

Galaxy Evolution

The galaxies that we observe in the universe at the present time exhibit a remarkable variety of properties, such as morphology, colors, luminosity and dynamics. There are galaxies which are one hundred times brighter than our own Milky Way and others that have one thousandth its luminosity. Some galaxies have very regular shapes, other are amorphous. There are galaxies with spheroidal shape that have little amount of rotation, and whose stars have large 'thermal motions' (like the molecules of a hot fluid), and galaxies that resemble rotating disks, where self-gravity and centrifugal force balance each other. There are galaxies with high amounts of gas (mostly hydrogen and helium) and dust, and a vigorous activity of STAR FORMATION. Other galaxies consist mostly of old, evolved stars, indicating that the activity of star formation in these systems ended in a remote past.

Why do galaxies show such a variety of properties? How could they form and diversify in this way, and when did they do so? Is a given property an indication of a specific evolutionary phase in the life of a galaxy, or have galaxies with different properties formed and evolved in different ways? Which physical mechanism is each property the manifestation of?

Such questions are among the most important in the study of cosmology, and yet they still elude a consistent and complete explanation. There was a time during the cosmic evolution that followed the BIG BANG when there were no galaxies, and the universe was filled with matter and radiation uniformly distributed throughout the space. Only very tenuous perturbations of the density of matter and radiation altered this otherwise almost perfect homogeneity. Then the universe underwent a transition from this phase to the highly structured one that characterizes the current epoch. GALAXY FORMATION and evolution are the results of a complex sequence of events that occurred during this transition. Their importance rests in the fact that the chronology and dynamics of these events are intimately linked to the global cosmological properties of the universe (i.e. the age of the universe, if it will expand for ever or if it will eventually collapse again in a Big Crunch, or if the expansion is an accelerated one or not) and to those of the matter that fills the cosmic space (e.g. the composition of the dark matter and its abundance relative to the ordinary, visible matter). The goal of galaxy evolution is to reconstruct back in time the physical mechanisms that led to the present-day galaxies and to explain them in terms of the properties of the matter and of the cosmology.

The morphology, structure and luminosity of galaxies

The galaxies in the present-day universe have been extensively studied for more than seventy years and, as a result, we have now a fairly accurate knowledge of their physical properties, such as the distribution of luminosity, abundances, morphology, age, chemical composition,

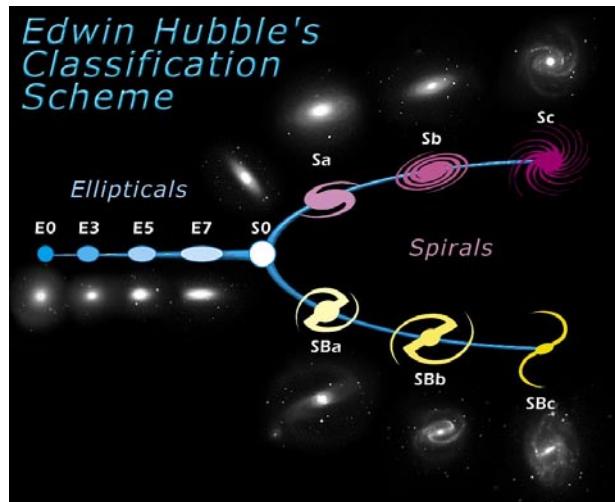


Figure 1. The Hubble morphological classification. Elliptical galaxies are ordered from E0 to E7, according to their ellipticity. The S0 type essentially represents elliptical galaxies that also have a disk. Spiral galaxies are divided into normal and barred types, and are ordered from the 'a' to the 'd' sub-types (including mixed type such as 'bc' and 'cd' not shown in the figure), according to the increasing importance of the disk component over the bulge. Irregular galaxies are not shown. (Photo credit: STScI OPO.)

and internal kinematics. This knowledge forms the background against which observations of the properties of galaxies at earlier cosmic epochs are compared to test and quantify galaxy evolution.

Galaxy morphology

The morphology of the luminous component of galaxies, namely stars, gas and dust, is probably the most obvious manifestation of the diversity of galaxy properties. During the first half of the 19th century, EDWIN P HUBBLE classified galaxy morphologies into the four principal types of ellipticals, lenticulars, spirals and irregulars. Figure 1 shows the classical representation of the Hubble classification, also referred to as the *Hubble sequence*.

Essentially, the HUBBLE CLASSIFICATION scheme is based on the relative proportions of the two major structural parts of the luminous component of galaxies, namely the bulge and the disk. The bulge is generally made of relatively old and evolved stars, which are characterized by red colors, and is shaped like an ellipsoid with various degrees of flattening and oblateness (ellipsoids with no flattening and oblateness are spheres). In general the bulge contains no appreciable amount of dust and gas, and therefore has no activity of star formation. The disk is composed by a mixture of old and young stars and by gas and dust, and it is frequently site of star formation activity. Its colors are generally bluer than those of the bulge. The most visually striking feature of the disk is the frequent presence of the 'SPIRAL ARMS', density waves that generate in the disk and which contain active regions of

star formation. Other articles in this encyclopedia discuss galaxy morphology in greater detail.

Bulge and disk do not only differ for the composition of their stellar populations and interstellar medium, but also for their dynamics. Galaxies are large concentrations of mass, and they would collapse under the effect of their own self-gravity if some opposing force were not there to keep them in dynamical equilibrium. Disks are kept in equilibrium by rotation, which provides the centrifugal force to oppose gravity. Bulges have modest or no rotation, and are sustained against gravity by the 'velocity dispersion' of their stars, which is similar to the motions of molecules in a hot gas. These motions provide the pressure that contrasts gravity and keeps the structure in dynamical equilibrium.

ELLIPTICAL GALAXIES consist only of the bulge component and have no disk. LENTICULAR GALAXIES are essentially ellipticals with a very thin, often almost invisible disk, which gives them the shape of a lentil. SPIRAL GALAXIES have both a bulge and a disk with spiral arms. If a bar is present, this contains the bulge at its center and has the spiral arms departing from its extremities. IRREGULAR GALAXIES tend to resemble spiral ones, except that the bulge, the disk and the spiral arms are not regularly defined. There are also amorphous galaxies, which have morphology that cannot be easily classified into any of the main types.

Initially, it was believed that the Hubble sequence was the manifestation of an evolutionary process such that galaxies originated as ellipticals, which were therefore called the 'early types', and evolved to become spirals and irregulars, which were called the 'late types'. Although it has been shown that such ideas were incorrect, nonetheless the nomenclature has remained, and today it is common, for example, to refer to ellipticals and lenticulars as 'early-type galaxies'.

The luminosity function

One very powerful tool that describes in a quantitative way two of the most important properties of galaxies is a statistic called the *luminosity function*, which provides information on the relative frequency in space of galaxies with given luminosity. The reason to couple together luminosity and abundance is an important one. As will become clear later, the luminosity of galaxies is a measure of either their stellar content or of their current activity of star formation or of both (depending on the evolutionary phase of the galaxy). The stellar content, in turn, is generally correlated to the total mass of the galaxy, to its star formation history as well as to the history of merging with other galaxies. Finally, the current activity of star formation provides a measure of the rate at which the stellar content of galaxy is increasing.

The abundance of galaxies in space (or volume density) reflects another set of important physical properties. Galaxies form within virialized structures, namely systems that have collapsed under the effect of their self-gravity and have decoupled their dynamical evolution from that of the cosmic mass distribution

(for example, they have stopped following the cosmic expansion). The theory of gravitational instability predicts that the number of virialized systems that have formed at any given time depends on their mass, with more massive systems being less abundant than less massive ones. It also predicts that, for a given mass, the number of systems depends on the density parameters Ω and cosmological constant Ω_Λ and on the spectrum of the primordial density perturbations that emerged from the Big Bang¹. However, only galaxies that have formed enough stars or that have enough star formation (and whose light is not obscured by dust) can be observed. Thus, by coupling together the information on the luminosity and abundances of galaxies, the luminosity function provides synthetic information on the formation and evolution of both the structural and visible components of galaxies as well as on the cosmological parameters. A more detailed discussion on the luminosity function can be found in the article LUMINOSITY FUNCTION OF GALAXIES.

The luminosity function of local galaxies has been directly measured with good accuracy. It is found that fainter galaxies are much more common than brighter ones. For example, although galaxies with luminosity similar to that of the Milky Way or the Andromeda galaxy² are relatively common, fainter galaxies such as the Magellanic Clouds are much more frequent. Bright galaxies such as the big ellipticals often found at the centers of clusters of galaxies are much rarer. Irregular galaxies are typically fainter and more abundant than galaxies with regular morphology.

Star formation and merging: the engines of galaxy evolution

Clearly, a morphological classification is useful only if it reflects some important physical property that can be directly linked to the origin and evolution of the galaxies. What are the physical properties of the Hubble sequence? We have seen that the age of the stellar populations and the dynamical state play an important role, but how do they relate to the history of star formation and how can we use them to reconstruct its evolution? How can we exploit the information contained in the luminosity function?

Star formation

Let us illustrate the role of star formation with an example. Let us consider the case of a spiral galaxy which contains 10^{11} stars similar to the Sun. This would be approximately the size of a galaxy like the Milky Way. Let us imagine that

¹ The cosmological parameters Ω and Ω_Λ measure the total mass content of the universe and the strength of the cosmological constant, a form of repulsive force that may cause the universe to keep expanding at an accelerated pace. Their combined value decides if the universe will expand for ever or it will collapse into a Big Crunch. Together with the Hubble constant, they determine the age of the universe

² Such galaxies are commonly referred to as ' L^* galaxies'. The symbol L^* is used to indicate the value of the blue luminosity of galaxies similar to the Milky Way. Its value is $L^* = 10^{12}L_\odot$, where L_\odot denotes the luminosity of the Sun or 4×10^{33} erg s⁻¹

this galaxy is experiencing an episode of star formation, and that it is generating new stars at a rate of, say, 10 solar masses per year or $M_{\odot} \text{ yr}^{-1}$ (this is the total stellar mass, not the number, of new stars). The STELLAR MASSES of freshly produced stars are not all equal, but vary approximately between 100 solar masses to about one tenth of solar mass (i.e. stars whose mass is between one hundred times and one tenth the mass of our Sun. The symbol that indicates the mass of the Sun is M_{\odot}). The relative proportions of stars of a given mass produced during an episode of star formation are, to first approximation, the same everywhere and are described by the so-called initial mass function (IMF). These proportions are such that for every star of mass M being produced, there are about 2.5 times as many stars with mass $M/2$ being also produced, implying that low-mass stars are formed much more frequently than high-mass stars. Another way to state this is that when a given total amount of stellar mass is being produced in a galaxy, this is distributed prevalently in low-mass stars.

While low-mass stars carry the bulk of the new stellar mass produced during an episode of star formation, they contribute very little to the luminosity of the event. This is because high-mass stars are much more luminous than low-mass ones. A star of $100 M_{\odot}$ is about 1000 000 times brighter than the Sun. Even if there is only one star of $100 M_{\odot}$ every 500 stars of $1 M_{\odot}$, it alone outshines all of them, whose contribution to the total luminosity is negligible. After the first million years of continuously forming stars at a rate of $10 M_{\odot}$ per year, the galaxy of our example has put together a total stellar mass equal to $10 \times 1000 000 = 10^7 M_{\odot}$, about one ten-thousandth of its initial stellar mass. Of these, about 20 000 are massive stars. Because each of these massive stars is 1000 000 times brighter than stars like the Sun, their total luminosity will approximately equal that of 2×10^{10} of these. Thus, during the star formation activity the galaxy increases its luminosity by $\sim 20\%$. Its colors also become bluer, because the radiation emitted by massive stars is very rich of light at short wavelengths, which include the blue and ultraviolet ones. Had the galaxy initially been fainter, containing say 10^{10} Sun-like stars, the same episode of star formation would have trebled its initial luminosity.

Although fainter, however, low-mass stars live much longer than high-mass ones. A star of $100 M_{\odot}$ lives about one million years, after which it explodes as supernova, while a star like the Sun keeps burning its hydrogen nuclear fuel for about 10 billion years, one hundred thousand times longer. Thus, about one million years or so after the end of the star formation, when all the massive stars have died out, the luminosity of the galaxy eventually becomes again dominated by low-mass stars, both the recently created new ones and the ones that existed before the episode of star formation. If star formation continued at the same rate for one million years, the galaxy of our example, which started with 10^{11} Sun-like stars, would create another 10^7 of them, or only 0.1% of the original amount. Once the massive stars have died out, the luminosity of the galaxy in its post-starburst phase is again

powered by these stars and it returns to the original value. Of course, star formation of longer duration would result in a larger increase of final luminosity and stellar mass. The luminosity of the galaxy during the phase of star formation, however, depends only on the rate at which stars are being produced because this is what determines the amount of massive stars that can be observed at a given time. Also, after star formation ends, the colors of the galaxy also return to be dominated by the light of the low-mass stars, which is redder than that of the massive stars whose colors dominated the spectral energy distribution of the galaxy during the star-forming phase.

Luminosity, stellar mass and color are not the only changes induced in a galaxy by star formation. Another fundamental product is the creation of the 'metals', namely of elements heavier than helium. The chemical composition of the universe as it emerged from the Big Bang practically consisted only of hydrogen and helium, in the proportion of about 80% and 20% by mass, respectively. Essentially all the heavier elements that we observe, both on Earth and elsewhere in the universe, have been synthesized in the interiors of stars by reactions of nuclear fusions at epochs subsequent to the Big Bang. The metals synthesized in the interiors of stars are released to interstellar space during SUPERNOVA explosions. Once the chemically enriched material has been dispersed, it becomes available to fuel further star formation and, with successive generations of star formation episodes, the general metal content of the galaxy (and of the universe as well) increases with time.

Merging

Another important mechanism that drives galaxy evolution is merging, which can transform the morphology of galaxies and promote a substantial amount of star formation over a short period of time. It also induces changes in the volume density and stellar content, and therefore affects the evolution of the luminosity function. Merging of galaxies is observed at all cosmological epochs where galaxies have been detected, and it is also predicted in all theories of galaxy evolution. See GALAXIES: INTERACTIONS AND MERGERS.

When two or more galaxies merge, the final structure that forms may not bear any resemblance to the original components. The outcome of the process depends on the relative masses of the mergers and on their types. If two galaxies with approximately the same mass merge, the perturbation to the gravitational equilibrium of the two systems is very large and takes place on a relatively short time scale (of the order of the free-fall time scale, which is $\sim 10^7$ – 10^8 years for a galaxy similar to the Milky Way). The gravitational potential of the forming system varies very rapidly throughout the whole structure during the event, a process known as *violent relaxation*, and when this happens the final equilibrium configuration of the system is a spheroid, similar to a bulge or an elliptical galaxy, supported against its self-gravity by the velocity dispersion. If the two merging systems were two spiral

galaxies, their disks would be destroyed during the event, and the gas contained in them transformed into stars on a time scale of the order of a few free-fall times, giving origin to a very powerful burst of star formation. In this case, depending on the gas content, the stellar mass of the final galaxy can increase to significantly larger values than the sum of the two mergers, and the luminosity during the event can reach prodigious values.

If the merging systems have largely different masses, and consist of one massive galaxy and one or more sub-massive ones, the event is not as traumatic as above. Although the small galaxy is destroyed and loses its identity, it only causes a relatively minor perturbation to the structure of the massive galaxy, which in general retains its properties and morphology. If there is available gas, either in the large galaxy or in the accreting one, the perturbation can induce episodes of star formation that may significantly increase the luminosity of the galaxy, even if they do not in general appreciably change its stellar mass.

Reconstructing the evolution from present to past
 The morphological and dynamical differences of present-day galaxies have been interpreted as evidence of different evolutionary history (Kennicutt 1998). The STELLAR POPULATIONS of elliptical and lenticular galaxies as well as the bulges of spirals are very old, and hence the activity of star formation must have ended rather early during cosmic evolution. Direct estimates of the age of the stars place the epoch of their formation to when the universe was about $\sim 20\%$ of its current age. For example, if the universe is 14 billion years old, most ellipticals, lenticulars and bulges already assembled the bulk of the stars by the time the universe was about 3 billion years old. This means, for example, that to form a total stellar mass of $\sim 10^{11} M_{\odot}$ in a few billion years, star formation rates must have been very high, of the order of $10^{11}/10^9 \sim 10^2 M_{\odot} \text{ yr}^{-1}$. Thus, the 'spheroids' must have formed their stars during intense but relatively short bursts of star formation.

In the disk of spiral galaxies star formation is observed to proceed at a relatively low pace but in a continuous fashion, and the age and composition of the stars suggest that this was the case in the past as well, although the rate of star formation was probably higher in the past. Luminosity and colors (the so-called spectral energy distribution) of DISK GALAXIES, therefore, largely reflect those of the forming stellar population. At the onset of the first episode of star formation, the spectral energy distribution closely resembled that of young massive stars, since only these objects populated the galaxy. At more advanced stages, when a significant older stellar population had built up, the spectral energy distribution became redder, reflecting the presence of the low-mass stars. For such a mechanism to take place on a time scale comparable to the age of the universe it requires a reservoir of gas to be continuously available to feed the activity of star formation. This gas is thought to accrete on the disk from the outer regions of the galaxy.

The morphological and dynamical properties of galaxies are decided when the body of the structure forms by gravitational collapse of a proto-cloud or of discrete sub-structures as in the case of merging. If the collapsing structure has enough angular momentum (in other words, enough rotation), it eventually settles on a final equilibrium configuration which has the shape of a disk, where the centrifugal force balances gravity. On the other hand, if the structure does not have enough angular momentum, there is no centrifugal force to oppose gravity, and the collapse continues until the velocity dispersion of the sub-components of the structure can halt it. The equilibrium configuration in this case is that of a spheroid, whose morphology is significantly more compact than that of a disk.

This difference in the collapse dynamics have consequences in the star formation mechanisms as well. The violent collapse that gives origin to a spheroid deeply changes the gravitational potential of the forming structure in a short time scale (of the order of the free-fall time). This large perturbation can drive a violent burst of star formation throughout the structure that converts all the available gas into stars on a time scale of several free-fall times (i.e. 10^8 – 10^9 yr). When the burst ends because of gas exhaustion, no more significant star formation is possible if the galaxy does not accrete additional gas by interaction and merging with other galaxies. The stars, which are essentially coeval, then begin a phase of passive evolution.

Observing the evolution from past to present

Directly comparing the properties of earlier galaxies with the present-day ones is another powerful avenue of investigation which has been made possible by the technological advances in telescopes and instrumentation achieved over the last 30 years. Thanks to the travel time imposed by the finite speed of light, approximately 1.6 billion miles per year, galaxies at earlier cosmic epochs can be studied simply by looking at distant galaxies. This is because a distant galaxy appears to an observer as it was in the past, at a look-back time equal to the travel time of the light from the galaxy to the observer. For very distant galaxies the look-back time is a large fraction of the entire age of the universe. Table 1 one shows the relationship between the luminosity distance³, redshift, cosmic age, look-back time, and fractional age of the universe. These quantities have been calculated assuming a universe with a Hubble constant equal to $H_0 = 65 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and with density parameter and cosmological constant equal to $\Omega = 0.3$ and $\Omega_{\Lambda} = 0.7$, respectively. It is common practice to use the redshift, rather than the distance, to provide a measure of the look-back time (for more information about the redshift see the article on GALAXIES AT HIGH REDSHIFT in this volume). Galaxies with redshift smaller than $z \sim 0.1$,

³ This is the distance that enters in the inverse-square law, which states that the observed flux ϕ of the source is inversely proportional to the square of the distance d from the source to the observer, or $\phi = L/4\pi d^2$, where L is the luminosity of the source.

or distance shorter than ~ 500 Mpc, are considered to be local systems observed in the present-day universe. Note that the epoch when the universe was half its present age (which is 14.5 billion years with the adopted cosmological parameters) corresponds to redshift $z \sim 0.7$, or to a distance of ~ 4600 Mpc, and that the bulk of the cosmic evolution, say $\sim 80\%$ of the cosmic age, took place from redshift $z \sim 2$ to the present time (redshift $z = 0$). Note also the little increment of look-back time at very high redshifts. Finally, we recall that the cosmic microwave background radiation originated at $z \sim 1300$.

Table 1. Relationship between luminosity distance D in Mpc, redshift z , cosmic age (in billion years), look-back time (LBT) (in billion years) and fractional age as % of the present age.
 $H_0 = 65 \text{ km s}^{-1} \text{ Mpc}^{-1}$; $\Omega = 0.3$; $\Omega_\Lambda = 0.7$.

D	z	Age	LBT	%
0	0.0	14.51	0.00	100.00
496	0.1	13.11	1.40	90.35
1672	0.3	10.83	3.68	74.64
3051	0.5	9.08	5.43	62.58
4587	0.7	7.72	6.79	53.20
7116	1.0	6.20	8.32	42.73
11 749	1.5	4.53	9.99	31.22
18 735	2.0	3.48	11.04	23.98
21 965	2.5	2.77	11.74	19.09
27 378	3.0	2.26	12.24	15.58
32 935	3.5	1.91	12.60	13.16
38 610	4.0	1.63	12.88	11.23
50 240	5.0	1.24	13.27	8.55
74 339	7.0	0.81	13.71	5.58
111 830	10.0	0.50	14.01	3.45
657 150	50.0	0.05	14.46	0.34
1370 345	100.0	0.02	14.50	0.14
19 222 400	1300.0	0.0004	14.51	0.003

Faint blue galaxies

Currently, the observations have studied relatively well galaxies at redshifts up to $z \sim 1$. From $z \sim 1$ up to $z \sim 2.5$ there is a gap of systematic observations due to the difficulty, with current technology, of measuring redshifts of faint sources in that range. At redshift higher than $z \sim 2.5$, and up to $z \sim 5$, star-forming galaxies have been observed and studied in large numbers (but not other types of galaxies, if they existed at those epochs) thanks to a special observing strategy called the ‘Lyman-break technique’ (see the article GALAXIES AT HIGH REDSHIFT in this volume), fine-tuned to be sensitive in that redshift range. More recently, there have been a few reports of identification of star forming galaxies with redshifts as high as $z \sim 6.2$, and also of candidates at redshifts $z \sim 10$. Great efforts are currently being put into the design and construction of very large telescopes, such as ground-based ones with mirrors of 30 meters or more, or the 8-meter space-based Next Generation Space Telescope (NGST), capable of detecting galaxies at $z \sim 20$, if they exist there.

The first surveys of very faint galaxies were carried out in the late 1970s and early 1980s, following the

advent of the first CCD detectors, and consisted mostly of deep imaging surveys (Tyson 1977). Although they did not provide direct information on the distance to the galaxies (and thus on the epoch at which they were observed), these observations found evidence that galaxies’ properties have changed over time. If the distribution of *intrinsic* luminosities and the abundances of galaxies did not evolve, then from the knowledge of the luminosity function at the present time one can accurately predict the numbers of distant (i.e. earlier) galaxies that have a specified distribution of *apparent luminosities*. This is due to the fact that the apparent luminosity of a galaxy is related to its intrinsic luminosity by the inverse-square law. Thus, under the assumption that the distribution and abundances of absolute luminosities (i.e. the luminosity function) is the same at any epoch, one can compute the distribution of apparent luminosities of galaxies expected in a deep survey simply by summing together the contribution of the luminosity function from each epoch. Very interestingly, the observations revealed that the number of faint galaxies is much higher than one would expect from such a calculation, with the excess increasing at fainter galaxies and being more pronounced at bluer wavelengths than at redder ones. Figure 2 shows this effect by plotting the number counts of galaxies per square degree observed in the sky at various wavelengths versus their apparent luminosity, expressed in magnitude⁴ (bottom) or physical units (top). Plotted in the graph are the observed counts in the K , I , B and U bands, respectively, corresponding to wavelengths of 22 000 (near-infrared), 8000 (red end of visible radiation), 4000 (blue radiation) and 3600 Å (violet end of visible radiation). The different symbols represent data from various observers, and provide an indication on the uncertainties of the observations. The curves are the theoretical predictions based on the local luminosity function. As can be seen, the excess in the counts over the predictions is very strong in the U and B bands, where it becomes noticeable already at around magnitude 22 and it sharply increases at fainter levels. The excess becomes progressively smaller towards redder wavelengths, and in the K band it is noticeable only at the faintest levels.

While deep multi-band imaging showed that the number density and/or the luminosity of blue galaxies is rapidly evolving going towards earlier epochs, without an indication of the distance to these faint galaxies, however, it is not possible to understand from these data alone when such evolution took place and whether the evolution of the blue galaxies was synchronous or not with other types.

The advent of more sensitive CCD detectors and advanced instrumentation during the late 1980s and early 1990s made it eventually possible to carry out large faint spectroscopic surveys and measure the redshifts of

⁴ The magnitude m of a galaxy with apparent luminosity f is defined as $m = -2.5 \log(f) + c$, where c is an arbitrary number that sets the zero point. Because of the negative sign, the scale of magnitude is such that brighter sources have smaller magnitudes and vice versa.

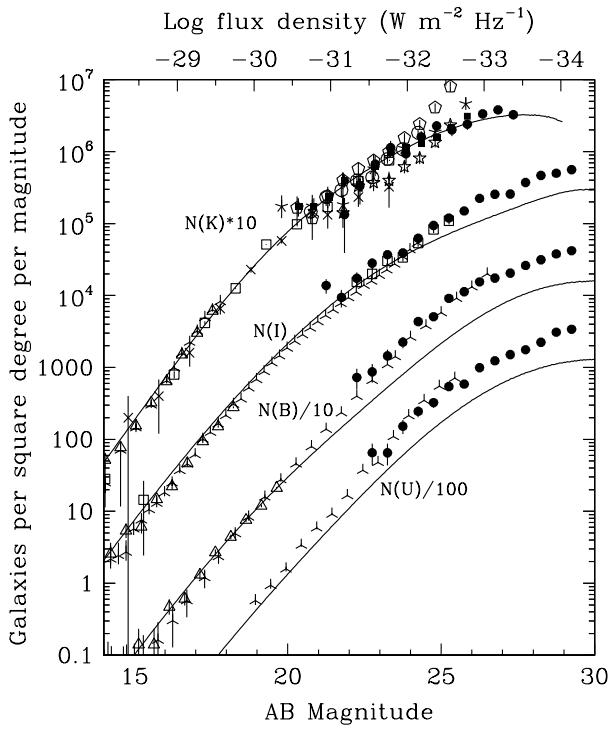


Figure 2. The excess of blue galaxies in the faint counts. The vertical axis reports the surface density of galaxies in the sky in units of galaxies per square degree per magnitude interval. The horizontal axis on the bottom reports the magnitude (faint is to the right, bright is to the left), the one on the top reports the luminosity in logarithmic scale expressed in physical units of Watts per square meter per Hertz. Counts in the K , I , B and U bands are plotted. Different symbols denote different observations. The curves are the theoretical predictions of the counts assuming that the galaxies of the present-day universe have not evolved. The faint excess is very strong in the U and B band and becomes progressively weaker in the I and K bands. (From Ferguson 1998.)

the galaxies, and hence their distance. This, in turn, made it possible to measure the intrinsic luminosity of the galaxies, and hence to derive their luminosity function at various cosmic epochs, mapping its evolution in time (Lilly *et al* 1995). These works showed that most of the faint blue galaxies observed in the imaging surveys are intrinsically relatively small systems located at moderate to intermediate redshifts, i.e. $z \sim 0.3$ to $z \sim 1$, which are undergoing a robust activity of star formation. Their closest counterparts in the local universe are the irregular galaxies, similar to the Magellanic Clouds. Interestingly, the luminosity function of blue galaxies was found to evolve very rapidly from the present epoch up to the earliest epoch studied ($z \sim 1.3$), implying that the luminosity (and hence the star formation rate) and abundance of these galaxies were progressively higher and higher in the past.

The observations also showed that the luminosity function of galaxies that have the same intrinsic spectral

energy distribution as the ellipticals and spirals of the local universe, hence which are redder than the faint blue galaxies, is characterized by a much more modest evolution over the same cosmic epochs. This is shown in figure 3, which plots the measures of the luminosity function of red and blue galaxies in a number of redshift intervals, as specified in the labels (note that there are no red galaxies in the range $z = 1.00$ to $z = 1.30$). The vertical axes represent the number density in units of galaxies per cubic Mpc, the horizontal ones the luminosity expressed as absolute magnitudes (this is the apparent magnitude a source would appear to have if placed at a distance of 10 pc from the observer). The circles show the data, the continuous line is the fit to the data, the dashed line is the fit to the data of the redshift bin $z = 0.2$ to $z = 0.5$ (used to show the relative evolution), and the dotted line is the fit to the luminosity function of local galaxies. Thus, galaxies with colors similar to those of the Milky Way (or redder) were already formed and in place by redshift $z \sim 1$ (namely for the last two-thirds of cosmic time) and have been evolving very modestly since then. Blue galaxies, on the other hand, have undergone a spectacular evolution over the same epoch, progressively diminishing their abundances and dimming their luminosity (i.e. reducing the star formation rate) towards the present time.

Solving one piece of the puzzle

With the repair of the *Hubble Space Telescope* (*HST*) in 1994 it eventually became possible to image distant galaxies with enough spatial resolution to classify their morphology and compare it to that of the local galaxies (Abraham *et al* 1996). The unprecedented angular resolution of the *HST* imaging has shown that galaxies up to redshifts $z \sim 1$ have morphologies that can still be classified according to the traditional types of ellipticals, lenticulars, spirals and irregulars (Glazebrook *et al* 1995). Consistently with what was found in the redshift surveys, the *HST* imaging surveys showed that elliptical and lenticular galaxies underwent essentially only passive evolution⁵ during the last half of the universe lifetime (Schade *et al* 1999). Over the same period of time, the disks of spiral galaxies showed some moderate evolution induced by an overall increased of the activity of star formation, although their abundance was found to be comparable to that in the present-day universe (Lilly *et al* 1999). In sharp contrast to the modest evolution of ellipticals and spirals, however, the *HST* images showed that irregular galaxies underwent a spectacular amount of evolution, both in luminosity and abundance, over the same period of time (Ellis 1997). Thus, *HST* has clarified the nature of the faint blue galaxies by showing that they consist of irregular galaxies, and has provided an independent confirmation that these systems were much more abundant in the past than they are today, and that they were also forming stars at a higher pace. These galaxies, in effect, have been the drivers of the

⁵ The expression ‘passive evolution’ denotes the evolution in luminosity of a galaxy induced by the natural evolution of its stars

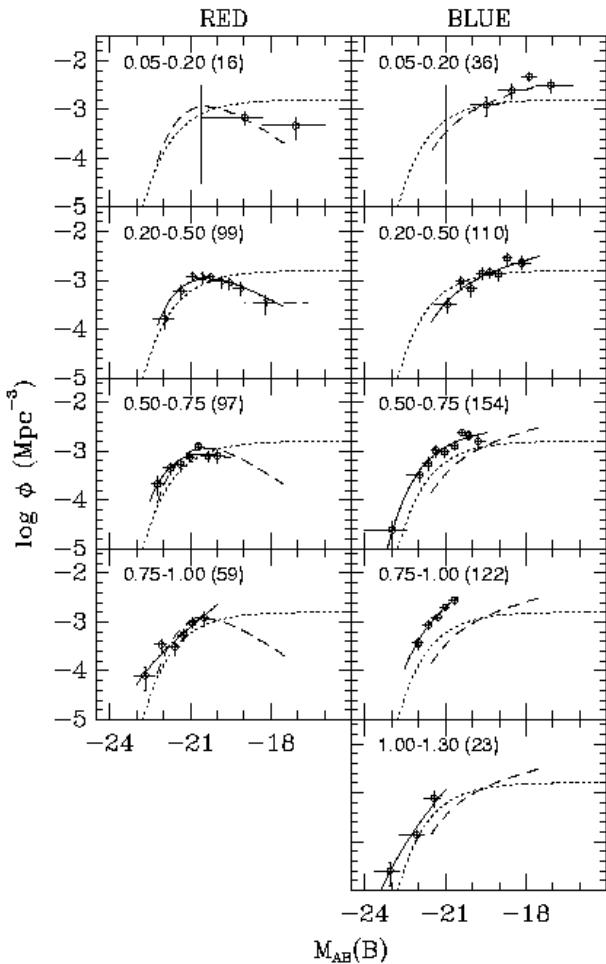


Figure 3. The evolution of the luminosity function of red and blue galaxies. Each panel contains measure of the luminosity function in the specified redshift range (data points) and a mathematical fit to the data point (continuous curve). Also shown is the luminosity function of local galaxies (dotted line) and the fit to the data of the redshift bin $z = 0.20-z = 0.50$ (dashed line), useful to see the differential evolution. The vertical axes represent the volume density of galaxies in units of galaxies per cubic Mpc. The horizontal represents the intrinsic luminosity, expressed in absolute magnitudes in the B band. The absolute magnitude is the magnitude a source would appear to have if placed at a distance of 10 pc from the observer. (From Lilly *et al* 1995.)

evolution of the overall luminosity function (i.e. that of all galaxies, regardless of their morphology or colors) and of the star formation history of the universe during the last 8 billion years of cosmic age. Figure 4 shows four examples of faint blue galaxies imaged with *HST*. Note the highly irregular and fragmented morphologies, which consist of compact regions that host active star formation embedded in more diffuse structures composed by a mixture of new and older stars and of gas and dust.

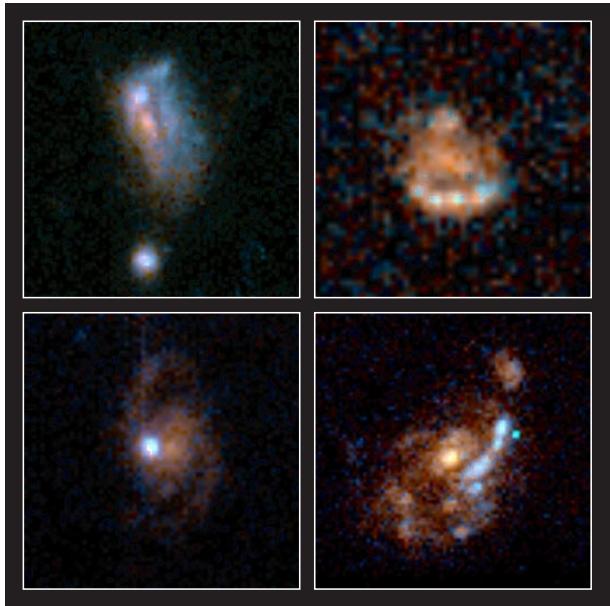


Figure 4. Four images of faint blue galaxies taken with the *Hubble Space Telescope* (Griffith *et al* 1998). Images like these helped understanding the nature of these population of galaxies, which underwent a spectacular evolution in luminosity and volume density over the last 8 billion years of the cosmic time. These galaxies are very similar to the Irregular types found in the present-day universe, however they were much more abundant in the past. Where the descendants of these rapidly evolving galaxies are at the present epoch is still precisely not known. (Photo credit: STScI OPO, R Griffiths (Carnegie Mellon University and NASA).)

Looking at the big picture

The study of galaxies over the last two-thirds of the cosmic age has shown clear patterns in the evolution of galaxies. What are today small galaxies seem to have undergone most of their evolution over this time. Although we do not yet know with certainty what and where the present-day counterparts of the faint blue galaxies are at the present epoch, there is mounting evidence that these systems have gradually merged with larger galaxies and have disappeared. As we have discussed earlier, the accretion of a small system onto a large one does not alter the properties of the latter as when a major merging event between two similar systems takes place. Another possibility that has been proposed is that at least a fraction of them have faded beyond detectability of the present instrumentation, once the burst of star formation has ended.

Over the same cosmic time, large galaxies such as ellipticals and spirals evolved much more modestly, being essentially already in place by redshift $z \sim 1$. This implies that they completed most of their formation at earlier epochs, and that to understand their evolution one must search at even earlier epochs.

The search for the progenitors of the so-called 'bright galaxies' has traditionally occupied a central role in observational cosmology (see the article on GALAXIES AT HIGH

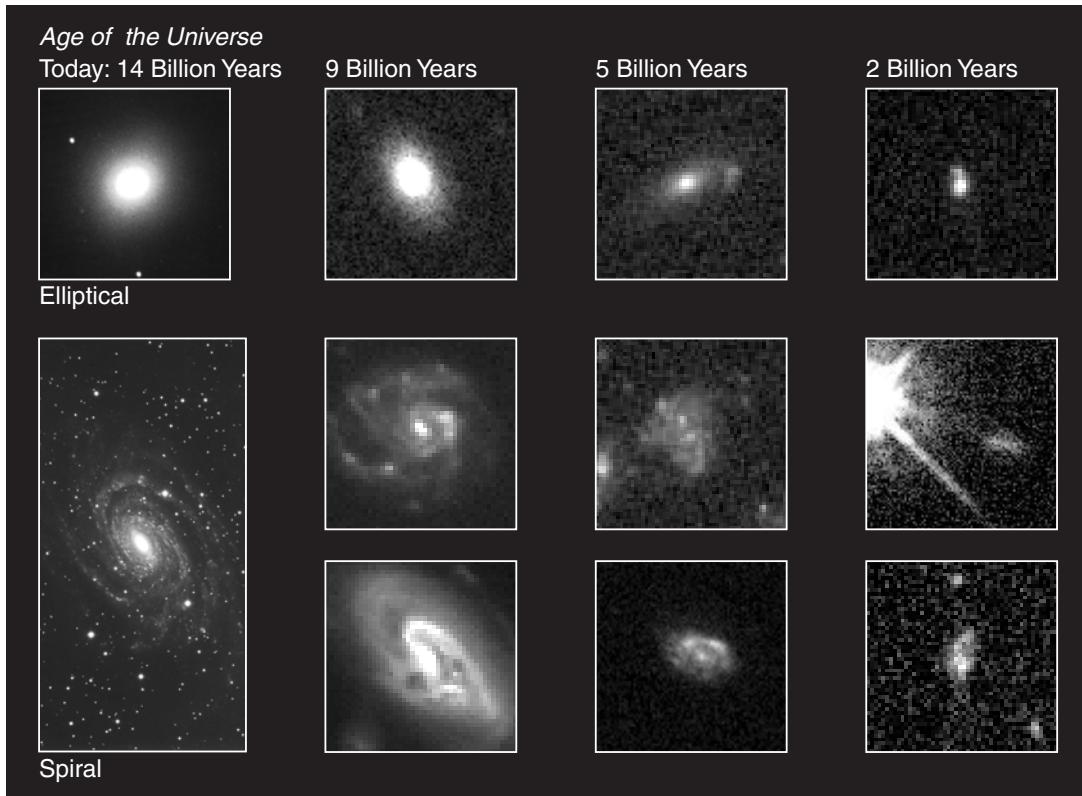


Figure 5. Comparison between galaxies observed at various cosmic epochs (expressed as time after the Big Bang). The traditional morphological types can be recognized already at ~ 5 billion years. At earlier epochs the morphology seems to differ from that of later epochs, although galaxies do exhibit various degrees of compactness and irregularity in their morphologies, ranging from galaxies with light profiles similar to that of ellipticals and bulges to those with disk-like morphology, in resemblance to the Hubble types. (Photo credit: STScI OPO, M Dickinson, M Giavalisco and NASA.)

REDSHIFT in this volume). One of the most important results of the extragalactic research of the last five years has been the identification of a massive population of star-forming galaxies at redshifts larger than $z \sim 2.5$, observed at a time when the universe was younger than 20% of its current age (Steidel *et al* 1996; Giavalisco *et al* 1996). These galaxies, also known as Lyman-break galaxies, are observed during an intense activity of star formation which has contributed a significant fraction of all the stars that we count in the present-day universe, between 20% and 50% according to current measures. Thus, the stars formed in these galaxies can account for a large fraction (if not all) of the stellar mass currently contained in spheroids since, having formed so early in the cosmic evolution, they must necessarily be very old today. These arguments strongly support the idea that the Lyman-break galaxies are the direct progenitors of the present-day spheroids, observed at a time when they were forming the bulk of their stars.

Unfortunately, because of the gap in the observations between redshift $z \sim 1$ and $z \sim 2.5$, this association currently cannot be tested by directly following the evolution of the galaxies at progressively lower redshifts. Equally unknown is whether the Lyman-break galaxies

were the first to appear in the universe or if other galaxies preceded them in evolution. These will be major activities of research when instruments of improved sensitivity for ground-based telescopes and the Next Generation Space Telescope will come on line.

Figure 5 shows a comparison between elliptical and spiral galaxies of the present-day universe (observed with ground-based telescopes) with their counterparts observed at progressively higher look-back times (observed with *HST*). The galaxies in the right panels are Lyman-break galaxies at redshift $z \sim 3$. Note that for these objects, a classification into morphological types is not really possible. However, in resemblance to the Hubble sequence, these galaxies do show various degrees of compactness and irregularity in their morphologies, ranging from systems with the light profile similar to that of ellipticals and bulges to systems with disk-like morphology. In the three panels on the right of figure 5 we have represented the cases of a galaxy with the same type of light profile as a present-day spheroid (top), of one with the same light profile as a disk (middle), and of one with an irregular profile (bottom).

One possible evolutionary path is that the Lyman-break galaxies have formed the bulk of their stars in intense

bursts of star formation at high redshifts and then evolved mostly passively. In this scenario, the final outcome will be an elliptical galaxy when at the end of the collapse the progenitor has exhausted all the gas reservoir. Or it will be a spiral galaxy when the proto-bulge has still sufficient gas around it (or it is able to strip gas from another galaxy it happens to interact with), to give origin to a disk where star formation can occur with low rate and over a prolonged period of time.

Another possibility is that the Lyman-break galaxies have evolved into the present-day galaxies through successive generations of hierarchical merging. In such a scenario, elliptical galaxies and bulges are easily explained, since we have seen that the outcome of a major merging episode is a spheroid. In order to form lenticular and spiral galaxies, it is necessary that, following the merging episode that originated the bulge, the nascent galaxy has not only a sufficient gas reservoir but also the possibility to evolve undisturbed, with no further merging events, so that the disk can develop (disks are destroyed in a merging event). According to this scheme, elliptical galaxies would be spheroids that have not developed a disk, either because of the lack of gas or because they formed recently and simply did not have the time to do so. They would be, therefore, structurally 'young' systems, even if their stellar populations are old. Spirals, on the other hand, must have had the time to develop the disk, and thus they are 'structurally' old even if they contain a large fraction of young stars in the disk. Interestingly, while the data show that the occurrence of merging seems to be higher in distant galaxies, its relative importance in the evolution of galaxies of different types has not yet been quantified. Thus, it is not known, from an empirical point of view, for which galaxies merging has been instrumental for their formation, and for which it only provided a 'perturbation' in an otherwise structurally passive evolution.

The picture that seems to emerge is that galaxy evolution proceeded along different paths, according to the different types of the galaxies, with massive galaxies forming first and smaller systems later. In other words, the mass seems to be the key parameter that decides the evolutionary destiny of galaxies. This is consistent with the paradigm that gravity has been the primary mechanism for the formation of structure at all scales in the universe, including galaxies. The paradigm states that the cosmic structures form as a result of the amplification and gravitational collapse of density perturbations in the mass distribution. Since perturbations grow as a result of their self-gravity, a prediction of the paradigm is that more massive perturbations grow faster than smaller ones, and hence they can virialize and begin the process of star formation sooner.

The importance of this paradigm goes beyond explaining the formation of galaxies, since it forms the conceptual ground for our understanding of many of the observations in cosmology and, ultimately, of the evolution of the universe as a whole. Yet, despite its

importance, this paradigm has never been subject to a rigorous empirical test. Reconstructing the history of the evolution of galaxies from their earliest appearance in space to the present time will provide one of the first and most important empirical constraints to the whole theoretical framework of cosmology.

Bibliography

- Abraham R *et al* 1996 *Astrophys. J.* **107** 1
 Kennicutt R 1988 *Ann. Rev. Astron. Astrophys.* **36** 189
 Ellis R S 1997 *Ann. Rev. Astron. Astrophys.* **35** 389
 Ferguson H C 1998 The Hubble Deep Field *Proc. STScI May Symp.* ed M Livio, S M Fall and P Madau p 181
 Giavalisco M, Steidel C C and Macchetto F D 1996 *Astrophys. J.* **470** 189
 Glazebrook K, Ellis R, Santiago B and Griffiths R 1995 *Mon. Not. R. Astron. Soc.* **275** 19
 Lilly S J *et al* 1999 *Astrophys. J.* **500** 75
 Lilly S J *et al* 1995 *Astrophys. J.* **455** 108
 Schade D *et al* 1999 *Astrophys. J.* **525** 31
 Steidel C C, Giavalisco M, Pettini M, Dickinson M and Adelberger K 1996 *Astrophys. J.* **462** 17
 Tyson J A 1988 *Astrophys. J.* **96** 1

Mauro Giavalisco