± 0.0005
Galaxies: HI	$(4.2 \pm 0.7)\times10^{-4}$
Galaxies: H_2	$(1.6 \pm 0.6)\times10^{-4}$
Galaxies: others	$(\approx 2.0)\times10^{-4}$
Intracluster gas	0.0018 ± 0.0007
IGM: (cold-warm)	0.013 ± 0.0023
IGM: (warm-hot)	≈ 0.016
$z > 2$	
Lyα forest clouds	> 0.035

The present–day abundance of baryons in virialized objects (normal stars, gas, white dwarfs, black holes, etc. in galaxies, and hot gas in clusters) is therefore $\Omega_B \approx 0.0037$, which accounts for $\approx 9\%$ of all the baryons at low redshifts. The gas in not virialized structures in the Intergalactic Medium (cold-warm Lyα/β gas clouds and the warm–hot phase) accounts for $\approx 73\%$ of all baryons. Instead, at $z > 2$ more than 88% of the universal baryonic fraction is in the Lyα forest composed of cold HI clouds. The baryonic budget's outstanding questions: *Why only $\approx 9\%$ of baryons are in virialized structures at the present epoch?*

2.3 Ecology

The properties of galaxies vary systematically as a function of environment. The environment can be relatively local (measured through the number of nearest neighborhoods) or of large scale (measured through counting in defined volumes around the galaxy). The morphological type of galaxies is earlier in the locally denser regions (morphology–density relation),the fraction of ellipticals being maximal in cluster cores [40] and enhanced in rich [96] and poor groups. The extension of the morphology–density relation to low local–density environment (cluster outskirts, low mass groups, field) has been a matter of debate. From an analysis of SDSS data, [54] have found that (i) in the sparsest regions both relations flatten out, (ii) in the intermediate density regions (e.g., cluster outskirts) the intermediate–type galaxy (mostly S0s) fraction increases towards denser regions whereas the late–type galaxy fraction decreases, and (iii) in the densest regions intermediate–type fraction decreases radically and early–type fraction increases. In a similar way, a study based on 2dFGRS data of the luminosity functions in clusters and voids shows that the population of faint late–type galaxies dominates in the latter, while, in contrast, very bright early–late galaxies are relatively overabundant in the former [34]. This and other studies suggest that the origin of the morphology–density (or morphology-radius) relation could be a combination of (i) *initial (cosmological) conditions* and (ii) of *external mechanisms* (ram-pressure and tidal stripping, thermal evaporation of the disk gas, strangulation, galaxy harassment, truncated star formation, etc.) that operate mostly in dense environments, where precisely the relation steepens significantly.

The morphology–environment relation evolves. It systematically flattens with z in the sense that the grow of the early-type (E+S0) galaxy fraction with density becomes less rapid ([97] and more references therein) the main change being in the high–density population fraction. Postman et al. conclude that the observed flattening of the relation up to $z \sim 1$ is due mainly to a deficit of S0 galaxies and an excess of Sp+Irr galaxies relative to the local galaxy population; the E fraction-density relation does not appear to evolve over the range $0 < z < 1.3$! Observational studies show that other properties besides morphology vary with environment. The galaxy properties most sensitive to environment are the integral color and specific star formation rate (e.g. [68, 114, 127]. The dependences of both properties on environment extend typically to lower densities than the dependence for morphology. These properties are tightly related to the galaxy star formation history, which in turn depends on internal formation/evolution processes related directly to initial cosmological conditions as well as to external astrophysical mechanisms able to inhibit or induce star formation activity.

2.4 Genetics

Galaxies definitively evolve. We can reconstruct the past of a given galaxy by matching the observational properties of its stellar populations and ISM with (parametric) spectro–photo–chemical models (inductive approach). These are well–established models specialized in following the spectral, photometrical and chemical evolution of stellar populations formed with different gas infall rates and star formation laws (e.g. [16] and the references therein). The inductive approach allowed to determine that spiral galaxies as our Galaxy can not be explained with closed–box models (a single burst of star formation); continuous infall of low–metallicity gas is required to reproduce the local and global colors, metal abundances, star formation rates, and gas fractions. On the other hand, the properties of massive ellipticals (specially their high α-elements/Fe ratios) are well explained by a single early fast burst of star formation and subsequent passive evolution.

A different approach to the genetical study of galaxies emerged after cosmology provided a reliable theoretical background. Within such a background it is possible to "handle" galaxies as physical objects that evolve according to the initial and boundary conditions given by cosmology. The deductive construction of galaxies can be confronted with observations corresponding to different stages of the proto-galaxy and galaxy evolution. The breakthrough for the deductive approach was the success of the inflationary theory and the consistency of the so–called Cold Dark Matter (CDM) scenario with particle physics and observational cosmology. The main goal of these notes is to describe the ingredients, predictions, and tests of this scenario.

Galaxy evolution in action

The dramatic development of observational astronomy in the last 15 years or so opened a new window for the study of galaxy genesis: the follow up of galaxy/protogalaxy populations and their environment at different redshifts. The Deep and Ultra Deep Fields of the Hubble Spatial Telescope and other facilities allowed to discover new populations of galaxies at high redshifts, as well as to measure the evolution of global (per unit of comoving volume) quantities associated with galaxies: the cosmic star formation rate density (SFRD), the cosmic density of neutral gas, the cosmic density of metals, etc. Overall, these global quantities change significantly with z, in particular the SFRD as traced by the UV–luminosity at rest of galaxies [79]: since $z \sim 1.5-2$ to the present it decreased by a factor close to ten (the Universe is literally lightening off), and for higher redshifts the SFRD remains roughly constant or slightly decreases ([51, 61] and the references therein). There exists indications that the SFRD at redshifts 2–4 could be approximately two times higher if considering Far Infrared/submmilimetric sources (SCUBA galaxies), where intense bursts of star formation take place in a dust–obscured phase.

Concerning populations of individual galaxies, the Deep Fields evidence a significant increase in the fraction of blue galaxies at $z \sim 1$ for the blue sequence that at these epochs look more distorted and with higher SFRs than their local counterparts. Instead, the changes observed in the red sequence are small; it seems that most red elliptical galaxies were in place long ago. At higher redshifts ($z \gtrsim 2$), galaxy objects with high SFRs become more and more common. The most abundant populations are:

Lyman Break Galaxies (LBG) , selected via the Lyman break at 912Å in the rest–frame. These are star–bursting galaxies (SFRs of $10 - 1000 M_\odot$/yr) with stellar masses of $10^9 - 10^{11} M_\odot$ and moderately clustered.

Sub-millimeter (SCUBA) Galaxies, detected with sub–millimeter bolometer arrays. These are strongly star–bursting galaxies (SFRs of $\sim 1000 M_\odot$/yr) obscured by dust; they are strongly clustered and seem to be merging galaxies, probably precursors of ellipticals.

Lyman α emitters (LAEs), selected in narrow–band studies centered in the Lyman α line at rest at $z > 3$; strong emission Lyman α lines evidence phases of rapid star formation or strong gas cooling. LAEs could be young (disk?) galaxies in the early phases of rapid star formation or even before, when the gas in the halo was cooling and infalling to form the gaseous disk.

Quasars (QSOs), easily discovered by their powerful energetics; they are associated to intense activity in the nuclei of galaxies that apparently will end as spheroids; QSOs are strongly clustered and are observed up to $z \approx 6.5$.

There are many other populations of galaxies and protogalaxies at high redshifts (Luminous Red Galaxies, Damped Lyα disks, Radiogalaxies, etc.). A major challenge now is to put together all the pieces of the high–redshift puzzle to come up with a coherent picture of galaxy formation and evolution.

3 Cosmic structure formation

In the previous section we have learn that galaxy formation and evolution are definitively related to cosmological conditions. Cosmology provides the theoretical framework for the initial and boundary conditions of the cosmic structure formation models. At the same time, the confrontation of model predictions with astronomical observations became the most powerful testbed for cosmology. As a result of this fruitful convergence between cosmology and astronomy, there emerged the current paradigmatic scenario of cosmic structure formation and evolution of the Universe called Λ Cold Dark Matter (ΛCDM). The ΛCDM scenario integrates nicely: (1) cosmological theories (Big Bang and Inflation), (2) physical models (standard and extensions of the particle physics models), (3) astrophysical models (gravitational cosmic structure growth, hierarchical clustering, gastrophysics), and (4) phenomenology (CMBR anisotropies, non-baryonic DM, repulsive dark energy, flat geometry, galaxy properties).

Nowadays, cosmology passed from being the Cinderella of astronomy to be one of the highest precision sciences. Let us consider only the Inflation/Big Bang cosmological models with the F-R-W metric and adiabatic perturbations. The number of parameters that characterize these models is high, around 15 to be more precise. No single cosmological probe constrain all of these parameters. By using multiple data sets and probes it is possible to constrain with precision several of these parameters, many of which correlate among them (degeneracy). The main cosmological probes used for precision cosmology are the CMBR anisotropies, the type–Ia SNe and long Gamma–Ray Bursts, the Lyα power spectrum, the large–scale power spectrum from galaxy surveys, the cluster of galaxies dynamics and abundances, the peculiar velocity surveys, the weak and strong lensing, the baryonic acoustic oscillation in the large–scale galaxy distribution. There is a model that is systematically consistent with most of these probes and one of the goals in the last years has been to improve the error bars of the parameters for this 'concordance' model. The geometry in the concordance model is flat with an energy composition dominated in $\sim 2/3$ by the cosmological constant Λ (generically called Dark Energy), responsible for the current accelerated expansion of the Universe. The other $\sim 1/3$ is matter, but $\sim 85\%$ of this 1/3 is in form of non–baryonic DM. Table 2 presents the central values of different parameters of the ΛCDM cosmology from combined model fittings to the recent 3–year *WMAP* CMBR and several other cosmological probes [109] (see the WMAP website).

Table 2. Constraints to the parameters of the ΛCDM model

Parameter	Constraint
Total density	$\Omega = 1$
Dark Energy density	$\Omega_\Lambda = 0.74$
Dark Matter density	$\Omega_{DM} = 0.216$
Baryon Matter dens.	$\Omega_B = 0.044$
Hubble constant	$h = 0.71$
Age	13.8 Gyr
Power spectrum norm.	$\sigma_8 = 0.75$
Power spectrum index	$n_s(0.002) = 0.94$

In the following, I will describe some of the ingredients of the ΛCDM scenario, emphasizing that most of these ingredients are well established aspects that any alternative scenario to ΛCDM should be able to explain.

3.1 Origin of fluctuations

The Big Bang[4] is now a mature theory, based on well established observational pieces of evidence. However, the Big Bang theory has limitations. One of them is namely the origin of fluctuations that should give rise to the highly inhomogeneous structure observed today in the Universe, at scales of less than $\sim$ 200Mpc. The smaller the scales, the more clustered is the matter. For example, the densities inside the central regions of galaxies, within the galaxies, cluster of galaxies, and superclusters are about 10^{11}, 10^6, 10^3 and few times the average density of the Universe, respectively.

The General Relativity equations that describe the Universe dynamics in the Big Bang theory are for an homogeneous and isotropic fluid (Cosmological Principle); inhomogeneities are not taken into account in this theory "by definition". Instead, the concept of fluctuations is inherent to the Inflationary theory introduced in the early 80's by A. Guth and A. Linde namely to

[4] It is well known that the name of 'Big Bang' is not appropriate for this theory. The key physical conditions required for an explosion are temperature and pressure gradients. These conditions contradict the Cosmological Principle of homogeneity and isotropy on which is based the 'Big Bang' theory.

overcome the Big Bang limitations. According to this theory, at the energies of Grand Unification ($\gtrsim 10^{14}$GeV or $T \gtrsim 10^{27}$K!), the matter was in the state known in quantum field theory as vacuum. Vacuum is characterized by quantum fluctuations –temporary changes in the amount of energy in a point in space, arising from Heisenberg uncertainty principle. For a small time interval Δt, a virtual particle–antiparticle pair of energy ΔE is created (in the GU theory, the field particles are supposed to be the X- and Y-bossons), but then the pair disappears so that there is no violation of energy conservation. Time and energy are related by $\Delta E \Delta t \approx \frac{h}{2\pi}$. The vacuum quantum fluctuations are proposed to be the seeds of present–day structures in the Universe.

How is that quantum fluctuations become density inhomogeneities? During the inflationary period, the expansion is described approximately by the de Sitter cosmology, $a \propto e^{Ht}$, $H \equiv \dot{a}/a$ is the Hubble parameter and it is constant in this cosmology. Therefore, the proper length of any fluctuation grows as $\lambda_p \propto e^{Ht}$. On the other hand, the proper radius of the horizon for de Sitter metric is equal to c/H =const, so that initially causally connected (quantum) fluctuations become suddenly supra–horizon (classical) perturbations to the spacetime metric. After inflation, the Hubble radius grows proportional to ct, and at some time a given curvature perturbation cross again the horizon (becomes causally connected, $\lambda_p < L_H$). It becomes now a true density perturbation. The interesting aspect of the perturbation 'trip' outside the horizon is that its amplitude remains roughly constant, so that if the amplitude of the fluctuations at the time of exiting the horizon during inflation is constant (scale invariant), then their amplitude at the time of entering the horizon should be also scale invariant. In fact, the computation of classical perturbations generated by a quantum field during inflation demonstrates that the amplitude of the scalar fluctuations at the time of crossing the horizon is nearly constant, $\delta\phi_H \propto$const. This can be understood on dimensional grounds: due to the Heisenberg principle $\delta\phi/\delta t \propto$ const, where $\delta t \propto H^{-1}$. Therefore, $\delta\phi_H \propto H$, but H is roughly constant during inflation, so that $\delta\phi_H \propto$const.

3.2 Gravitational evolution of fluctuations

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 epochs) by gravitational instability in an expanding frame. The fluctuation or perturbation is characterized by its density contrast,

$$\delta \equiv \frac{\delta\rho}{\overline{\rho}} = \frac{\rho - \overline{\rho}}{\overline{\rho}}, \tag{1}$$

where $\overline{\rho}$ is the average density of the Universe and ρ is the perturbation density. At early epochs, $\delta << 1$ for perturbation of all scales, otherwise the homogeneity condition in the Big Bang theory is not anymore obeyed. When $\delta << 1$, the perturbation is in the *linear* regime and its physical size grows with the expansion proportional to $a(t)$. The perturbation analysis in the linear approximation shows whether a given perturbation is stable ($\delta \sim$ const or even $\rightarrow 0$) or unstable (δ grows). In the latter case, when $\delta \rightarrow 1$, the linear approximation is not anymore valid, and the perturbation "separates" from the expansion, collapses, and becomes a self–gravitating structure. The gravitational evolution in the *non–linear regime* is complex for realistic cases and is studied with numerical N–body simulations. Next, a pedagogical review of the linear evolution of perturbations is presented. More detailed explanations on this subject can be found in the books [72, 94, 90, 30, 77, 92].

Relevant times and scales.

The important times in the problem of linear gravitational evolution of perturbations are: (a) the epoch when inflation finished ($t_{inf} \approx 10^{-34}$s, at this time the primordial fluctuation field is established); (b) the epoch of matter–radiation equality t_{eq} (corresponding to æ $\approx 1/3.9 \times 10^4(\Omega_0 h^2)$, before t_{eq} the dynamics of the universe is dominated by radiation density, after t_{eq} dominates matter density); (c) the epoch of recombination t_{rec}, when radiation decouples from baryonic matter (corresponding to $a_{rec} = 1/1080$, or $t_{rec} \approx 3.8 \times 10^5$yr for the concordance cosmology).

Scales: first of all, we need to characterize the size of the perturbation. In the linear regime, its physical size expands with the Universe: $\lambda_p = a(t)\lambda_0$, where λ_0 is the comoving size, by convention fixed (extrapolated) to the present epoch, $a(t_0) = 1$. In a given (early) epoch, the size of the perturbation can be larger than the so–called *Hubble radius*, the typical radius over which physical processes operate coherently (there is causal connection): $L_H \equiv (a/\dot{a})^{-1} = H^{-1} = n^{-1}ct$. For the radiation or matter dominated cases, $a(t) \propto t^n$, with $n = 1/2$ and $n = 2/3$, respectively, that is $n < 1$. Therefore, L_H grows faster than λ_p and at a given "crossing" time t_{cross}, $\lambda_p < L_H$. Thus, the perturbation is supra–horizon sized at epochs $t < t_{cross}$ and sub–horizon sized at $t > t_{cross}$. Notice that if $n > 1$, then at some time the perturbation "exits" the Hubble radius. This is what happens in the inflationary epoch, when $a(t) \propto e^t$: causally–connected fluctuations of any size are are suddenly "taken out" outside the Hubble radius becoming causally disconnected.

For convenience, in some cases it is better to use masses instead of sizes. Since in the linear regime $\delta << 1$ ($\rho \approx \overline{\rho}$), then $M \approx \rho_M(a)\ell^3$, where ℓ is the size of a given region of the Universe with average matter density ρ_M. The mass of the perturbation, M_p, is invariant.

Supra–horizon sized perturbations.

In this case, causal, microphysical processes are not possible, so that it does not matter what perturbations are made of (baryons, radiation, dark matter, etc.); they are in general just perturbations to the metric. To study the gravitational growth of metric perturbations, a General Relativistic analysis is necessary. A major issue in carrying out this program is that the metric perturbation is not a gauge invariant quantity. See e.g., [72] for an outline of how E. Lifshitz resolved brilliantly this difficult problem in 1946. The result is quite simple and it shows that the amplitude of metric perturbations outside the horizon grows *kinematically* at different rates, depending on the dominant component in the expansion dynamics. For the critical cosmological model (at early epochs all models approach this case), the growing modes of metric perturbations according to what dominates the background are:

$$\delta_{m,+} \propto a(t) \propto t^{2/3}, ... matter \quad (2)$$
$$\delta_{m,+} \propto a(t)^2 \propto t, ... radiation$$
$$\delta_{m,+} \propto a(t)^{-2} \propto e^{-2Ht}, .. \Lambda \; (deSitter) \quad (3)$$

Sub–horizon sized perturbations.

Once perturbations are causally connected, microphysical processes are switched on (pressure, viscosity, radiative transport, etc.) and the gravitational evolution of the perturbation depends on what it is made of. Now, we deal with true *density* perturbations. For them applies the classical perturbation analysis for a fluid, originally introduced by J. Jeans in 1902, in the context of the problem of star formation in the ISM. But unlike in the ISM, in the cosmological context the fluid is expanding. What can prevent the perturbation amplitude from growing gravitationally? The answer is pressure support. If the fluid pressure gradient can re–adjust itself in a timescale t_{press} smaller than the gravitational collapse timescale, t_{grav}, then pressure prevents the gravitational growth of δ. Thus, the condition for gravitational instability is:

$$t_{grav} \approx \frac{1}{(G\rho)^{1/2}} < t_{press} \approx \frac{\lambda_p}{v}, \quad (4)$$

where ρ is the density of the component that is most gravitationally dominant in the Universe, and v is the sound speed (collisional fluid) or velocity dispersion (collisionless fluid) of the perturbed component. In other words, if the perturbation scale is larger than a critical scale $\lambda_J \sim v(G\rho)^{-1/2}$, then pressure loses, gravity wins.

The perturbation analysis applied to the hydrodynamical equations of a fluid at rest shows that δ grows *exponentially* with time for perturbations obeying the Jeans instability criterion $\lambda_p > \lambda_J$, where the exact value of λ_J is $v(\pi/G\rho)^{1/2}$. If $\lambda_p < \lambda_J$, then the perturbations are described by stable *gravito–acustic oscillations*. The situation is conceptually similar for perturbations in an expanding cosmological fluid, but the growth of δ in the unstable regime is *algebraical* instead of exponential. Thus, the cosmic structure formation process is relatively slow. Indeed, the typical epochs of galaxy and cluster of galaxies formation are at redshifts $z \sim 1-5$ (ages of $\sim 1.2-6$ Gyrs) and $z < 1$ (ages larger than 6 Gyrs), respectively.

Baryonic matter. The Jeans instability analysis for a relativistic (plasma) fluid of baryons *ideally* coupled to radiation and expanding at the rate $H = \dot{a}/a$ shows that there is an instability critical scale $\lambda_J = v(3\pi/8G\rho)^{1/2}$, where the sound speed for adiabatic perturbations is $v = p/\rho = c/\sqrt{3}$; the latter equality is due to pressure radiation. At the epoch when *radiation dominates*, $\rho = \rho_r \propto a^{-4}$ and then $\lambda_J \propto a^2 \propto ct$. It is not surprising that at this epoch λ_J approximates the Hubble scale $L_H \propto ct$ (it is in fact ~ 3 times larger). Thus, perturbations that might collapse gravitationally are in fact outside the horizon, and those that already entered the horizon, have scales smaller than λ_J: they are stable gravito–acoustic oscillations. When *matter dominates*, $\rho = \rho_M \propto a^{-3}$, and $a \propto t^{2/3}$. Therefore, $\lambda_J \propto a \propto t^{2/3} \lesssim L_H$, but still radiation is coupled to baryons, so that radiation pressure is dominant and λ_J remains large. However, when radiation decouples from baryons at t_{rec}, the pressure support drops dramatically by a factor of $P_r/P_b \propto n_r T/n_b T \approx 10^8$! Now, the Jeans analysis for a gas mix of H and He at temperature $T_{\rm rec} \approx 4000$ K shows that baryonic clouds with masses $\gtrsim 10^6 M_\odot$ can collapse gravitationally, i.e. all masses of cosmological interest. But this is literally too "ideal" to be true.

The problem is that as the Universe expands, radiation cools ($T_r = T_0 a^{-1}$) and the photon–baryon fluid becomes less and less perfect: the mean free path for scattering of photons by electrons (which at the same time are coupled electrostatically to the protons) increases. Therefore, photons can diffuse out of the bigger and bigger density perturbations as the photon mean free path increases. If perturbations are in the gravito–acoustic oscillatory regime, then the oscillations are damped out and the perturbations disappear. The "ironing out" of perturbations continues until the epoch of recombination. In a pioneering work, J. Silk [104] carried out a perturbation analysis of a relativistic cosmological fluid taking into account radiative transfer in the diffusion approximation. He showed that all photon–baryon perturbations of masses smaller than M_S are "ironed out" until t_{rec} by the (Silk) damping process. The first crisis in galaxy formation theory emerged: calculations showed that M_S is of the order of $10^{13} - 10^{14} M_\odot h^{-1}$! If somebody (god, inflation, ...) seeded primordial fluctuations in the Universe, by Silk damping all galaxy–sized perturbation are "ironed out". [5]

Non–baryonic matter. The gravito–acoustic oscillations and their damping by photon diffusion refer to baryons. What happens for a fluid of non–baryonic DM? After all, astronomers, since Zwicky in the 1930s, find routinely pieces

[5] In the 1970s Y. Zel'dovich and collaborators worked out a scenario of galaxy formation starting from very large perturbations, those that were not affected by Silk damping. In this elegant scenario, the large–scale perturbations, considered in a first approximation as ellipsoids, collapse most rapidly along their shortest axis, forming flattened structures ("pancakes"), which then fragment into galaxies by gravitational or thermal instabilities. In this 'top-down' scenario, to obtain galaxies in place at $z \sim 1$, the amplitude of the large perturbations at recombination should be $\geq 3 \times 10^{-3}$. Observations of the CMBR anisotropies showed that the amplitudes are 1–2 order of magnitudes smaller than those required.