Cambridge University Press & Assessment
978-1-107-13476-8 — Introduction to Galaxy Formation and Evolution
Andrea Cimatti , Filippo Fraternali , Carlo Nipoti
Excerpt
More Information

1 Introduction

1.1 Galaxies: a Very Brief History

Galaxies are gravitationally bound systems made of stars, interstellar matter (gas and dust), stellar remnants (white dwarfs, neutron stars and black holes) and a large amount of dark matter. They are varied systems with a wide range of morphologies and properties. For instance, the characteristic sizes of their luminous components are from ~ 0.1 kpc to tens of kiloparsecs, whereas the optical luminosities and stellar masses are in the range 10^3–10^{12} in solar units. Roughly spherical halos of dark matter dominate the mass budget of galaxies. As a reference, the size of the stellar disc of our Galaxy[1] is about 20 kpc, whereas the dark matter halo is thought to be extended out to $\approx$300 kpc. The total mass of the Galaxy, including dark matter, is $\sim 10^{12}\ \mathcal{M}_\odot$, whereas the stellar and gas masses amount to only $\approx 5 \times 10^{10}\ \mathcal{M}_\odot$ and $\approx 6 \times 10^9\ \mathcal{M}_\odot$, respectively.

The discovery of galaxies (without knowing their nature) dates back to when the first telescope observations showed the presence of objects, originally called nebulae, whose light appeared diffuse and fuzzy. The first pioneering observations of these nebulae were done with telescopes by C. Huygens in the mid-seventeenth century, and by E. Halley and N.-L. de Lacaille in the first half of the eighteenth century. Interestingly, in 1750, T. Wright published a book in which he interpreted the Milky Way as a flat layer of stars and suggested that nebulae could be similar systems at large distances. The philosopher I. Kant was likely inspired by these ideas to the extent that, in 1755, he explained that these objects (e.g. the Andromeda galaxy) appear nebulous because of their large distances which make it impossible to discern their individual stars. In this context, the Milky Way was interpreted as one of these many stellar systems (island universes).

In 1771, C. Messier started to catalogue the objects which appeared fuzzy based on his telescope observations. These objects were identified by the letter M (for Messier) followed by a number. Now we know that some of these objects are located within our Galaxy (star clusters and emission nebulae; e.g. M 42 is the Orion nebula), but some are nearby galaxies bright enough to be visible with small telescopes (e.g. M 31 is the Andromeda galaxy). However, Messier did not express any opinion about the nature and the distance of these systems. Since late 1700, W. Herschel, C. Herschel and J. Herschel increased the sample of diffuse objects thanks to their larger telescopes, and classified them depending on their

[1] The terms Galaxy (with the capital G) or Milky Way are used to indicate the galaxy where the Sun, the authors and the readers of this book are located.

observed features. In 1850, W. Parsons (Lord Rosse) noticed that some of these nebulae exhibited a clear spiral structure (e.g. M 51).

Since late 1800, the advent of astronomical photography allowed more detailed observations to be performed, and these studies triggered a lively discussion about the nature of the spiral nebulae and their distance. This led to the so-called Great Debate, or the Shapley–Curtis debate referring to the names of the two astronomers who, in 1920, proposed two widely different explanations about spiral nebulae. On the one hand, H. Shapley argued that these objects were interstellar gas clouds located within one large stellar system. On the other hand, according to H. Curtis, spiral nebulae were external systems, and our Galaxy was one of them. Clearly, this debate involved not only the very nature of these objects, but also the size and the extent of the Universe itself. The issue was resolved soon after with deeper observations. In 1925, using the 100-inch telescope at Mount Wilson Observatory, E. Hubble identified individual stars in M 31 and M 33 and discovered variable stars such as Cepheids and novae. In particular, Cepheids are pulsating giant stars that can be exploited as distance indicators. These stars are what astronomers call 'standard candles', i.e. objects whose intrinsic luminosity is known *a priori*, and that therefore can be used to estimate their distance. In 1912, it was found by H. Leavitt that the intrinsic luminosity of Cepheids is proportional to the observed period of their flux variation. Thus, once the period is measured, the intrinsic luminosity is derived and, therefore, the distance can be estimated. Based on these results, Hubble demonstrated that spiral nebulae were at very large distances, well beyond the size of our Galaxy, and that therefore they were indeed external galaxies.

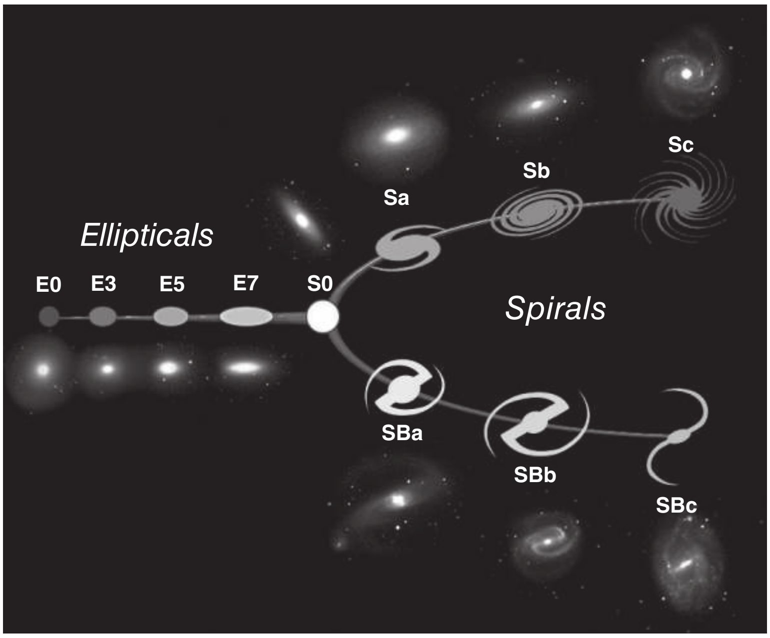
The term 'galaxy' originates from the Greek $\gamma\acute{\alpha}\lambda\alpha$, which means milk, and it refers to the fuzzy and 'milky' appearance of our own Milky Way when observed with the naked eye. Also external galaxies look 'milky' when observed with small telescopes. Discovering that galaxies were external systems also implied that the Universe was much larger than our Galaxy, and this was crucial to open a new window on cosmology in general. In modern astrophysics, the term 'nebula' is still used, but it refers only to objects within the interstellar medium of galaxies. Notable examples are the emission nebulae where the gas is photoionised by hot massive stars, dark nebulae which host cold and dense molecular gas mixed with interstellar dust, and planetary nebulae produced by the gas expelled by stars with low to intermediate mass during their late evolutionary phases. Since the discovery of Hubble, spiral nebulae have therefore been called spiral galaxies. In 1927–1929, based on galaxy samples for which radial velocities and distances were available, G. Lemaître and Hubble found that galaxies are systematically receding from us. In particular, their radial velocity is proportional to their distance: the farther away the galaxies, the higher the redshift of their spectral lines, and therefore the velocity at which they move away from us. This crucial discovery led to the Hubble–Lemaître law[2] which is the experimental proof that the Universe is expanding.

[2] In October 2018, the members of the International Astronomical Union (IAU) voted and recommended to rename the Hubble law as the Hubble–Lemaître law.

1.2 Galaxies as Astrophysical Laboratories

Present-day galaxies display a variety of properties and span a very broad range of luminosities, sizes and masses. At first sight, this already suggests that galaxy formation and evolution is not a simple process. However, the existence of tight scaling relations involving galaxy masses, sizes and characteristic velocities (e.g. the Tully–Fisher relation and the fundamental plane) indicates some regularities in the formation and assembly of these systems.

The first distinctive feature of a galaxy is its morphology. The shape of a galaxy as observed on the sky plane is a combination of the intrinsic three-dimensional (3D) structure and its orientation relative to the line of sight. Present-day galaxies show a broad range of shapes. Understanding the physical formation and evolution of the morphological types remains one of the most important, and still open, questions in extragalactic astrophysics. The first systematic study in the optical waveband dates back to 1926, when Hubble started a classification of galaxy morphologies following an approximate progression from simple to complex forms. In particular, Hubble proposed a tuning fork diagram on which the main galaxy types can be placed. Based on this classification, galaxies were divided into three main classes: ellipticals, lenticulars and spirals, plus a small fraction of irregulars. As shown in Fig. 1.1, the Hubble sequence starts from the left with the class of ellipticals (E). This class is further divided into subclasses as a function of their observed flattening.



The Hubble classification of galaxy morphology. © NASA and ESA, reproduced with permission.

Fig. 1.1

Perfectly round ellipticals are called E0, whereas the most flattened are the E7. If the observed shape of these galaxies is approximated by ellipses, their flattening is related to the ellipticity $\epsilon = (a - b)/a$, where a and b are the observed semi-major and semi-minor axes, respectively. The number written after the letter E is the integer closest to 10ϵ. Proceeding beyond the E7 class, galaxies start to display morphologies with a central dominant spheroidal structure (the so-called bulge) surrounded by a fainter disc without spiral arms. These systems are classified as lenticulars (S0) and represent a morphological transition from ellipticals to spirals. Proceeding further to the right, the tuning fork is bifurcated in two prongs populated by the two main classes of spiral (S) galaxies. In both prongs, spirals have the common characteristic of having a disc-like appearance with well defined spiral arms originating from the centre and extending throughout the outer regions. The top prong includes the so-called normal spirals characterised by a central bulge surrounded by a disc. These spirals are classified Sa, Sb and Sc as a function of decreasing prominence of the bulge (with respect to the disc) and increasing importance of the spiral arms. The bottom prong includes the barred spirals (SB) which show a central bar-like structure which connects the bulge with the regions where the spiral arms begin. Moving further to the right, i.e. beyond Sc types, all galaxies not falling into the previous classes are classified as irregulars (Irr).

Subsequent studies showed that ellipticals and lenticulars are red systems, made of old stars, with weak or absent star formation, high stellar masses, with a wide range of kinematic properties (from fast to absent rotation), and preferentially located in regions of the Universe where the density of galaxies is higher. On the other side of the tuning fork, spirals are bluer, have ongoing star formation, larger fractions of cold gas, stellar populations with a wide range of ages, kinematics dominated by rotation, and are found preferentially in regions with lower density of galaxies. Given the wide range of properties displayed by present-day galaxies, it is crucial to investigate the physical processes which led to their formation and evolution. The study of galaxies involves a wide range of galactic and sub-galactic scales ranging from hundreds of kiloparsecs down to sub-parsec level depending on the processes that are considered. In this respect, galaxies can be seen as 'laboratories' where a plethora of astrophysical processes can be investigated.

1.3 Galaxies in the Cosmological Context

Besides their role as astrophysical laboratories, galaxies can be placed in a broader context and exploited as point-like luminous 'particles' which trace the distribution of matter on scales much larger than the size of individual galaxies. This distribution, called large-scale structure, is the 3D spatial distribution of matter in the Universe on scales from tens of megaparsecs to gigaparsecs. Due to its characteristic shape, the large-scale structure is also called the cosmic web. The study of galaxies on these large scales has deep connections with cosmology, the branch of physics and astrophysics that studies the general properties, the matter–energy content and the evolution of the Universe as a whole. Modern cosmology rests on two major observational pillars. The first is the expansion of the Universe (Hubble–Lemaître law). The second is the nearly uniform radiation background observed in the microwaves, the cosmic microwave background (CMB), discovered in 1965 by A. Penzias and R. Wilson. The spectrum of the CMB is an almost perfect black body with a temperature $T \simeq 2.726$ K. The CMB radiation is interpreted as the thermal relic of the Big Bang that occurred about 13.8 Gyr ago when the Universe originated as a hot plasma with virtually infinite temperature and density. Although the detailed properties of the Big Bang itself are unknown, an expanding Universe can be described using the Einstein equations of general relativity together with the Friedmann–Lemaître–Robertson–Walker metric. The current view of the Universe relies on the Big Bang model and on the so-called ΛCDM cosmological framework. In this scenario, also known as standard cosmology, the Universe is homogeneous and isotropic on large scales, and it is made of ordinary matter (i.e. baryonic matter), neutrinos, photons and a mysterious component of cold dark matter (CDM). CDM is dominant with respect to ordinary matter as it amounts to about 84% of the whole matter present in the Universe. CDM is thought to be composed of non-relativistic massive particles that interact with each other and with ordinary matter only through the gravitational force. However, the nature and individual mass of these particles are currently unknown. For this reason, this is one of the main open questions of modern physics. In addition, a further component, called dark energy, is required to explain the current acceleration of the Universe expansion that S. Perlmutter, B. Schmidt and A. Riess discovered in 1998 exploiting distant supernovae as standard candles. In standard cosmology, the space-time geometry is flat (Euclidean), and dark energy is assumed to be a form of energy density (known as vacuum energy) which is constant in space and time. This form of dark energy is indicated by Λ and called the cosmological constant. However, other possibilities (e.g. a scalar field) are not excluded, and the nature of dark energy is currently unknown. This represents another big mystery of modern physics.

The ΛCDM model can be fully described by a small number of quantities called cosmological parameters which measure the relative fractions of the matter–energy components and constrain the geometry of the Universe. The ΛCDM model is now supported by a variety of cosmological probes such as the CMB, the Hubble expansion rate estimated from Type Ia supernovae, the properties of the large-scale structure and the mass of galaxy clusters. If the ΛCDM model is assumed, the observational results constrain the cosmological parameters with extremely high accuracy. In particular, in the present-day Universe, dark energy contributes ≈70% of the matter–energy budget of the Universe, whereas the contributions of dark matter and baryons amount to ≈25% and ≈5%, respectively, plus a negligible fraction of photons and neutrinos. The relative uncertainties on these fractions are very small (sub-per cent level). For this reason, modern cosmology is also called precision cosmology. However, it remains paradoxical that the nature of dark matter and dark energy, which together make 95% of the Universe, is still completely unknown despite the accuracy with which we know their relative importance.

Once the cosmological framework has been established, present-day galaxies can be seen as the endpoints that enclose crucial information on how baryonic and dark matter evolved as a function of cosmic time. In this regard, galaxies are also useful to test the ΛCDM cosmology. For instance, the current age of the oldest stars in galaxies should not be older than the age of the Universe itself, estimated to be about 13.8 Gyr based on observational cosmology. This key requirement is met by the age estimates of the Galactic globular clusters based on the Hertzsprung–Russell diagram.

1.4 Galaxies: from First Light to Present-Day Galaxies

Galaxies were originated from the primordial gas present in the early Universe. Fig. 1.2 shows a sketch of the main cosmic epochs that are treated in this textbook. Soon after the Big Bang, the baryonic matter was fully ionised and coupled with a 'bath' of black-body photons. In this photon–baryon fluid, the Universe was opaque because photons could not propagate freely due to the incessant Thomson scattering with free electrons. As the Universe expanded, its temperature and density gradually decreased and, about three minutes after the Big Bang, the nuclei of elements heavier than hydrogen (basically only helium and lithium) formed through a process called primordial nucleosynthesis. About 400 000 years after the Big Bang, the temperature and density dropped enough to allow lithium, helium and hydrogen to gradually recombine with electrons and form neutral atoms. This phase is called cosmological recombination. This is the epoch when the Universe became transparent because photons started to propagate freely thanks to the negligible role of Thomson scattering. The CMB radiation observed in the present-day Universe was originated in this phase and therefore represents the earliest possible image of the Universe. After recombination, the Universe was filled of dark matter and diffuse neutral gas composed of hydrogen, helium and lithium only. It is from the evolution of this primordial gas that the first luminous objects and galaxies began to form.

Understanding galaxy formation and evolution is a complex task because it involves several physical processes, their mutual interactions, and their evolution as a function of cosmic time. This is one of the most multi-disciplinary areas of astrophysics as it requires

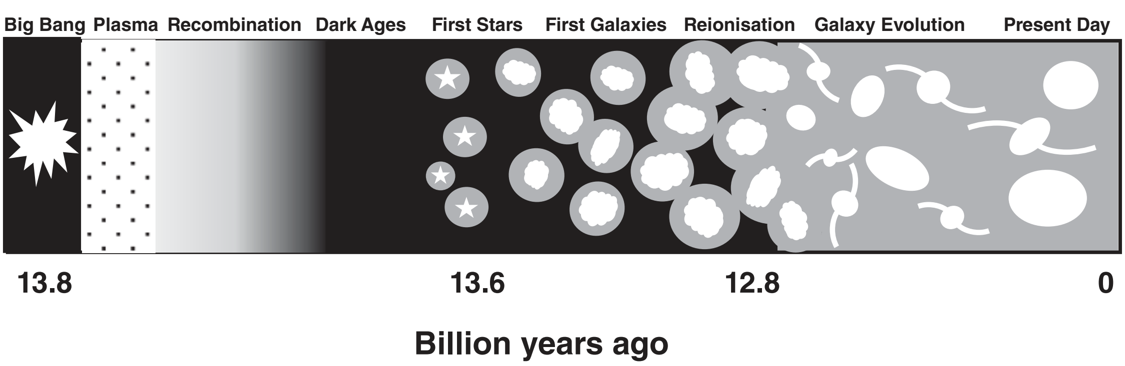


Fig. 1.2 A sketch of the main epochs which characterised the evolution of the Universe, starting from the Big Bang. After the formation of the first stars and galaxies, galaxies followed different evolutionary paths which led to the assembly of the galaxy types that we observe in the present-day Universe.

the cross-talk among a wide range of fields such as cosmology, particle physics (including dark matter) and the physics of baryonic matter. Galaxy formation and evolution is also a relatively young research field because galaxies were recognised as such only about a century ago, and their observation at cosmological distances became possibile only in the mid-1990s thanks to the advent of ground-based 8–10 m diameter telescopes in the optical and near-infrared spectral ranges, in synergy with the *Hubble Space Telescope* (*HST*).

The first step in the study of galaxy formation and evolution requires the definition of a cosmological framework (currently the ΛCDM model) within which galaxies form and evolve. The second step is to include the formation and evolution of dark matter halos which will host the first luminous objects and galaxies. In the ΛCDM model, dark matter halos are the results of the gravitational collapse of CDM in the locations where the matter density is high enough to locally prevail over the expansion of the Universe. As a matter of fact, the competition between the expansion of the Universe and gravity is one of the key processes in galaxy formation. On the one hand, if we take a large volume of the Universe at a given time, the mean matter density decreases with increasing cosmic time due to the expansion of the volume itself. On the other hand, the masses present in the same volume attract each other due to the reciprocal gravitational forces. In the early Universe, the typical masses of these halos were small, but they subsequently grew hierarchically with cosmic time through the merging with other halos and with the accretion of diffuse dark matter. Part of the gas is expected to follow the gravitational collapse of dark matter halos, and then to settle into their potential wells. The possibility to form a galaxy depends on whether this gas can have a rapid gravitational collapse. First of all, gravity must prevail over the internal pressure of the gaseous matter. However, this is not sufficient because the temperature rises as soon as the contraction proceeds. Gas heating is the enemy of galaxy formation because it increases the internal pressure and hampers gravitational collapse. This is why the second key requirement for galaxy formation is that gas cooling prevails over heating. Gas cooling can be produced by the emission of continuum radiation and spectral lines. The emitted photons abandon the gas cloud, carrying energy away, and therefore making the gas cooler and more prone to further gravitational collapse.

The cosmic epoch before the formation of the first collapsed objects (known as first stars or Population III stars) is named the dark ages because the Universe was made only of neutral gas, and luminous sources were completely absent (Fig. 1.2). We think that Population III stars began to form about 100 million years after the Big Bang from the collapse of pristine gas (H, He, Li) within dark matter halos with masses around $10^6\ \mathcal{M}_\odot$. At these early epochs, the main radiative coolants of the gas were primordial molecules such as LiH, HD and H_2 previously formed through gas-phase chemical reactions. This collapse led to the formation of protostellar objects and the subsequent ignition of the first thermonuclear reactions in the cores of Population III stars. When these systems started to shine, their strong ultraviolet radiation photoionised the surrounding gas. This was the beginning of the reionisation era. Population III stars ended their life very rapidly and vanished with the expulsion of most of their gas from their dark matter halos by violent supernova explosions. Thus, having lost most of the initial gas, these halos could not host further episodes of star formation. It is therefore thought that the formation of the first galaxies occurred later (a few hundred million years after the Big Bang) in larger dark matter halos with masses around $10^8\ \mathcal{M}_\odot$. These objects are called galaxies because they were massive enough to gravitationally retain a substantial fraction of the gas to prolong star formation without losing and/or heating it excessively due to supernova explosions.

After these early phases, galaxy formation proceeded following a wide range of evolutionary paths depending on the local conditions, the properties of the gas and the interactions with other systems (e.g. merging of their host halos). This is why the full understanding of galaxy formation and evolution is complex and requires a self-consistent treatment of the physical processes of baryonic matter (gas, stars and dust), their kinematics, their evolution within an expanding Universe, and the gravitational interactions with the dark matter component. The physics of baryonic matter is particularly complicated as it involves a variety of ingredients such as radiative processes, multi-phase gas physics and dynamics, gas cooling and heating, radiative transfer, star formation, stellar evolution, metal enrichment and feedback. Moreover, galaxy evolution involves the formation of supermassive black holes, the associated accretion of matter, and also the consequent feedback processes on the surrounding environment. A further complication is that all these processes and their evolution must be investigated on very wide ranges of spatial scales (from sub-parsec to megaparsec) and timescales, say from the lifetime of the most massive stars ($\sim 10^6$ yr) to the age of the Universe ($\sim 10^{10}$ yr). Fig. 1.3 illustrates the main ingredients that need to be included for the physical description of galaxy formation and evolution.

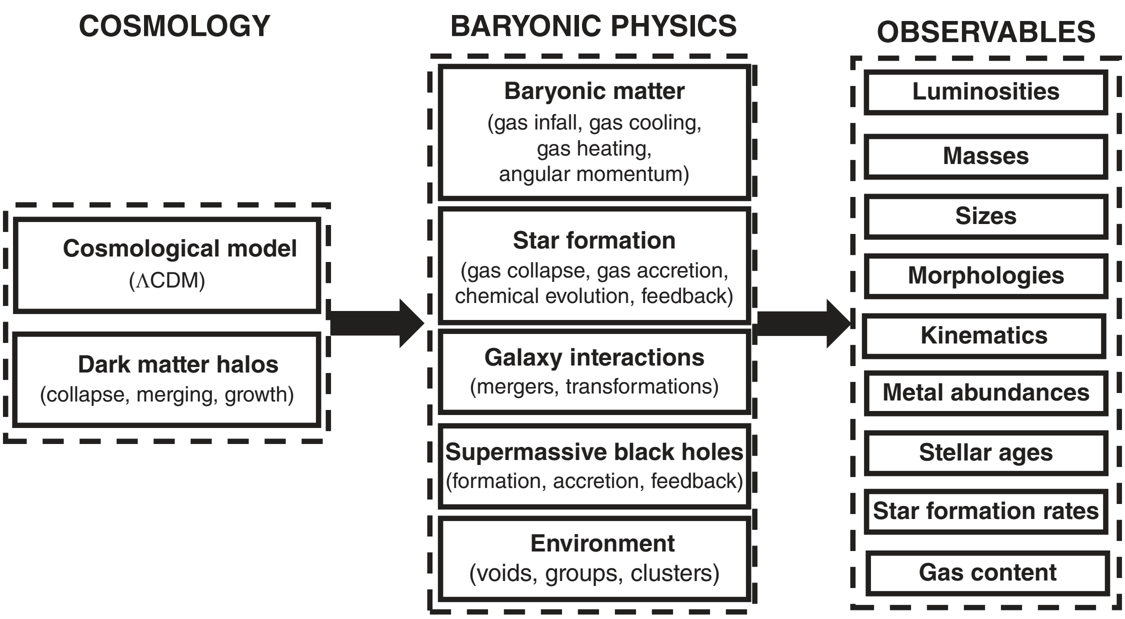


Fig. 1.3 The main ingredients of models of galaxy formation and evolution. *Left*. The cosmological model and the properties of dark matter halos define the 'skeleton' within which galaxies form and evolve. *Centre*. The main processes that drive the evolution of baryonic matter and galaxy formation. *Right*. The predicted properties of galaxies that are compared with the observations to verify the reliability of theoretical models.

1.5 Galaxies: Near and Far, Now and Then

Given the above complexity, how can we study galaxy formation and evolution? One approach is through theoretical models which describe coherently the physical processes involved in the formation of galaxies and their subsequent evolution from the smallest to the largest scales. In these models, the ΛCDM cosmology framework provides the initial conditions (e.g. the dark and baryonic matter fractions, the expansion rate of the Universe as a function of time, the properties of CDM halos and the hierarchical evolution of their masses). Once the cosmological framework is defined, galaxies can be modelled with two main methodologies. The first is based on cosmological hydrodynamic simulations, which follow as much as possible self-consistently the evolution of gas, star formation and feedback processes within dark matter halos. These simulations are very time consuming. This implies that sub-galactic scales can be simulated at the price of not covering large volumes of the Universe due to the limited computational resources. The second approach, called semi-analytic, consists in treating the physics of baryonic matter with a set of analytic prescriptions that, combined with the theoretically predicted evolution of dark matter halos, are tuned to reproduce the observed properties of present-day galaxies. The semi-analytic approach is cheaper from the computational point of view and therefore allows one to simulate large volumes of the Universe up to gigaparsec scales. However, the price to pay is that only the global properties of galaxies can be studied, and limited spatially resolved information is available. For these reasons, the two methods are complementary to each other. A further possibility is to perform analytic/numerical modelling of specific processes which take place within galaxies. An example is given by the chemical evolution models applied to the Milky Way.

The other approach to study galaxy formation and evolution is complementary to the theoretical modelling, and consists in the direct observation of galaxies in order to obtain data (images and spectra) from which the physical and structural properties can be extracted. A first possibility is the so-called archaeological approach where present-day galaxies are exploited as ‘fossils’[3] from which it is possible to reconstruct their past history based on what is observed today. For instance, the ages and metal abundances of the stellar populations present in a galaxy allow us to infer how star formation and the enrichment of heavy elements evolved as a function of cosmic time. With this approach, the most reliable results are obtained when the stars within a galaxy can be observed individually and therefore can be placed on the Hertzsprung–Russell diagram. Unfortunately, with the current telescopes, this can be done only within the Milky Way and for galaxies in the Local Group, a ≈1 Mpc size region where the Galaxy is located together with its neighbours. The study of our Galaxy is so important as a benchmark of galaxy evolution studies that the *Gaia* space mission has been designed to obtain distances and proper motions of more than a billion stars, with radial velocity measurements for a fraction of them. *Gaia* allowed us to derive a kinematic map of our Galaxy that is essential to investigate its formation and

[3] As present-day galaxies are considered ‘fossils’, the archaeological approach should be more appropriately called ‘palaeontological’.

evolution. Beyond the Local Group, galaxies become rapidly too faint and their angular sizes are too small to observe their stars individually. In these cases, one has to rely on the 'average' information that can be extracted from the so-called integrated light, i.e. the sum of the radiation emitted by the entire galaxy (or by a region of it).

Besides the archaeological studies in the present-day Universe, galaxy formation and evolution can also be investigated with the so-called look-back approach. This consists in the observation of galaxies at cosmological distances. Since light travels at a finite speed, the photons emitted from more distant galaxies reach us after a longer time interval. This means that distant galaxies appear today to us as they were in the past. Thus, it is possible to observe directly the evolution of galaxy properties if we observe galaxies at increasing distances. The fundamental assumption that makes the look-back time approach possible is that the Universe is homogeneous on large scales, so the global properties of the galaxy population on sufficiently large volumes are independent of the position in the Universe. This implies that galaxies in the local volume, in which our Galaxy is located, are representative of the general population of present-day galaxies. Similarly the galaxies observed in a distant volume are assumed to be representative of the past population of galaxies. For instance, if we want to investigate the evolution of spiral galaxies, we need to observe samples of this type of galaxies at increasing distances (i.e. larger look-back times) and to study how their properties (e.g. size, rotation velocity, mass, star formation) change with cosmic time. With this approach, it is truly possible to trace the detailed evolution of galaxies billions of years ago.

The archaeological and look-back approaches are complementary to each other, and their results are essential to build theoretical models and verify their predictions. However, in both cases multi-wavelength data are needed to provide a complete view of galaxy properties and their evolution. The reason is that galaxies are multi-component systems which emit radiation in different regions of the electromagnetic spectrum through diverse processes. For instance, due to the typical temperatures of the stellar photospheres, the starlight is concentrated from the ultraviolet to the near-infrared. Instead, the study of the interstellar molecular gas and dust requires observations from the far-infrared to the millimetre, the atomic hydrogen must be investigated in the radio, and the hot gas and the supermassive black hole activity in the ultraviolet and X-rays. The multi-wavelength approach is limited by the terrestrial atmosphere which is opaque and/or too bright in several spectral ranges. Ground-based telescopes can observe only in the optical, near-infrared and in a few transparent windows of the submillimetre, millimetre and radio. The other spectral ranges are accessible with space-based telescopes. The major advance in multi-wavelength studies of galaxy evolution at cosmological distances became possible thanks to the concurrence of ground- and space-based telescopes which, for the first time, allowed the identification of galaxies at cosmological distances. In the realm of space telescopes, the main contributions to galaxy evolution studies have come from the *Chandra X-ray Observatory*, *XMM-Newton* (X-rays), *Galaxy Evolution Explorer* (*GALEX*; ultraviolet), *HST* (optical/near-infrared), *Spitzer* (mid-infrared) and *Herschel* (far-infrared). In ground-based observations, the look-back approach became a reality with the advent of the 8–10 m diameter Keck telescopes and the Very Large Telescope (VLT) operating since the mid-1990s in the optical and near-infrared, followed by other facilities of