3. The World of Galaxies

The insight that our Milky Way is just one of many galaxies in the Universe is less than 100 years old, despite the fact that many had already been known for a long time. The catalog by Charles Messier (1730–1817), for instance, lists 103 diffuse objects. Among them M31, the Andromeda galaxy, is listed as the 31st entry in the Messier catalog. Later, this catalog was extended to 110 objects. John Dreyer (1852–1926) published the *New General Catalog (NGC)* that contains nearly 8000 objects, most of them galaxies. In 1912, Vesto Slipher found that the spiral nebulae are rotating, using spectroscopic analysis. But the nature of these extended sources, then called nebulae, was still unknown at that time; it was unclear whether they are part of our Milky Way or outside it.

The year 1920 saw a public debate (the Great Debate) between Harlow Shapley and Heber Curtis. Shapley believed that the nebulae are part of our Milky Way, whereas Curtis was convinced that the nebulae must be objects located outside the Galaxy. The arguments which the two opponents brought forward were partly based on assumptions which later turned out to be invalid, as well as on incorrect data. We will not go into the details of their arguments which were partially linked to the assumed size of the Milky Way since, only a few years later, the question of the nature of the nebulae was resolved.

In 1925, Edwin Hubble discovered Cepheids in Andromeda (M31). Using the period-luminosity relation for these pulsating stars (see Sect. 2.2.7) he derived a distance of 285 kpc. This value is a factor of ~ 3 smaller than the distance of M31 known today, but it provided clear evidence that M31, and thus also other spiral nebulae, must be extragalactic. This then immediately implied that they consist of innumerable stars, like our Milky Way. Hubble's results were considered conclusive by his contemporaries and marked the beginning of extragalactic astronomy. It is not coincidental that at this time George Hale began to arrange the funding for an ambitious project. In 1928 he obtained six

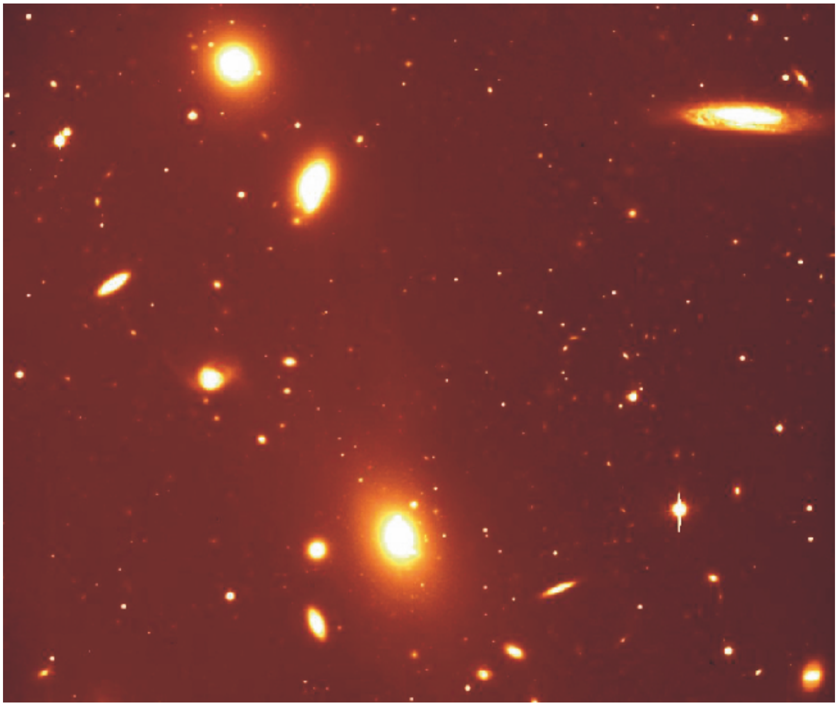

Fig. 3.1. Galaxies occur in different shapes and sizes, and often they are grouped together in groups or clusters. This cluster, ACO 3341, at a redshift of $z = 0.037$, contains numerous galaxies of different types and luminosities

Peter Schneider, The World of Galaxies.
In: Peter Schneider, Extragalactic Astronomy and Cosmology. pp. 87–140 (2006)
DOI: 10.1007/11614371_3 © Springer-Verlag Berlin Heidelberg 2006

million dollars for the construction of the 5-m telescope on Mt. Palomar which was completed in 1949.

This chapter is about galaxies. We will confine the consideration here to "normal" galaxies in the local Universe; galaxies at large distances, some of which are in a very early evolutionary state, will be discussed in Chap. 9, and active galaxies, like quasars for example, will be discussed later in Chap. 5.

3.1 Classification

The classification of objects depends on the type of observation according to which this classification is made. This is also the case for galaxies. Historically, optical photometry was the method used to observe galaxies. Thus, the morphological classification defined by Hubble is still the best-known today. Besides morphological criteria, color indices, spectroscopic parameters (based on emission or absorption lines), the broad-band spectral distribution (galaxies with/without radio- and/or X-ray emission), as well as other features may also be used.

3.1.1 Morphological Classification: The Hubble Sequence

Figure 3.2 shows the classification scheme defined by Hubble. According to this, three main types of galaxies exist:

- *Elliptical galaxies* (E's) are galaxies that have nearly elliptical isophotes[1] without any clearly defined structure. They are subdivided according to their ellipticity $\epsilon \equiv 1 - b/a$, where a and b denote the semimajor and the semiminor axes, respectively. Ellipticals are found over a relatively broad range in ellipticity, $0 \leq \epsilon \lesssim 0.7$. The notation E$n$ is commonly used to classify the ellipticals with respect to ϵ, with $n = 10\epsilon$; i.e., an E4 galaxy has an axis ratio of $b/a = 0.6$, and E0's have circular isophotes.
- *Spiral galaxies* consist of a disk with spiral arm structure and a central bulge. They are divided into two subclasses: *normal spirals* (S's) and *barred spirals* (SB's). In each of these subclasses, a sequence is defined that is ordered according to the brightness ratio of bulge and disk, and that is denoted by a, ab, b, bc, c, cd, d. Objects along this sequence are often referred to as being either an early-type or a late-type; hence, an Sa galaxy is an early-type spiral, and an SBc galaxy is a late-type barred spiral. We stress explicitly that this nomenclature is not a statement of the evolutionary stage of the objects but is merely a nomenclature of purely historical origin.
- *Irregular galaxies* (Irr's) are galaxies with only weak (Irr I) or no (Irr II) regular structure. The classification of Irr's is often refined. In particular, the sequence of spirals is extended to the classes Sdm, Sm, Im, and Ir (m stands for Magellanic; the Large Magellanic Cloud is of type SBm).
- *S0 galaxies* are a transition between ellipticals and spirals. They are also called lenticulars as they are lentil-shaped galaxies which are likewise subdivided into S0 and SB0, depending on whether or not they show a bar. They contain a bulge and a large enveloping region of relatively unstructured brightness which often appears like a disk without spiral arms. Ellipticals and S0 galaxies are referred to as early-type galaxies, spirals as late-type galaxies. As before,

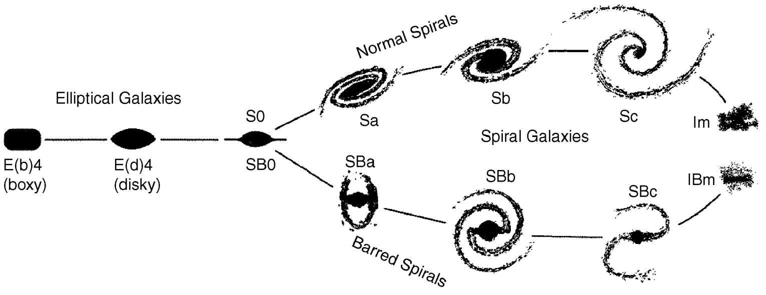

Fig. 3.2. Hubble's "tuning fork" for galaxy classification

[1] Isophotes are contours along which the surface brightness of a sources is constant. If the light profile of a galaxy is elliptical, then its isophotes are ellipses.

these names are only historical and are not meant to describe an evolutionary track!

Obviously, the morphological classification is at least partially affected by projection effects. If, for instance, the spatial shape of an elliptical galaxy is a triaxial ellipsoid, then the observed ellipticity ϵ will depend on its orientation with respect to the line-of-sight. Also, it will be difficult to identify a bar in a spiral that is observed from its side ("edge-on").

Besides the aforementioned main types of galaxy morphologies, others exist which do not fit into the Hubble scheme. Many of these are presumably caused by interaction between galaxies (see below). Furthermore, we observe galaxies with radiation characteristics that differ significantly from the spectral behavior of "normal" galaxies. These galaxies will be discussed next.

3.1.2 Other Types of Galaxies

The light from "normal" galaxies is emitted mainly by stars. Therefore, the spectral distribution of the radiation from such galaxies is in principle a superposition of the spectra of their stellar population. The spectrum of stars is, to a first approximation, described by a Planck function (see Appendix A) that depends only on the star's surface temperature. A typical stellar population covers a temperature range from a few thousand Kelvin up to a few tens of thousand Kelvin. Since the Planck function has a well-localized maximum and from there steeply declines to both sides, most of the energy of such "normal" galaxies is emitted in a relatively narrow frequency interval that is located in the optical and NIR sections of the spectrum.

In addition to these, other galaxies exist whose spectral distribution cannot be described by a superposition of stellar spectra. One example is the class of active galaxies which generate a significant fraction of their luminosity from gravitational energy that is released in the infall of matter onto a supermassive black hole, as was mentioned in Sect. 1.2.4. The activity of such objects can be recognized in various ways. For example, some of them are very luminous in the radio and/or in the X-ray portion of the spectrum (see Fig. 3.3), or they show strong emission lines with a width of several thousand km/s if the line width is interpreted as due to Doppler broadening, i.e., $\Delta\lambda/\lambda = \Delta v/c$. In many cases, by far the largest fraction of luminosity is produced in a very small central region: the active galactic nucleus (AGN) that gave this class of galaxies its name. In quasars, the central luminosity can be of the order of $\sim 10^{13} L_\odot$, about a thousand times as luminous as the total luminosity of our Milky Way. We will discuss active galaxies, their phenomena, and their physical properties in detail in Chap. 5.

Another type of galaxy also has spectral properties that differ significantly from those of "normal" galaxies, namely the starburst galaxies. Normal spiral galaxies like our Milky Way form new stars at a star-formation rate of $\sim 3 M_\odot$/yr which can be derived, for instance, from the Balmer lines of hydrogen generated in the HII regions around young, hot stars. By contrast, elliptical galaxies show only marginal star formation or none at all. However, there are galaxies which have a much higher star-formation rate, reaching values of

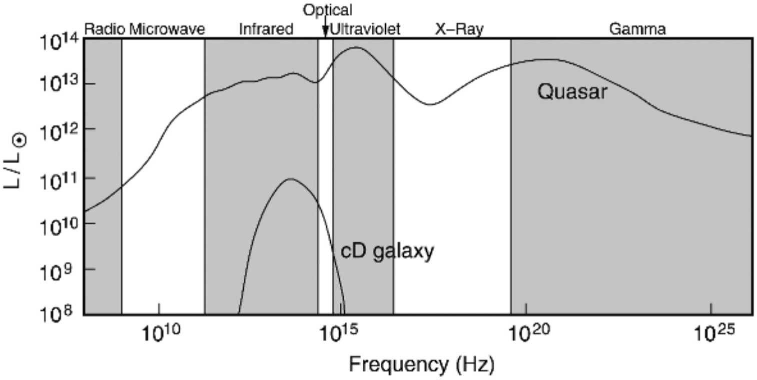

Fig. 3.3. The spectrum of a quasar (3C273) in comparison to that of an elliptical galaxy. While the radiation from the elliptical is concentrated in a narrow range spanning less than two decades in frequency, the emission from the quasar is observed over the full range of the electromagnetic spectrum, and the energy per logarithmic frequency interval is roughly constant. This demonstrates that the light from the quasar cannot be interpreted as a superposition of stellar spectra, but instead has to be generated by completely different sources and by different radiation mechanisms

$100 M_\odot/\mathrm{yr}$ and more. If many young stars are formed we would expect these starburst galaxies to radiate strongly in the blue or in the UV part of the spectrum, corresponding to the maximum of the Planck function for the most massive and most luminous stars. This expectation is not fully met though: star formation takes place in the interior of dense molecular clouds which often also contain large amounts of dust. If the major part of star formation is hidden from our direct view by layers of absorbing dust, these galaxies will not be very prominent in blue light. However, the strong radiation from the young, luminous stars heats the dust; the absorbed stellar light is then emitted in the form of thermal dust emission in the infrared and submillimeter regions of the electromagnetic spectrum – these galaxies can thus be extremely luminous in the IR. They are called ultra-luminous infrared galaxies (ULIRGs). We will describe the phenomena of starburst galaxies in more detail in Sect. 9.2.1. Of special interest is the discovery that the star-formation rate of galaxies seems to be closely related to interactions between galaxies – many ULIRGs are strongly interacting galaxies (see Fig. 3.4).

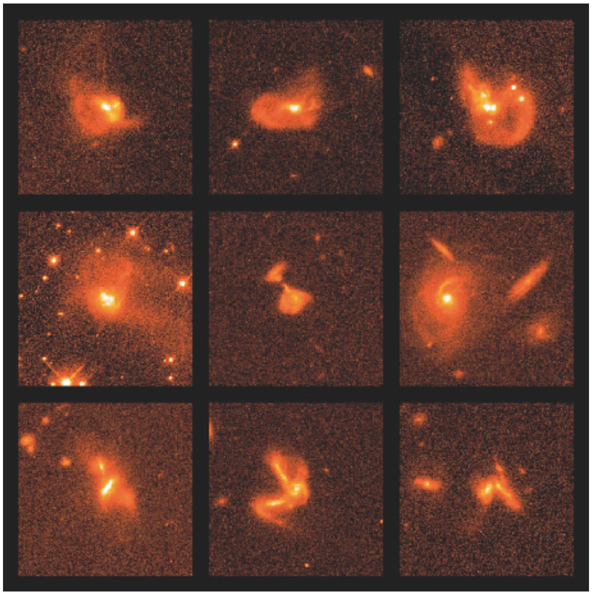

Fig. 3.4. This mosaic of nine HST images shows different ULIRGs in collisional interaction between two or more galaxies

3.2 Elliptical Galaxies

3.2.1 Classification

The general term “elliptical galaxies” (or ellipticals, for short) covers a broad class of galaxies which differ in their luminosities and sizes – some of them are displayed in Fig. 3.5. A rough subdivision is as follows:

- *Normal ellipticals*. This class includes giant ellipticals (gE’s), those of intermediate luminosity (E’s), and compact ellipticals (cE’s), covering a range in absolute magnitudes from $M_B \sim -23$ to $M_B \sim -15$. In addition, S0 galaxies are often assigned to this class of early-type galaxies.
- *Dwarf ellipticals* (dE’s). These differ from the cE’s in that they have a significantly smaller surface brightness and a lower metallicity.
- *cD galaxies*. These are extremely luminous (up to $M_B \sim -25$) and large (up to $R \lesssim 1\,\mathrm{Mpc}$) galaxies that are only found near the centers of dense clusters of galaxies. Their surface brightness is very high close to the center, they have an extended diffuse envelope, and they have a very high M/L ratio.
- *Blue compact dwarf galaxies*. These “blue compact dwarfs” (BCD’s) are clearly bluer (with $\langle B-V \rangle$ between 0.0 and 0.3) than the other ellipticals, and contain an appreciable amount of gas in comparison.
- *Dwarf spheroidals* (dSph’s) exhibit a very low luminosity and surface brightness. They have been observed down to $M_B \sim -8$. Due to these properties, they have thus far only been observed in the Local Group.

Thus elliptical galaxies span an enormous range (more than 10^6) in luminosity and mass, as is shown by the compilation in Table 3.1.

3.2.2 Brightness Profile

The brightness profiles of normal E’s and cD’s follow a de Vaucouleurs profile (see (2.39) or (2.41), respectively) over a wide range in radius, as is demonstrated in Fig. 3.6. The effective radius R_e is strongly correlated with the absolute magnitude M_B, as can be seen in Fig. 3.7, with rather little scatter. In comparison, the dE’s and the dSph’s clearly follow a different distribution. Owing to the relation (2.42) between luminosity,

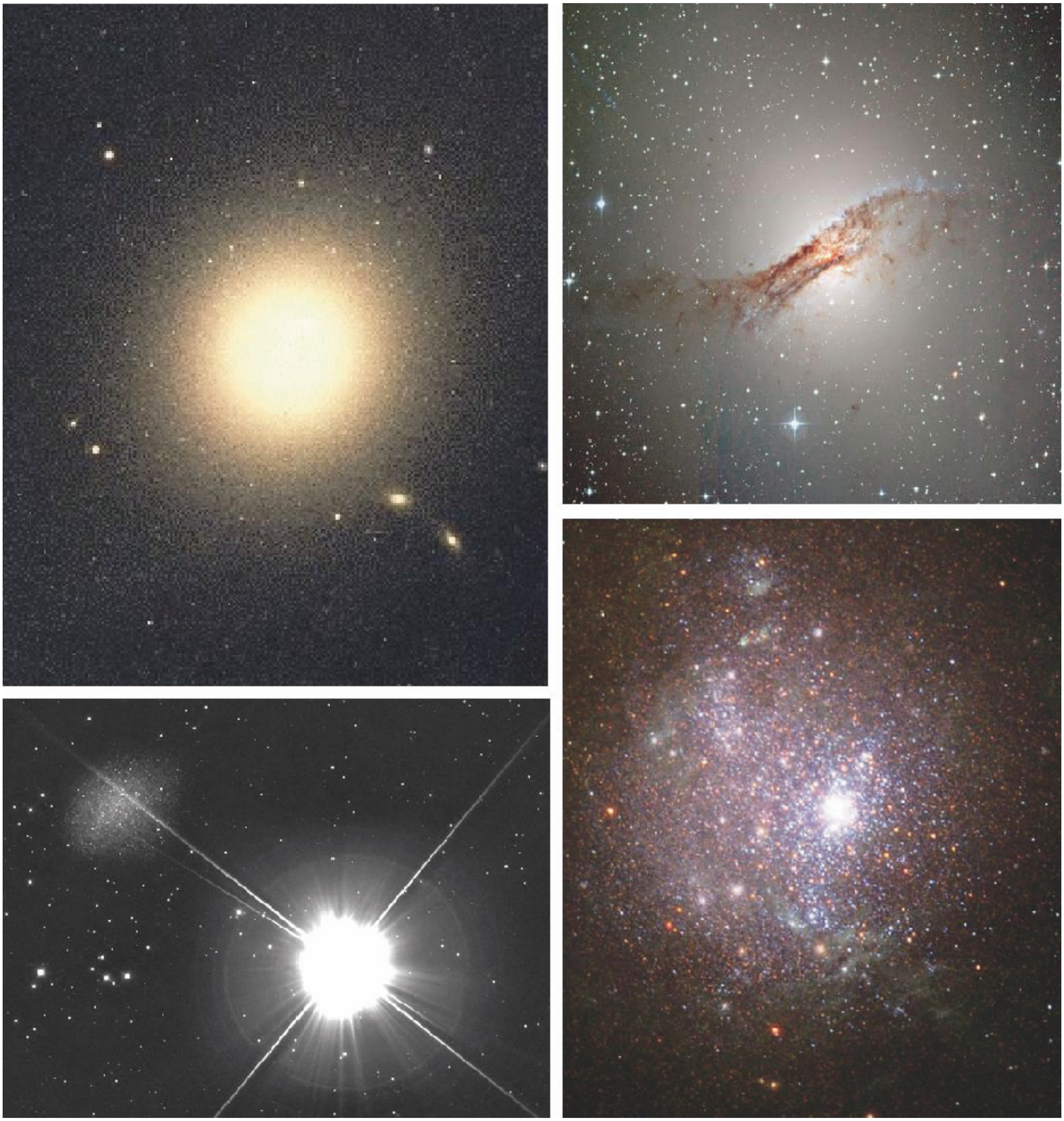

Fig. 3.5. Different types of elliptical galaxies. Upper left: the cD galaxy M87 in the center of the Virgo galaxy cluster; upper right: Centaurus A, a giant elliptical galaxy with a very distinct dust disk and an active galactic nucleus; lower left: the galaxy Leo I belongs to the nine known *dwarf spheroidals* in the Local Group; lower right: NGC 1705, a dwarf irregular, shows indications of massive star formation – a super star cluster and strong galactic winds

Table 3.1. Characteristic values for elliptical galaxies. D_{25} denotes the diameter at which the surface brightness has decreased to 25 B-mag/arcsec^2, $S_{\rm N}$ is the "specific frequency", a measure for the number of globular clusters in relation to the visual luminosity (see (3.13)), and M/L is the mass-to-light ratio in Solar units (the values of this table are taken from the book by Carroll & Ostlie, 1996)

	S0	cD	E	dE	dSph	BCD
M_B	-17 to -22	-22 to -25	-15 to -23	-13 to -19	-8 to -15	-14 to -17
$M(M_\odot)$	10^{10} to 10^{12}	10^{13} to 10^{14}	10^{8} to 10^{13}	10^{7} to 10^{9}	10^{7} to 10^{8}	$\sim 10^{9}$
D_{25} (kpc)	10–100	300–1000	1–200	1–10	0.1–0.5	< 3
$\langle M/L_{\rm B}\rangle$	~ 10	> 100	10–100	1–10	5–100	0.1–10
$\langle S_{\rm N}\rangle$	~ 5	~ 15	~ 5	4.8 ± 1.0	–	–

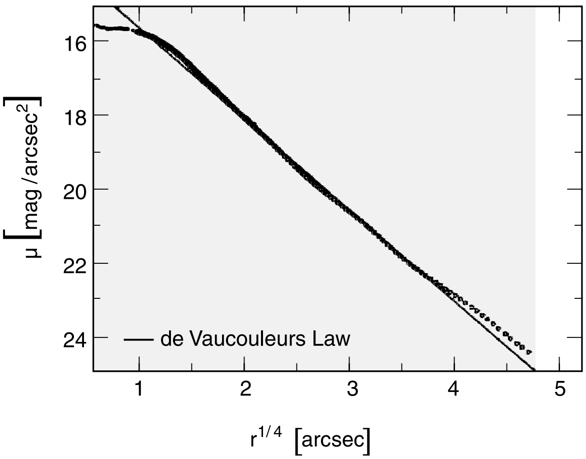

Fig. 3.6. Surface brightness profile of the galaxy NGC 4472, fitted by a de Vaucouleurs profile. The de Vaucouleurs profile describes a linear relation between the logarithm of the intensity (i.e., linear on a magnitude scale) and $r^{1/4}$; for this reason, it is also called an $r^{1/4}$-law

effective radius and central surface brightness, an analogous relation exists for the average surface brightness $\mu_{\rm ave}$ (unit: $\rm B - mag/arcsec^2$) within $R_{\rm e}$ as a function of M_B. In particular, the surface brightness in normal E's decreases with increasing luminosity, while it increases for dE's and dSph's.

Yet another way of expressing this correlation is by eliminating the absolute luminosity, thus obtaining a relation between effective radius $R_{\rm e}$ and surface brightness $\mu_{\rm ave}$. This form is then called the Kormendy relation.

The de Vaucouleurs profile provides the best fits for normal E's, whereas for E's with exceptionally high (or low) luminosity the profile decreases more slowly (or rapidly) for larger radii. The profile of cD's extends much farther out and is not properly described by a de Vaucouleurs profile (Fig. 3.8), except in its innermost part. It appears that cD's are similar to E's but embedded in a very extended, luminous halo. Since cD's are only found in the centers of massive clusters of galaxies, a connection must exist between this morphology and the environment of these galaxies. In contrast to these classes of ellipticals, diffuse dE's are often better described by an exponential profile.

3.2.3 Composition of Elliptical Galaxies

Except for the BCD's, elliptical galaxies appear red when observed in the optical, which suggests an old stellar population. It was once believed that ellipticals contain neither gas nor dust, but these components have now been found, though at a much lower mass-fraction than in spirals. For example, in some ellipticals hot gas ($\sim 10^7$ K) has been detected by its X-ray emission. Furthermore, Hα emission lines of warm gas ($\sim 10^4$ K) have been observed, as well as cold gas (~ 100 K) in the HI (21-cm) and CO molecular lines. Many of the normal ellipticals contain visible amounts of dust, partially manifested as a dust disk. The metallicity of ellipticals and S0 galaxies increases towards the galaxy center, as derived from color gradients. Also in S0 galaxies the bulge appears redder than the disk. The Spitzer Space Telescope, launched in 2003, has detected a spatially extended distribution of warm dust in S0 galaxies, organized in some sort of spiral structure. Cold dust has also been found in ellipticals and S0 galaxies.

This composition of ellipticals clearly differs from that of spiral galaxies and needs to be explained by mod-

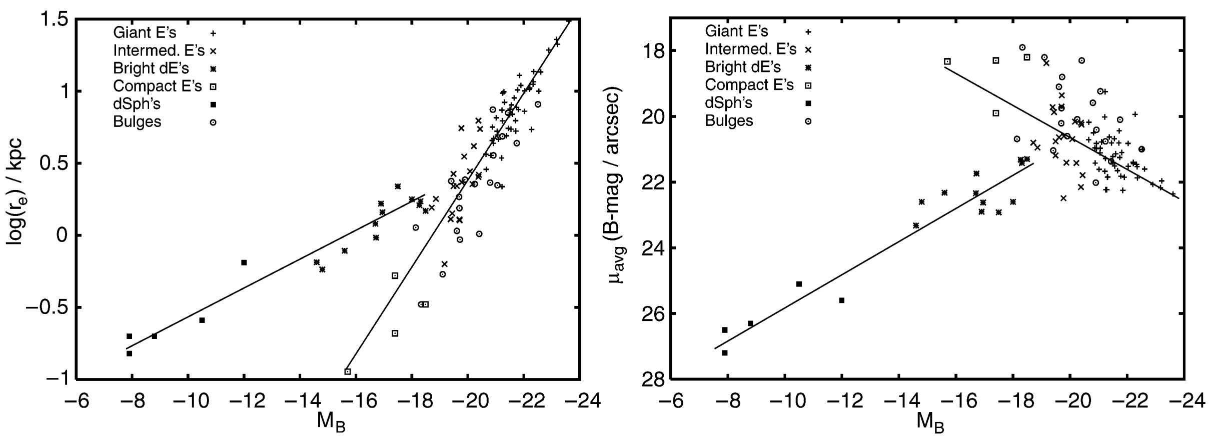

Fig. 3.7. Left panel: effective radius R_e versus absolute magnitude M_B; the correlation for normal ellipticals is different from that of dwarfs. Right panel: average surface brightness μ_{ave} versus M_B; for normal ellipticals, the surface brightness decreases with increasing luminosity while for dwarfs it increases

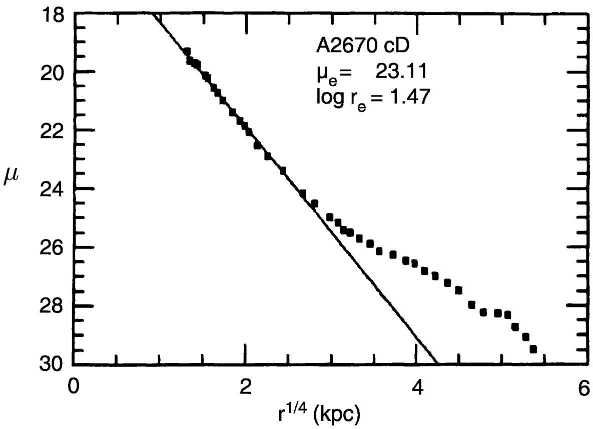

Fig. 3.8. Comparison of the brightness profile of a cD galaxy, the central galaxy of the cluster of galaxies Abell 2670, with a de Vaucouleurs profile. The light excess for large radii is clearly visible

els of the formation and evolution of galaxies. We will see later that the cosmic evolution of elliptical galaxies is also observed to be different from that of spirals.

3.2.4 Dynamics of Elliptical Galaxies

Analyzing the morphology of elliptical galaxies raises a simple question: *Why are ellipticals not round?* A simple explanation would be rotational flattening, i.e., as in a rotating self-gravitating gas ball, the stellar distribution bulges outwards at the equator due to centrifugal forces, as is also the case for the Earth. If this explanation were correct, the rotational velocity v_{rot}, which is measurable in the relative Doppler shift of absorption lines, would have to be of about the same magnitude as the velocity dispersion of the stars σ_v that is measurable through the Doppler broadening of lines. More precisely, by means of stellar dynamics one can show that for the rotational flattening of an axially symmetric, oblate[2] galaxy, the relation

$$\left(\frac{v_{\text{rot}}}{\sigma_v}\right)_{\text{iso}} \approx \sqrt{\frac{\epsilon}{1-\epsilon}} \tag{3.1}$$

has to be satisfied, where "iso" indicates the assumption of an isotropic velocity distribution of the stars. However, for luminous ellipticals one finds that, in general, $v_{\text{rot}} \ll \sigma_v$, so that rotation cannot be the major cause of their ellipticity (see Fig. 3.9). In addition, many ellipticals are presumably triaxial, so that no unambiguous rotation axis is defined. Thus, luminous ellipticals are in general *not* rotationally flattened. For less luminous ellipticals and for the bulges of disk galaxies, however, rotational flattening can play an important role. The question remains of how to explain a stable elliptical distribution of stars without rotation.

[2] If $a \geq b \geq c$ denote the lengths of the major axes of an ellipsoid, then it is called an oblate spheroid (= rotational ellipsoid) if $a = b > c$, whereas a prolate spheroid is specified by $a > b = c$. If all three axes are different, it is called triaxial ellipsoid.

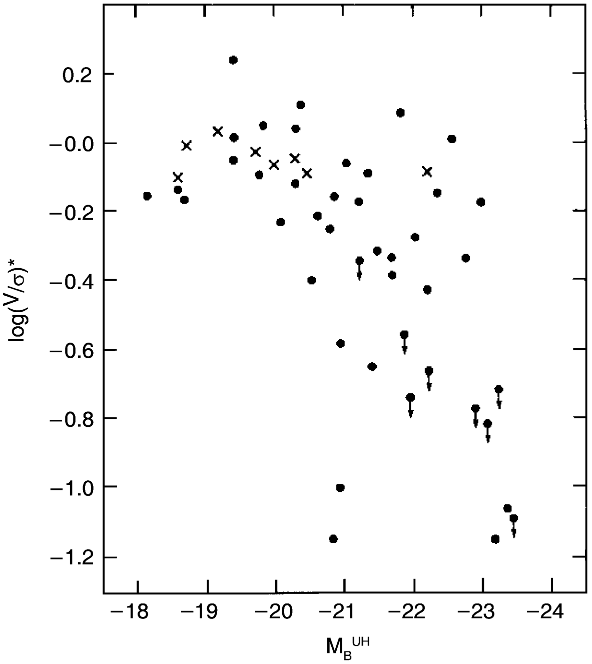

Fig. 3.9. The rotation parameter $\left(\frac{v_{\rm rot}}{\sigma_v}\right) / \left(\frac{v_{\rm rot}}{\sigma_v}\right)_{\rm iso}$ of elliptical galaxies, here denoted by $(V/\sigma)^*$, plotted as a function of absolute magnitude. Dots denote elliptical galaxies, crosses the bulges of disk galaxies

The brightness profile of an elliptical galaxy is defined by the distribution of its stellar orbits. Let us assume that the gravitational potential is given. The stars are then placed into this potential, with the initial positions and velocities following a specified distribution. If this distribution is not isotropic in velocity space, the resulting light distribution will in general not be spherical. For instance, one could imagine that the orbital planes of the stars have a preferred direction, but that an equal number of stars exists with positive and negative angular momentum L_z, so that the total stellar distribution has no angular momentum and therefore does not rotate. Each star moves along its orbit in the gravitational potential, where the orbits are in general not closed. If an initial distribution of stellar orbits is chosen such that the statistical properties of the distribution of the orbits are invariant in time, then one will obtain a stationary system. If, in addition, the distribution is chosen such that the respective mass distribution of the stars will generate exactly the originally chosen gravitational potential, one arrives at a self-gravitating equilibrium system. In general, it is a difficult mathematical problem to construct such self-gravitating equilibrium systems.

Relaxation Time-Scale. The question now arises whether such an equilibrium system can also be stable in time. One might expect that close encounters of pairs of stars would cause a noticeable disturbance in the distribution of orbits. These pair-wise collisions could then lead to a “thermalization” of the stellar orbits.[3] To examine this question we need to estimate the time-scale for such collisions and the changes in direction they cause.

For this purpose, we consider the relaxation time-scale by pair collisions in a system of N stars of mass m, total mass $M = Nm$, extent R, and a mean stellar density of $n = 3N/(4\pi R^3)$. We define the relaxation time $t_{\rm relax}$ as the characteristic time in which a star changes its velocity direction by $\sim 90°$ due to pair collisions with other stars. By simple calculation (see below), we find that

$$t_{\rm relax} \approx \frac{R}{v} \frac{N}{\ln N} , \tag{3.2}$$

or

$$\boxed{t_{\rm relax} = t_{\rm cross} \frac{N}{\ln N}} , \tag{3.3}$$

where $t_{\rm cross} = R/v$ is the crossing time-scale, i.e., the time it takes a star to cross the stellar system. If we now consider a typical galaxy, with $t_{\rm cross} \sim 10^8$ yr, $N \sim 10^{12}$ (thus $\ln N \sim 30$), then we find that the relaxation time is much longer than the age of the Universe. This means that *pair collisions do not play any role in the evolution of stellar orbits.* The dynamics of the orbits are determined solely by the large-scale gravitational field of the galaxy. In Sect. 7.5.1, we will describe a process called violent relaxation which most likely plays a central role in the formation of galaxies and which is probably also responsible for the stellar orbits establishing an equilibrium configuration.

The stars behave like a collisionless gas: elliptical galaxies are stabilized by (dynamical) pressure, and they are elliptical because the stellar distribution is

[3] Note that in a gas like air, scattering between molecules occurs frequently, which drives the velocity distribution of the molecules towards an isotropic Maxwellian, i.e., the thermal, distribution.

anisotropic in velocity space. This corresponds to an anisotropic pressure – where we recall that the pressure of a gas is nothing but the momentum transport of gas particles due to their thermal motions.

Derivation of the Collisional Relaxation Time-Scale. We consider a star passing by another one, with the impact parameter b being the minimum distance between the two. From gravitational deflection, the star attains a velocity component perpendicular to the incoming direction of

$$\boxed{v_\perp^{(1)} \approx a\,\Delta t \approx \left(\frac{Gm}{b^2}\right)\left(\frac{2b}{v}\right) = \frac{2Gm}{bv}}\,, \tag{3.4}$$

where a is the acceleration at closest separation and Δt the "duration of the collision", estimated as $\Delta t = 2b/v$ (see Fig. 3.10). Equation (3.4) can be derived more rigorously by integrating the perpendicular acceleration along the orbit. A star undergoes many collisions, through which the perpendicular velocity components will accumulate; these form two-dimensional vectors perpendicular to the original direction. After a time t we have $\boldsymbol{v}_\perp(t) = \sum_i \boldsymbol{v}_\perp^{(i)}$. The expectation value of this vector is $\langle \boldsymbol{v}_\perp(t)\rangle = \sum_i \left\langle \boldsymbol{v}_\perp^{(i)}\right\rangle = 0$ since the directions of the individual $\boldsymbol{v}_\perp^{(i)}$ are random. But the mean square velocity perpendicular to the incoming direction does not vanish,

$$\left\langle |\boldsymbol{v}_\perp|^2(t)\right\rangle = \sum_{ij}\left\langle \boldsymbol{v}_\perp^{(i)}\cdot\boldsymbol{v}_\perp^{(j)}\right\rangle = \sum_i \left\langle \left|\boldsymbol{v}_\perp^{(i)}\right|^2\right\rangle \neq 0\,, \tag{3.5}$$

where we set $\left\langle \boldsymbol{v}_\perp^{(i)}\cdot\boldsymbol{v}_\perp^{(j)}\right\rangle = 0$ for $i \neq j$ because the directions of different collisions are assumed to be uncorrelated. The velocity $\boldsymbol{v}_\perp$ performs a so-called *random walk*. To compute the sum, we convert it into an integral where we have to integrate over all collision parameters b. During time t, all collision partners with impact parameters within $\mathrm{d}b$ of b are located in a cylindrical shell of volume $(2\pi b\,\mathrm{d}b)\,(vt)$, so that

$$\begin{aligned}\left\langle |\boldsymbol{v}_\perp|^2(t)\right\rangle &= \int 2\pi\, b\,\mathrm{d}b\, v\, t\, n\, \left|v_\perp^{(1)}\right|^2 \\ &= 2\pi\left(\frac{2Gm}{v}\right)^2 v\,t\,n \int \frac{\mathrm{d}b}{b}\,.\end{aligned} \tag{3.6}$$

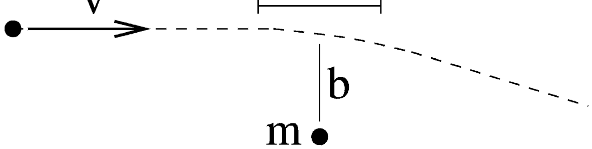

Fig. 3.10. Sketch related to the derivation of the dynamical time-scale

The integral cannot be performed from 0 to ∞. Thus, it has to be cut off at $b_{\min}$ and $b_{\max}$ and then yields $\ln(b_{\max}/b_{\min})$. Due to the finite size of the stellar distribution, $b_{\max} = R$ is a natural choice. Furthermore, our approximation which led to (3.4) will certainly break down if $v_\perp^{(1)}$ is of the same order of magnitude as v; hence we choose $b_{\min} = 2Gm/v^2$. With this, we obtain $b_{\max}/b_{\min} = Rv^2/(2Gm)$. The exact choice of the integration limits is not important, since $b_{\min}$ and $b_{\max}$ appear only logarithmically. Next, using the virial theorem, $|E_{\rm pot}| = 2E_{\rm kin}$, and thus $GM/R = v^2$ for a typical star, we get $b_{\max}/b_{\min} \approx N$. Thus,

$$\boxed{\left\langle |\boldsymbol{v}_\perp|^2(t)\right\rangle = 2\pi\left(\frac{2Gm}{v}\right)^2 v\,t\,n\,\ln N}\,. \tag{3.7}$$

We define the relaxation time $t_{\rm relax}$ by $\left\langle |\boldsymbol{v}_\perp|^2(t_{\rm relax})\right\rangle = v^2$, i.e., the time after which the perpendicular velocity roughly equals the infall velocity:

$$\begin{aligned} t_{\rm relax} &= \frac{1}{2\pi n v}\left(\frac{v^2}{2Gm}\right)^2 \frac{1}{\ln N} \\ &= \frac{1}{2\pi n v}\left(\frac{M}{2Rm}\right)^2 \frac{1}{\ln N} \approx \frac{R}{v}\frac{N}{\ln N}\,,\end{aligned} \tag{3.8}$$

from which we finally obtain (3.3).

3.2.5 Indicators of a Complex Evolution

The isophotes (that is, the curves of constant surface brightness) of many of the normal elliptical galaxies are well approximated by ellipses. These elliptical isophotes with different surface brightnesses are concentric to high accuracy, with the deviation of the isophote's center from the center of the galaxy being typically $\lesssim 1\%$ of its extent. However, in many cases the ellipticity varies with radius, so that the value for ϵ is not a constant. In addition, many ellipticals show a so-called isophote twist: the orientation of the semi-major axis of the isophotes changes with the radius.

This indicates that elliptical galaxies are not spheroidal, but triaxial systems (or that there is some intrinsic twist of their axes).

Although the light distribution of ellipticals appears rather simple at first glance, a more thorough analysis reveals that the kinematics can be quite complicated. For example, dust disks are not necessarily perpendicular to any of the principal axes, and the dust disk may rotate in a direction opposite to the galactic rotation. In addition, ellipticals may also contain (weak) stellar disks.

Boxiness and Diskiness. The so-called boxiness parameter describes the deviation of the isophotes' shape from that of an ellipse. Consider the shape of an isophote. If it is described by an ellipse, then after a suitable choice of the coordinate system, $\theta_1 = a \cos t$, $\theta_2 = b \sin t$, where a and b are the two semi-axes of the ellipse and $t \in [0, 2\pi]$ parametrizes the curve. The distance $r(t)$ of a point from the center is

$$r(t) = \sqrt{\theta_1^2 + \theta_2^2} = \sqrt{\frac{a^2+b^2}{2} + \frac{a^2-b^2}{2} \cos(2t)} \; .$$

Deviations of the isophote shape from this ellipse are now expanded in a Taylor series, where the term $\propto \cos(4t)$ describes the lowest-order correction that preserves the symmetry of the ellipse with respect to reflection in the two coordinate axes. The modified curve is then described by

$$\boldsymbol{\theta}(t) = \left(1 + \frac{a_4 \cos(4t)}{r(t)}\right) \begin{pmatrix} a \cos t \\ b \sin t \end{pmatrix} , \tag{3.9}$$

with $r(t)$ as defined above. The parameter a_4 thus describes a deviation from an ellipse: if $a_4 > 0$, the isophote appears more disk-like, and if $a_4 < 0$, it becomes rather boxy (see Fig. 3.11). In elliptical galaxies we typically find $|a_4/a| \sim 0.01$, thus only a small deviation from the elliptical form.

Correlations of a_4 with Other Properties of Ellipticals. Surprisingly, we find that the parameter a_4/a is strongly correlated with other properties of ellipticals (see Fig. 3.12). The ratio $\left(\frac{v_{\rm rot}}{\sigma_v}\right) \Big/ \left(\frac{v_{\rm rot}}{\sigma_v}\right)_{\rm iso}$ (upper left in Fig. 3.12) is of order unity for disky ellipses ($a_4 > 0$) and, in general, significantly smaller than 1 for boxy ellipticals. From this we conclude that "diskies" are in part rotationally supported, whereas the flattening of "boxies" is mainly caused by the anisotropic distribution of their stellar orbits in velocity space. The mass-to-light ratio is also correlated with a_4: boxies (diskies) have a value of M/L in their core which is larger (smaller) than the mean elliptical of comparable luminosity. A very strong correlation exists between a_4/a and the radio luminosity of ellipticals: while diskies are weak radio emitters, boxies show a broad distribution in $L_{\rm radio}$. These correlations are also seen in the X-ray luminosity, since diskies are weak X-ray emitters and boxies have a broad distribution in $L_{\rm x}$. This bimodality becomes even more obvious if the radiation contributed by compact sources (e.g., X-ray binary stars) is subtracted from the total X-ray luminosity, thus considering only the diffuse X-ray emission. Ellipticals with a different sign of a_4 also differ in the kinematics of their stars: boxies often have cores spinning against the general direction of rotation (counter-rotating cores), which is rarely observed in diskies.

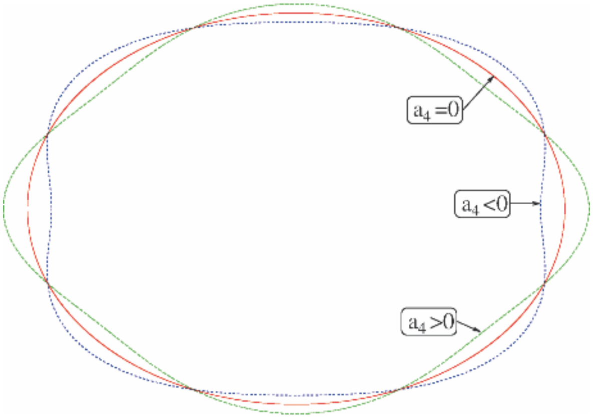

Fig. 3.11. Sketch to illustrate boxiness and diskiness. The solid red curve shows an ellipse ($a_4 = 0$), the green dashed curve a disky ellipse ($a_4 > 0$), and the blue dotted curve a boxy ellipse ($a_4 < 0$). In elliptical galaxies, the deviations in the shape of the isophotes from an ellipse are considerably smaller than in this sketch

About 70% of the ellipticals are diskies. The transition between diskies and S0 galaxies may be continuous along a sequence of varying disk-to-bulge ratio.

Shells and Ripples. In about 40% of the early-type galaxies that are not member galaxies of a cluster, sharp discontinuities in the surface brightness are found,

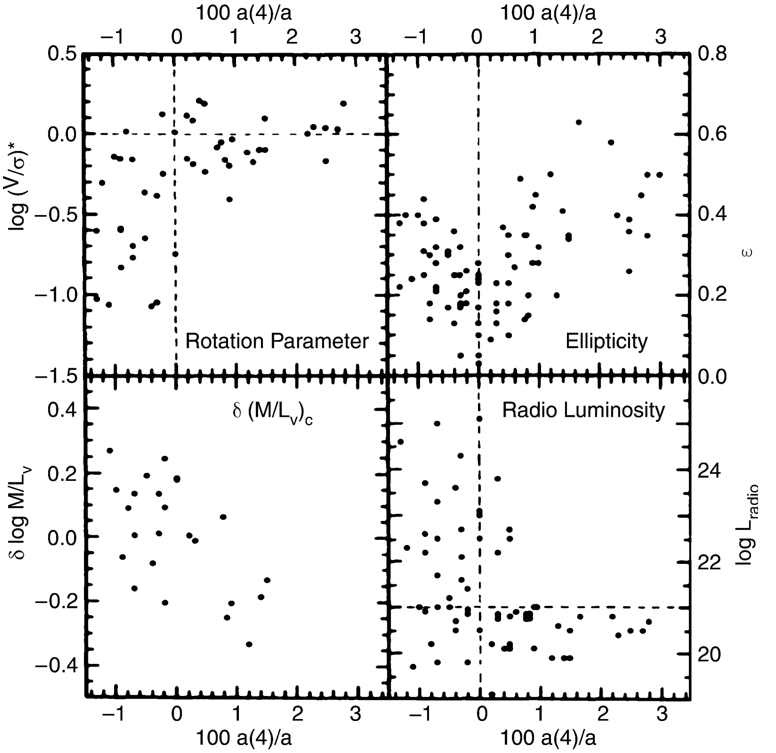

Fig. 3.12. Correlations of a_4/a with some other properties of elliptical galaxies. $100a(4)/a$ (corresponding to a_4/a) describes the deviation of the isophote shape from an ellipse in percent. Negative values denote boxy ellipticals, positive values disky ellipticals. The upper left panel shows the rotation parameter discussed in Sect. 3.2.4; at the lower left, the deviation from the average mass-to-light ratio is shown. The upper right panel shows the ellipticity, and the lower right panel displays the radio luminosity at 1.4 GHz. Obviously, there is a correlation of all these parameters with the boxiness parameter

a kind of shell structure ("shells" or "ripples"). They are visible as elliptical arcs curving around the center of the galaxy (see Fig. 3.13). Such sharp edges can only be formed if the corresponding distribution of stars is "cold", i.e., they must have a very small velocity dispersion, since otherwise such coherent structures would smear out on a very short time-scale. As a comparison, we can consider disk galaxies that likewise contain sharp structures, namely the thin stellar disk. Indeed, the stars in the disk have a very small velocity dispersion, ~ 20 km/s, compared to the rotational velocity of typically 200 km/s.

These peculiarities in ellipticals are not uncommon. Indicators for shells can be found in about half of the early-type galaxies, and about a third of them show boxy isophotes.

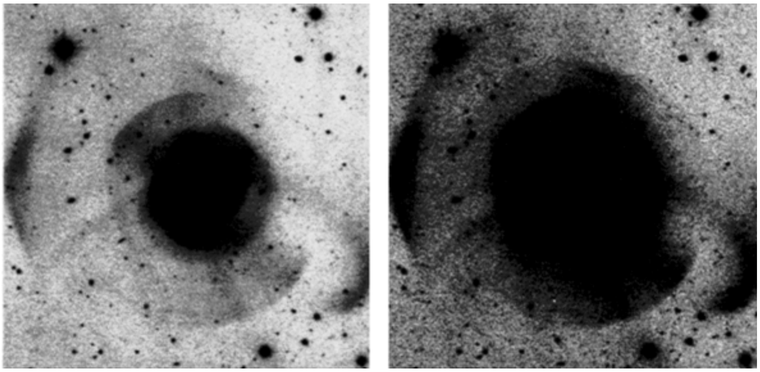

Fig. 3.13. In the galaxy NGC 474, here shown in two images of different contrast, a number of sharp-edged elliptical arcs are visible around the center of the galaxy, the so-called ripples or shells. The displayed image corresponds to a linear scale of about 90 kpc

> Boxiness, counter-rotating cores, and shells and ripples are all indicators of a complex evolution that is probably caused by past mergers with other galaxies.

We will proceed with a discussion of this interpretation in Chap. 9.

3.3 Spiral Galaxies

3.3.1 Trends in the Sequence of Spirals

Looking at the sequence of early-type spirals (i.e., Sa's or SBa's) to late-type spirals, we find a number of differences that can be used for classification (see also Fig. 3.14):

- a decreasing luminosity ratio of bulge and disk, with $L_{\rm bulge}/L_{\rm disk} \sim 0.3$ for Sa's and ~ 0.05 for Sc's;
- an increasing opening angle of the spiral arms, from $\sim 6°$ for Sa's to $\sim 18°$ for Sc's;
- and an increasing brightness structure along the spiral arms: Sa's have a "smooth" distribution of stars along the spiral arms, whereas the light distribution in the spiral arms of Sc's is resolved into bright knots of stars and HII regions.

Compared to ellipticals, the spirals cover a distinctly smaller range in absolute magnitude (and mass). They are limited to $-16 \gtrsim M_B \gtrsim -23$ and $10^9 M_\odot \lesssim M \lesssim 10^{12} M_\odot$, respectively. Characteristic parameters of the various types of spirals are compiled in Table 3.2.

Bars are common in spiral galaxies, with $\sim 70\%$ of all disk galaxies containing a large-scale stellar bar. Such a bar perturbs the axial symmetry of the gravitational potential in a galaxy, which may have a number of consequences. One of them is that this perturbation can lead to a redistribution of angular momentum of the stars, gas, and dark matter. In addition, by perturbing the orbits, gas can be driven towards the center of the galaxy which may have important consequences for triggering nuclear activity (see Chap. 5).

3.3.2 Brightness Profile

The light profile of the bulge of spirals is described by a de Vaucouleurs profile to a good approximation – see (2.39) and (2.41) – while the disk follows an exponential brightness profile, as is the case for our Milky Way. Expressing these distributions of the surface brightness in $\mu \propto -2.5 \log(I)$, measured in mag/arcsec^2, we obtain

$$\boxed{\mu_{\rm bulge}(R) = \mu_{\rm e} + 8.3268 \left[\left(\frac{R}{R_{\rm e}} \right)^{1/4} - 1 \right]} \qquad (3.10)$$

and

$$\boxed{\mu_{\rm disk}(R) = \mu_0 + 1.09 \left(\frac{R}{h_r} \right)} \, . \qquad (3.11)$$

Table 3.2. Characteristic values for spiral galaxies. $V_{\rm max}$ is the maximum rotation velocity, thus characterizing the flat part of the rotation curve. The opening angle is the angle under which the spiral arms branch off, i.e., the angle between the tangent to the spiral arms and the circle around the center of the galaxy running through this tangential point. $S_{\rm N}$ is the specific abundance of globular clusters as defined in (3.13). The values in this table are taken from the book by Carroll & Ostlie (1996)

	Sa	Sb	Sc	Sd/Sm	Im/Ir
M_B	-17 to -23	-17 to -23	-16 to -22	-15 to -20	-13 to -18
M $(M_\odot)$	10^9–10^{12}	10^9–10^{12}	10^9–10^{12}	10^8–10^{10}	10^8–10^{10}
$\langle L_{\rm bulge}/L_{\rm tot}\rangle_B$	0.3	0.13	0.05	–	–
Diam. (D_{25}, kpc)	5–100	5–100	5–100	0.5–50	0.5–50
$\langle M/L_{\rm B}\rangle$ $(M_\odot/L_\odot)$	6.2 ± 0.6	4.5 ± 0.4	2.6 ± 0.2	~ 1	~ 1
$\langle V_{\rm max}\rangle$(km s^{-1})	299	222	175	–	–
$V_{\rm max}$ range (km s^{-1})	163–367	144–330	99–304	–	50–70
Opening angle	$\sim 6°$	$\sim 12°$	$\sim 18°$	–	–
$\mu_{0,\rm B}$ (mag arcsec^{-2})	21.52 ± 0.39	21.52 ± 0.39	21.52 ± 0.39	22.61 ± 0.47	22.61 ± 0.47
$\langle B-V\rangle$	0.75	0.64	0.52	0.47	0.37
$\langle M_{\rm gas}/M_{\rm tot}\rangle$	0.04	0.08	0.16	0.25 (Scd)	–
$\langle M_{\rm H_2}/M_{\rm HI}\rangle$	2.2 ± 0.6 (Sab)	1.8 ± 0.3	0.73 ± 0.13	0.19 ± 0.10	–
$\langle S_{\rm N}\rangle$	1.2 ± 0.2	1.2 ± 0.2	0.5 ± 0.2	0.5 ± 0.2	–

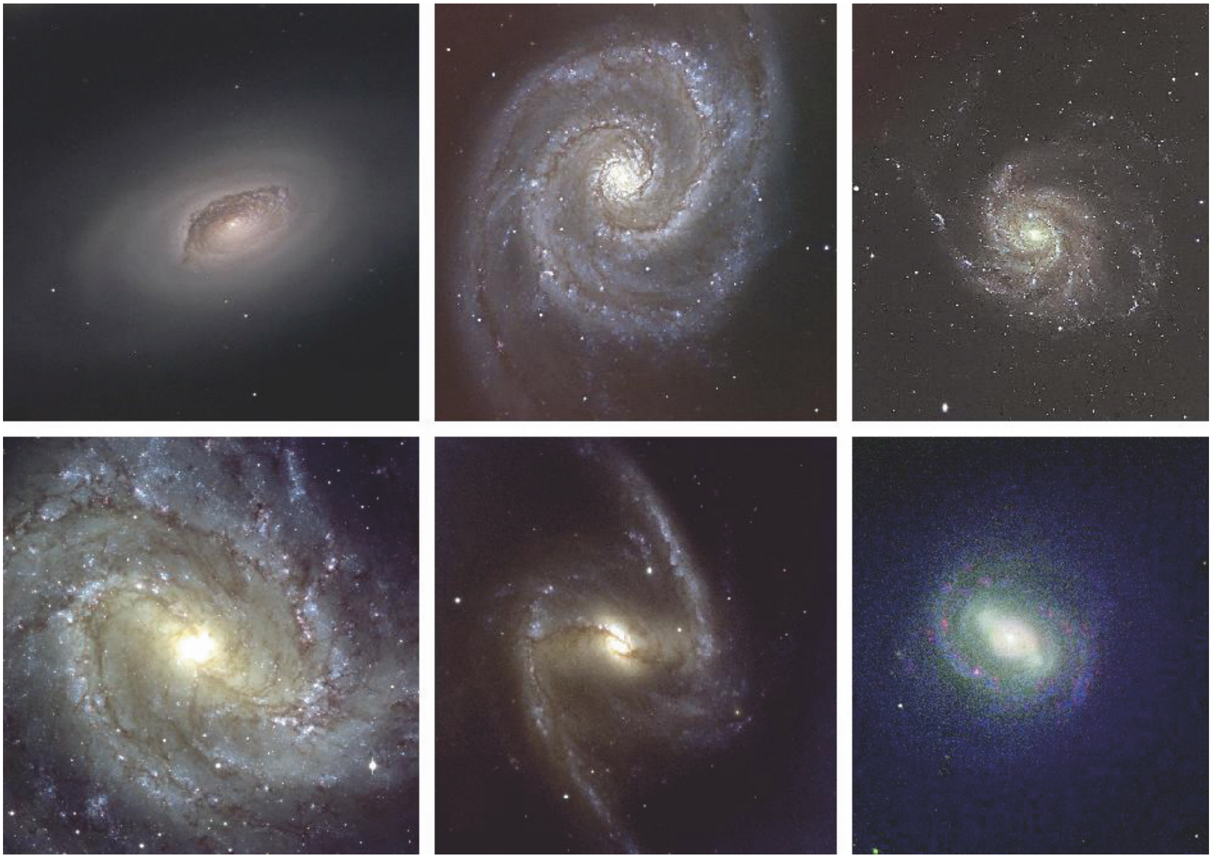

Fig. 3.14. Types of spiral galaxies. Top left: M94, an Sab galaxy. Top middle: M51, an Sbc galaxy. Top right: M101, an Sc galaxy. Lower left: M83, an SBa galaxy. Lower middle: NGC 1365, an SBb galaxy. Lower right: M58, an SBc galaxy

Here, μ_e is the surface brightness at the effective radius R_e which is defined such that half of the luminosity is emitted within R_e (see (2.40)). The central surface brightness and the scale-length of the disk are denoted by μ_0 and h_r, respectively. It has to be noted that μ_0 is not directly measurable since μ_0 is *not* the central surface brightness of the galaxy, only that of its disk component. To determine μ_0, the exponential law (3.11) is extrapolated from large R inwards to $R = 0$.

When Ken Freeman analyzed a sample of spiral galaxies, he found the remarkable result that the central surface brightness μ_0 of disks has a very low spread, i.e., it is very similar for different galaxies (*Freeman's law, 1970*). For Sa's to Sc's, a value of $\mu_0 = 21.52 \pm 0.39$ B-mag/arcsec^2 is observed, and for Sd spirals and later types, $\mu_0 = 22.61 \pm 0.47$ B-mag/arcsec^2. This result was critically discussed, for example with regard to its possible dependence on selection effects. Their importance is not implausible since the determination of precise photometry of galaxies is definitely a lot easier for objects with a high surface brightness. After accounting for such selection effects in the statistical analysis of galaxy samples, Freeman's law was confirmed for "normal" spiral galaxies. However, galaxies exist which have a significantly lower surface brightness, the *low surface brightness galaxies (LSBs)*. They seem to form a separate class of galaxies whose study is substantially more difficult compared to normal spirals because of their low surface brightness. In fact, the central surface brightness of LSBs is much lower than the brightness of the night sky, so that searching for these LSBs is problematic and requires very accurate data reduction and subtraction of the sky background.

Whereas the bulge and the disk can be studied in spirals even at fairly large distances, the stellar halo has too low a surface brightness to be seen in distant galaxies. However, our neighboring galaxy M31, the Andromeda galaxy, can be studied in quite some detail. In particular, the brightness profile of its stellar halo can be studied more easily than that of the Milky Way, taking advantage of our "outside" view. This galaxy should be quite similar to our Galaxy in many respects; for example, tidal streams from disrupted accreted galaxies were also clearly detected in M31.

A stellar halo of red giant branch stars was detected in M31, which extends out to more than 150 kpc from its center. The brightness profile of this stellar distribution indicates that for radii $r \lesssim 20$ kpc it follows the extrapolation from the brightness profile of the bulge, i.e., a de Vaucouleurs profile. However, for larger radii it exceeds this extrapolation, showing a power-law profile which corresponds to a radial density profile of approximately $\rho \propto r^{-3}$, not unlike that observed in our Milky Way. It thus seems that stellar halos form a generic property of spirals. Unfortunately, the corresponding surface brightness is so small that there is little hope of detecting such a halo in other spirals for which individual stars can no longer be resolved and classified.

The thick disk in other spirals can only be studied if they are oriented edge-on. In these cases, a thick disk can indeed be observed as a stellar population outside the plane of the disk and well beyond the scale-height of the thin disk. As is the case for the Milky Way, the scale-height of a stellar population increases with its age, increasing from young main-sequence stars to old asymptotic giant branch stars. For luminous disk galaxies, the thick disk does not contribute substantially to the total luminosity; however, in lower-mass disk galaxies with rotational velocities $\lesssim 120\,\mathrm{km/s}$, the thick disk stars can contribute nearly half the luminosity and may actually dominate the stellar mass. In this case, the dominant stellar population of these galaxies is old, despite the fact that they appear blue.

3.3.3 Rotation Curves and Dark Matter

The rotation curves of other spiral galaxies are easier to measure than that of the Milky Way because we are able to observe them "from outside". These measurements are achieved by utilizing the Doppler effect, where the inclination of the disk, i.e., its orientation with respect to the line-of-sight, has to be accounted for. The inclination angle is determined from the observed axis ratio of the disk, assuming that disks are intrinsically axially symmetric (except for the spiral arms). Mainly the stars and HI gas in the galaxies are used as luminous tracers, where the observable HI disk is in general significantly more extended than the stellar disk. Therefore, the rotation curves measured from the 21-cm line typically extend to much larger radii than those from optical stellar spectroscopy.

Like our Milky Way, other spirals also rotate considerably faster in their outer regions than one would expect from Kepler's law and the distribution of visible matter (see Fig. 3.15).

> The rotation curves of spirals do not decrease for $R \geq h_r$, as one would expect from the light distribution, but are basically flat. We therefore conclude that spirals are surrounded by a halo of dark matter. The density distribution of this dark halo can be derived from the rotation curves.

Indeed, the density distribution of the dark matter can be derived from the rotation curves. The force balance between gravitation and centrifugal acceleration yields the Kepler rotation law,

$$v^2(R) = GM(R)/R \,,$$

from which one directly obtains the mass $M(R)$ within a radius R. The rotation curve expected from the visible matter distribution is[4]

$$v^2_{\mathrm{lum}}(R) = GM_{\mathrm{lum}}(R)/R \,.$$

$M_{\mathrm{lum}}(R)$ can be determined by assuming a constant, plausible value for the mass-to-light ratio of the luminous matter. This value is obtained either from the spectral light distribution of the stars, together with knowledge of the properties of stellar populations, or by fitting the innermost part of the rotation curve (where

[4] This consideration is strongly simplified insofar as the given relations are only valid in this form for spherical mass distributions. The rotational velocity produced by an oblate (disk-shaped) mass distribution is more complicated to calculate; for instance, for an exponential mass distribution in a disk, the maximum of v_{lum} occurs at $\sim 2.2 h_r$, with a Kepler decrease, $v_{\mathrm{lum}} \propto R^{-1/2}$, at larger radii.

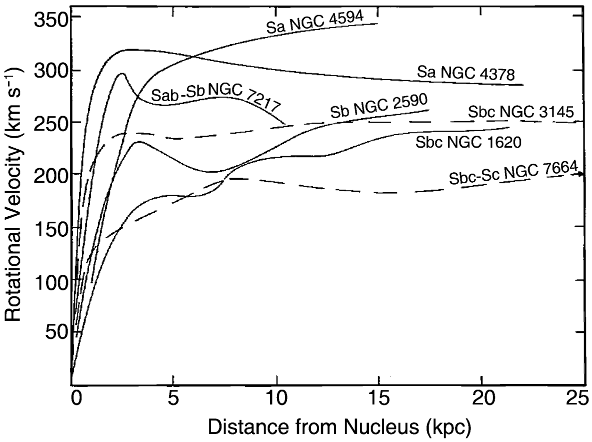

Fig. 3.15. Examples of rotation curves of spiral galaxies. They are all flat in the outer region and do not behave as expected from Kepler's law if the galaxy consisted only of luminous matter. Also striking is the fact that the amplitude of the rotation curve is higher for early types than for late types.

the mass contribution of dark matter can presumably be neglected), assuming that M/L is independent of radius for the stellar population. From this estimate of the mass-to-light ratio, the discrepancy between $v_{\rm lum}^2$ and v^2 yields the distribution of the dark matter, $v_{\rm dark}^2 = v^2 - v_{\rm lum}^2 = GM_{\rm dark}/R$, or

$$\boxed{M_{\rm dark}(R) = \frac{R}{G}\left[v^2(R) - v_{\rm lum}^2(R)\right]} \, . \qquad (3.12)$$

An example of this decomposition of the mass contributions is shown in Fig. 3.16.

The corresponding density profiles of the dark matter halos seem to be flat in the inner region, and decreasing as R^{-2} at large radii. It is remarkable that $\rho \propto R^{-2}$ implies a mass profile $M \propto R$, i.e., the mass of the halo increases linearly with the radius for large R. As long as the extent of the halo is undetermined the total mass of a galaxy will be unknown. Since the observed rotation curves are flat out to the largest radius for which 21-cm emission can still be observed, a lower limit for the radius of the dark halo can be obtained, $R_{\rm halo} \gtrsim 30h^{-1}$ kpc.

To derive the density profile out to even larger radii, other observable objects in an orbit around the galaxies are needed. Potential candidates for such luminous tracers are satellite galaxies – companions of other spirals, like the Magellanic Clouds are for the Milky Way. Because we cannot presume that these satellite galaxies move on circular orbits around their parent galaxy, conclusions can be drawn based only on a statistical sample of satellites. These analyses of the relative velocities of satellite galaxies around spirals still give no indication of an "edge" to the halo, leading to a lower limit for the radius of $R_{\rm halo} \gtrsim 100\,h^{-1}$ kpc.

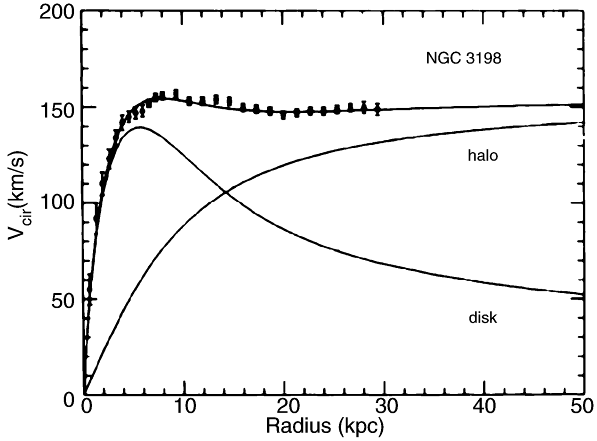

Fig. 3.16. The flat rotation curves of spiral galaxies cannot be explained by visible matter alone. The example of NGC 3198 demonstrates the rotation curve which would be expected from the visible matter alone (curve labeled "disk"). To explain the observed rotation curve, a dark matter component has to be present (curve labeled "halo"). However, the decomposition into disk and halo mass is not unambiguous because for it to be so it would be necessary to know the mass-to-light ratio of the disk. In the case considered here, a "maximum disk" was assumed, i.e., it was assumed that the innermost part of the rotation curve is produced solely by the visible matter in the disk

For elliptical galaxies the mass estimate, and thus the detection of a possible dark matter component, is significantly more complicated, since the orbits of stars are substantially more complex than in spirals. In particular, the mass estimate from measuring the stellar velocity dispersion via line widths depends on the anisotropy of the stellar orbits, which is a priori unknown. Nevertheless, in recent years it has been unambiguously proven that dark matter also exists in ellipticals. First, the degeneracy between the anisotropy of the orbits and the mass determination was broken by detailed kinematic analysis. Second, in some ellipticals hot gas was detected from its X-ray emission. As we will see in Sect. 6.3 in the context of clusters of galaxies, the temperature of the gas allows an estimate of the depth of the potential well, and therefore the mass. Both methods reveal that ellipticals are also surrounded by a dark halo.

The weak gravitational lens effect, which we will discuss in Sect. 6.5.2 in a different context, offers another way to determine the masses of galaxies up to very large radii. With this method we cannot study individual galaxies but only the mean mass properties of a galaxy population. The results of these measurements confirm the large size of dark halos in spirals and in ellipticals.

Correlations of Rotation Curves with Galaxy Properties. The form and amplitude of the rotation curves of spirals are correlated with their luminosity and their Hubble type. The larger the luminosity of a spiral, the steeper the rise of $v(R)$ in the central region, and the larger the maximum rotation velocity $v_{\max}$. This latter fact indicates that the mass of a galaxy increases with luminosity, as expected. For the characteristic values of the various Hubble types, one finds $v_{\max} \sim 300\,\mathrm{km/s}$ for Sa's, $v_{\max} \sim 175\,\mathrm{km/s}$ for Sc's, whereas Irr's have a much lower $v_{\max} \sim 70\,\mathrm{km/s}$. For equal luminosity, $v_{\max}$ is higher for earlier types of spirals. However, the shape (not the amplitude) of the rotation curves of different Hubble types is similar, despite the fact that they have a different brightness profile as seen, for instance, from the varying bulge-to-disk ratio. This point is another indicator that the rotation curves cannot be explained by visible matter alone.

These results leave us with a number of obvious questions. What is the nature of the dark matter? What are the density profiles of dark halos, how are they determined, and where is the "boundary" of a halo? Does the fact that galaxies with $v_{\rm rot} \lesssim 100\,\mathrm{km/s}$ have no prominent spiral structure mean that a minimum halo mass needs to be exceeded in order for spiral arms to form?

Some of these questions will be examined later, but here we point out that the major fraction of the mass of (spiral) galaxies consists of non-luminous matter. The fact that we do not know what this matter consists of leaves us with the question of whether this invisible matter is a new, yet unknown, form of matter. Or is the dark matter less exotic, normal (baryonic) matter that is just not luminous for some reason (for example, because it did not form any stars)? We will see in Chap. 4 that the problem of dark matter is not limited to galaxies, but is also clearly present on a cosmological scale; furthermore, the dark matter cannot be baryonic. A currently unknown form of matter is, therefore, revealing itself in the rotation curves of spirals.

3.3.4 Stellar Populations and Gas Fraction

The color of spiral galaxies depends on their Hubble type, with later types being bluer; e.g., one finds $B-V \sim 0.75$ for Sa's, 0.64 for Sb's, 0.52 for Sc's, and 0.4 for Irr's. This means that the fraction of massive young stars increases along the Hubble sequence towards later spiral types. This conclusion is also in agreement with the findings for the light distribution along spiral arms where we clearly observe active star-formation regions in the bright knots in the spiral arms of Sc's. Furthermore, this color sequence is also in agreement with the decreasing bulge fraction towards later types.

The formation of stars requires gas, and the mass fraction of gas is larger for later types, as can be measured, for instance, from the 21-cm emission of HI, from Hα and from CO emission. Characteristic values for the ratio $\langle M_{\rm gas}/M_{\rm tot}\rangle$ are about 0.04 for Sa's, 0.08 for Sb's, 0.16 for Sc's, and 0.25 for Irr's. In addition, the fraction of molecular gas relative to the total gas mass is smaller for later Hubble types. The dust mass is less than 1% of the gas mass.

Dust, in combination with hot stars, is the main source of far-infrared (FIR) emission from galaxies. Sc galaxies emit a larger fraction of FIR radiation than Sa's, and barred spirals have stronger FIR emission than normal spirals. The FIR emission arises due to dust being heated by the UV radiation of hot stars and then reradiating this energy in the form of thermal emission.

A prominent color gradient is observed in spirals: they are red in the center and bluer in the outer regions. We can identify at least two reasons for this trend. The first is a metallicity effect, as the metallicity is increasing inwards and metal-rich stars are redder than metal-poor ones, due to their higher opacity. Second, the color gradient can be explained by star formation. Since the gas fraction in the bulge is lower than in the disk, less star formation takes place in the bulge, resulting in a stellar population that is older and redder in general. Furthermore, it is found that the metallicity of spirals increases with luminosity.

Abundance of Globular Clusters. The number of globular clusters is higher in early types and in more luminous galaxies. The *specific abundance* of globular clusters in a galaxy is defined as their number, normalized to a galaxy of absolute magnitude $M_V = -15$. This can be done by scaling the observed number N_t of globular clusters in a galaxy of visual luminosity L_V or absolute magnitude M_V, respectively, to that of a hypothetical galaxy with $M_V = -15$:

$$S_N = N_\mathrm{t}\,\frac{L_{15}}{L_V} = N_\mathrm{t}\,10^{0.4(M_V+15)}\ . \tag{3.13}$$

If the number of globular clusters were proportional to the luminosity (and thus roughly to the stellar mass) of a galaxy, then this would imply $S_N = \mathrm{const}$. However, this is not the case: For Sa's and Sb's we find $S_N \sim 1.2$, whereas $S_N \sim 0.5$ for Sc's. S_N is larger for ellipticals and largest for cD galaxies.

3.3.5 Spiral Structure

The spiral arms are the bluest regions in spirals and they contain young stars and HII regions. For this reason, the brightness contrast of spiral arms increases as the wavelength of the (optical) observation decreases. In particular, the spiral structure is very prominent in a blue filter, as is shown impressively in Fig. 3.17.

Naturally, the question arises as to the nature of the spiral arms. Probably the most obvious answer would be that they are material structures of stars and gas, rotating around the galaxy's center together with the rest of the disk. However, this scenario cannot explain spiral arm structure since, owing to the differential rotation, they would wind up much more tightly than observed within only a few rotation periods.

Rather, it is suspected that spiral arms are a wave structure, the velocity of which does not coincide with the physical velocity of the stars. Spiral arms are quasi-stationary density waves, regions of higher density (possibly 10–20% higher than the local disk environment). If the gas, on its orbit around the center of the galaxy, enters a region of higher density, it is compressed, and this compression of molecular clouds results in an enhanced star-formation rate. This accounts for the blue color of spiral arms. Since low-mass (thus red) stars live longer, the brightness contrast of spiral arms is lower in red light, whereas massive blue stars are born in the spiral arms and soon after explode there as SNe. Indeed, only few blue stars are found outside spiral arms.

In order to better understand density waves we may consider, for example, the waves on the surface of a lake. Peaks at different times consist of different water particles, and the velocity of the waves is by no means the bulk velocity of the water.

3.3.6 Corona in Spirals?

Hot gas resulting from the evolution of supernova remnants may expand out of the disk and thereby be ejected to form a gaseous halo of a spiral galaxy. We might

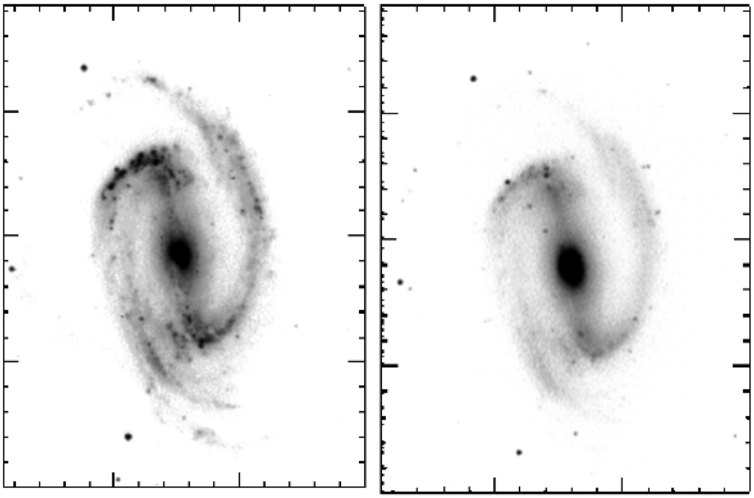

Fig. 3.17. The galaxy NGC 1300 in the B filter (left panel) and in the I filter (right panel). The spiral arms are much more prominent in the blue than in the red. Also, the tips of the bar are more pronounced in the blue – an indicator of an enhanced star-formation rate

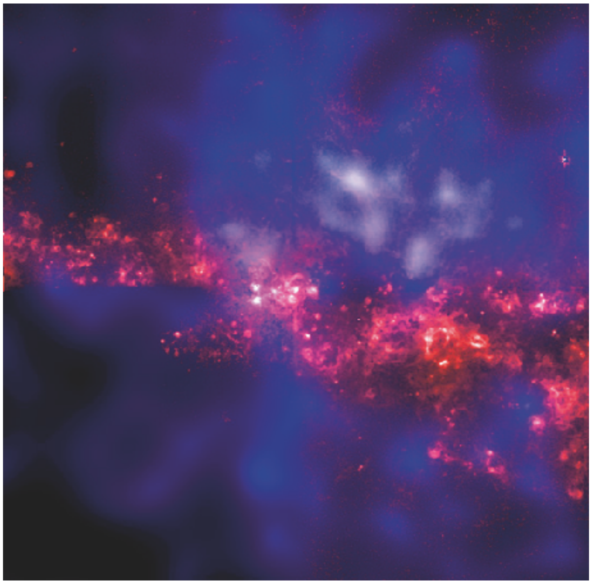

Fig. 3.18. The spiral galaxy NGC 4631. The optical (HST) image of the galaxy is shown in red; the many luminous areas are regions of very active star formation. The SN explosions of massive stars eject hot gas into the halo of the galaxy. This gas (at a temperature of $T \sim 10^6$ K) emits X-ray radiation, shown as the blue diffuse emission as observed by the Chandra satellite. The image has a size of $2\overset{\prime}{.}5$

therefore suspect that such a "coronal" gas exists outside the galactic disk. While the existence of this coronal gas has long been suspected, the detection of its X-ray emission was first made possible with the ROSAT satellite in the early 1990s. However, the limited angular resolution of ROSAT rendered the distinction between diffuse emission and clusters of discrete sources difficult. Finally, the Chandra observatory unambiguously detected the coronal gas in a number of spiral galaxies. As an example, Fig. 3.18 shows the spiral galaxy NGC 4631.

3.4 Scaling Relations

The kinematic properties of spirals and ellipticals are closely related to their luminosity. As we shall discuss below, spirals follow the *Tully–Fisher relation* (Sect. 3.4.1), whereas elliptical galaxies obey the *Faber–Jackson relation* (Sect. 3.4.2) and are located in the *fundamental plane* (Sect. 3.4.3). These scaling relations are a very important tool for distance estimations, as will be discussed in Sect. 3.6. Furthermore, these scaling relations express relations between galaxy properties which any successful model of galaxy evolution must be able to explain. Here we will describe these scaling relations and discuss their physical origin.

3.4.1 The Tully–Fisher Relation

Using 21-cm observations of spiral galaxies, in 1977 R. Brent Tully and J. Richard Fisher found that the maximum rotation velocity of spirals is closely related to their luminosity, following the relation

$$\boxed{L \propto v_{\rm max}^{\alpha}}\,, \tag{3.14}$$

where the slope of the Tully–Fisher relation is about $\alpha \sim 4$. The larger the wavelength of the filter in which the luminosity is measured, the smaller the dispersion of the Tully–Fisher relation (see Fig. 3.19). This is to be expected because radiation at larger wavelengths is less affected by dust absorption and by the current star-formation rate, which may vary to some extent between individual spirals. Furthermore, it is found that the value of α increases with the wavelength of the filter; the Tully–Fisher relation is steeper in the red. The dispersion of galaxies around the relation (3.14) in the near infrared (e.g., in the H-band) is about 10%.

Because of this close correlation, the luminosity of spirals can be estimated quite precisely by measuring the rotational velocity. The determination of the (maximum) rotational velocity is independent of the galaxy's distance. By comparing the luminosity, as determined from the Tully–Fisher relation, with the measured flux one can then estimate the distance of the galaxy – without utilizing the Hubble relation!

The measurement of $v_{\rm max}$ is obtained either from a spatially resolved rotation curve, by measuring $v_{\rm rot}(\theta)$, which is possible for relatively nearby galaxies, or by observing an integrated spectrum of the 21-cm line of HI that has a Doppler width corresponding to about $2v_{\rm max}$ (see Fig. 3.20). The Tully–Fisher relation shown in Fig. 3.19 was determined by measuring the width of the 21-cm line.

Explaining the Tully–Fisher Relation. The shapes of the rotation curves of spirals are very similar to each

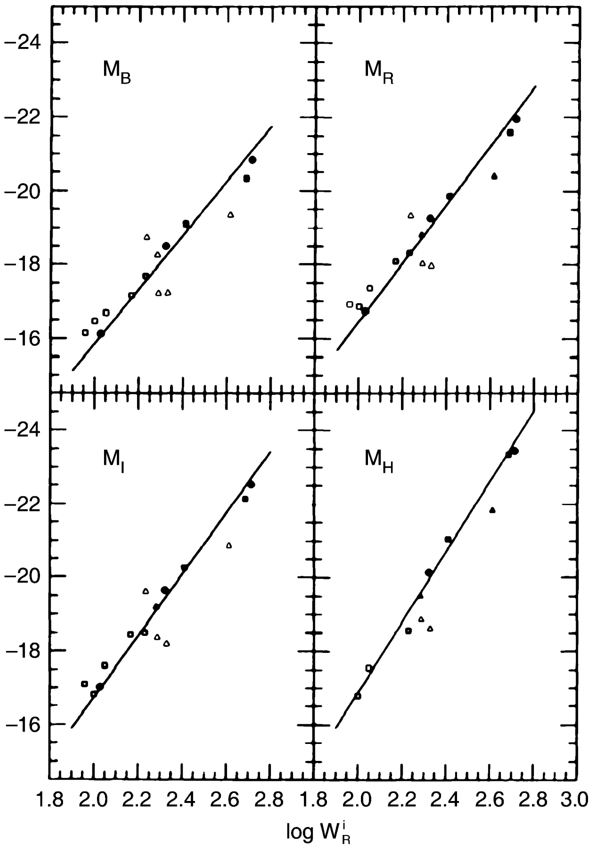

Fig. 3.19. The Tully–Fisher relation for galaxies in the Local Group (dots), in the Sculptor group (triangles), and in the M81 group (squares). The absolute magnitude is plotted as a function of the width of the 21-cm profile which indicates the maximum rotation velocity (see Fig. 3.20). Filled symbols represent galaxies for which independent distance estimates were obtained, either from RR Lyrae stars, Cepheids, or planetary nebulae. For galaxies represented by open symbols, the average distance of the respective group is used. The solid line is a fit to similar data for the Ursa-Major cluster, together with data of those galaxies for which individual distance estimates are available (filled symbols). The larger dispersion around the mean relation for the Sculptor group galaxies is due to the group's extent along the line-of-sight

other, in particular with regard to their flat behavior in the outer part. The flat rotation curve implies

$$\boxed{M = \frac{v_{\max}^2 R}{G}}\ , \tag{3.15}$$

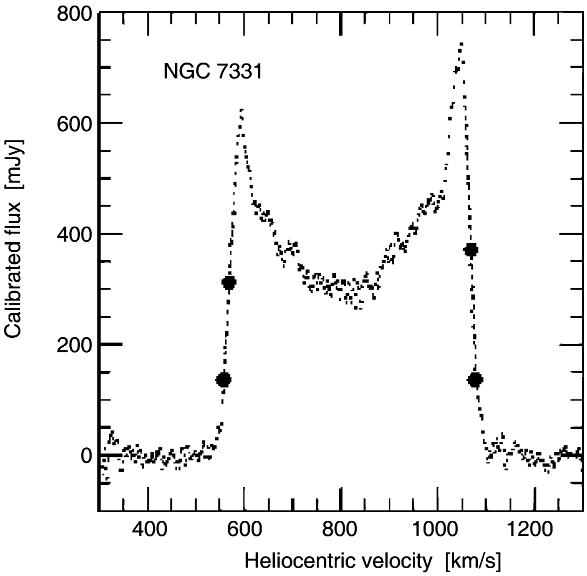

Fig. 3.20. 21 cm profile of the galaxy NGC 7331. The bold dots indicate 20% and 50% of the maximum flux; these are of relevance for the determination of the line width from which the rotational velocity is derived

where the distance R from the center of the galaxy refers to the flat part of the rotation curve. The exact value is not important, though, if only $v(R) \approx$ const. By re-writing (3.15),

$$L = \left(\frac{M}{L}\right)^{-1} \frac{v_{\max}^2 R}{G}\ , \tag{3.16}$$

and replacing R by the mean surface brightness $\langle I \rangle = L/R^2$, we obtain

$$L = \left(\frac{M}{L}\right)^{-2} \left(\frac{1}{G^2 \langle I \rangle}\right) v_{\max}^4\ . \tag{3.17}$$

This is the Tully–Fisher relation *if* M/L and $\langle I \rangle$ are the same for all spirals. The latter is in fact suggested by Freeman's law (Sect. 3.3.2). Since the shapes of rotation curves for spirals seem to be very similar, the radial dependence of the ratio of luminous to dark matter may also be quite similar among spirals. Furthermore, since the red or infrared mass-to-light ratios of a stellar population do not depend strongly on its age, the constancy of M/L could also be valid if dark matter is included.

Although the line of argument presented above is far from a proper derivation of the Tully–Fisher-relation, it nevertheless makes the existence of such a scaling relation plausible.

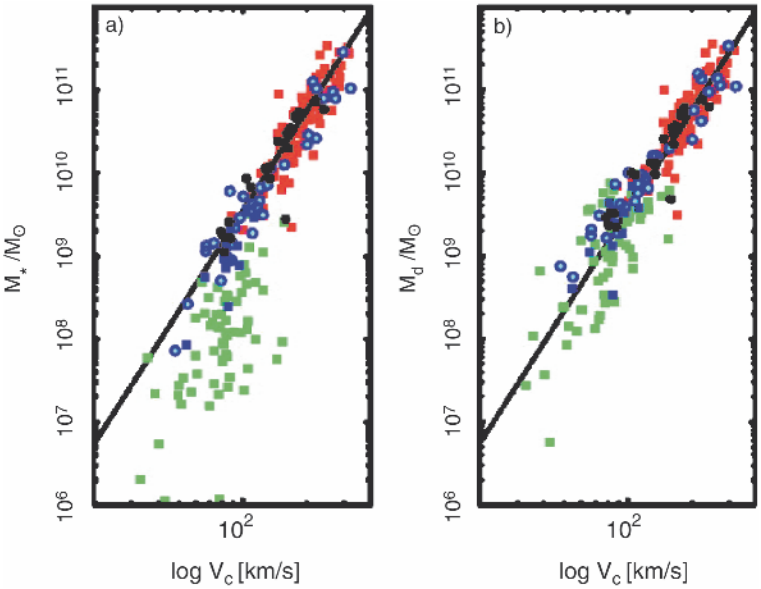

Fig. 3.21. Left panel: the mass contained in stars as a function of the rotational velocity V_c for spirals. This stellar mass is computed from the luminosity by multiplying it with a suitable stellar mass-to-light ratio which depends on the chosen filter and which can be calculated from stellar population models. This is the "classical" Tully–Fisher relation. Squares and circles denote galaxies for which V_c was determined from the 21-cm line width or from a spatially resolved rotation curve, respectively. The colors of the symbols indicate the filter band in which the luminosity was measured: H (red), K′ (black), I (green), B (blue). Right panel: instead of the stellar mass, here the sum of the stellar and gaseous mass is plotted. The gas mass was derived from the flux in the 21-cm line, $M_{\rm gas} = 1.4 M_{\rm HI}$, corrected for helium and metals. Molecular gas has no significant contribution to the baryonic mass. The line in both plots is the Tully–Fisher relation with a slope of $\alpha = 4$

Mass-to-Light Ratio of Spirals. We are unable to determine the total mass of a spiral because the extent of the dark halo is unknown. Thus we can measure M/L only within a fixed radius. We shall define this radius as R_{25}, the radius at which the surface brightness attains the value of 25 mag/arcsec^2 in the B-band;[5] then spirals follow the relation

$$\boxed{\log\left(\frac{R_{25}}{\rm kpc}\right) = -0.249 M_B - 4.00}\ , \qquad (3.18)$$

independently of their Hubble type. Within R_{25} one finds $M/L_B = 6.2$ for Sa's, 4.5 for Sb's, and 2.6 for Sc's. This trend does not come as a surprise because late types of spirals contain more young, blue and luminous stars.

The Baryonic Tully–Fisher Relation. The above "derivation" of the Tully–Fisher relation is based on the assumption of a constant M/L value, where M is the total mass (i.e., including dark matter). Let us assume that (i) the ratio of baryons to dark matter is constant, and furthermore that (ii) the stellar populations in spirals are similar, so that the ratio of stellar mass to luminosity is a constant. Even under these assumptions we would expect the Tully–Fisher relation to be valid only if the gas does not, or only marginally, contribute to the baryonic mass. However, low-mass spirals contain a significant fraction of gas, so we should expect that the Tully–Fisher relation does not apply to these galaxies. Indeed, it is found that spirals with a small $v_{\rm max} \lesssim 100$ km/s deviate significantly from the Tully–Fisher relation – see Fig. 3.21(a).

Since the luminosity is approximately proportional to the stellar mass, $L \propto M_*$, the Tully–Fisher relation is a relation between $v_{\rm max}$ and M_*. Adding the mass of the

[5] We point out explicitly once more that the surface brightness does not depend on the distance of a source.

gas, which can be determined from the strength of the 21-cm line, to the stellar mass a much tighter correlation is obtained, see Fig. 3.21(b). It reads

$$\boxed{M_{\rm disk} = 2 \times 10^9\, h^{-2}\, M_\odot \left(\frac{v_{\rm max}}{100\,{\rm km/s}}\right)^4} \,, \qquad (3.19)$$

and is valid over five orders of magnitude in disk mass $M_{\rm disk} = M_* + M_{\rm gas}$. If no further baryons exist in spirals (such as, e.g., MACHOs), this close relation means that the ratio of baryons and dark matter in spirals is constant over a very wide mass range.

3.4.2 The Faber–Jackson Relation

A relation for elliptical galaxies, analogous to the Tully–Fisher relation, was found by Sandra Faber and Roger Jackson. They discovered that the velocity dispersion in the center of ellipticals, σ_0, scales with luminosity (see Fig. 3.22),

$$\boxed{L \propto \sigma_0^4} \,,$$

or

$$\boxed{\log(\sigma_0) = -0.1 M_B + \text{const}} \,. \qquad (3.20)$$

"Deriving" the Faber–Jackson scaling relation is possible under the same assumptions as the Tully–Fisher relation. However, the dispersion of ellipticals about this relation is larger than that of spirals about the Tully–Fisher relation.

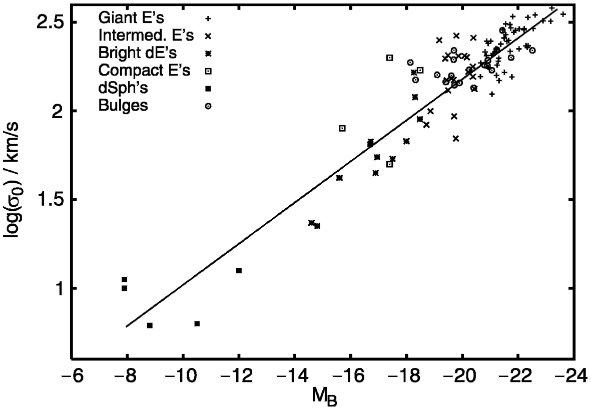

Fig. 3.22. The Faber–Jackson relation expresses a relation between the velocity dispersion and the luminosity of elliptical galaxies. It can be derived from the virial theorem

3.4.3 The Fundamental Plane

The Tully–Fisher and Faber–Jackson relations specify a connection between the luminosity and a kinematic property of galaxies. As we discussed previously, various relations exist between the parameters of elliptical galaxies. Thus one might wonder whether a relation exists between observables of elliptical galaxies for which the dispersion is smaller than that of the Faber–Jackson relation. Such a relation was indeed found and is known as the *fundamental plane*.

To explain this relation, we will consider the various relations between the parameters of ellipticals. In Sect. 3.2.2 we saw that the effective radius of normal ellipticals is related to the luminosity (see Fig. 3.7). This implies a relation between the surface brightness and the effective radius,

$$\boxed{R_{\rm e} \propto \langle I \rangle_{\rm e}^{-0.83}} \,, \qquad (3.21)$$

where $\langle I \rangle_{\rm e}$ is the average surface brightness within the effective radius, so that

$$\boxed{L = 2\pi R_{\rm e}^2 \,\langle I \rangle_{\rm e}} \,. \qquad (3.22)$$

From this, a relation between the luminosity and $\langle I \rangle_{\rm e}$ results,

$$L \propto R_{\rm e}^2 \,\langle I \rangle_{\rm e} \propto \langle I \rangle_{\rm e}^{-0.66}$$

or

$$\boxed{\langle I \rangle_{\rm e} \propto L^{-1.5}} \,. \qquad (3.23)$$

Hence, more luminous ellipticals have smaller surface brightnesses, as is also shown in Fig. 3.7. By means of the Faber–Jackson relation, L is related to σ_0, the central velocity dispersion, and therefore, σ_0, $\langle I \rangle_{\rm e}$, and $R_{\rm e}$ are related to each other. The distribution of elliptical galaxies in the three-dimensional parameter space ($R_{\rm e}$, $\langle I \rangle_{\rm e}$, σ_0) is located close to a plane defined by

$$\boxed{R_{\rm e} \propto \sigma_0^{1.4} \,\langle I \rangle_{\rm e}^{-0.85}} \,. \qquad (3.24)$$

Writing this relation in logarithmic form, we obtain

$$\boxed{\log R_{\rm e} = 0.34 \,\langle \mu \rangle_{\rm e} + 1.4 \log \sigma_0 + \text{const}} \,, \qquad (3.25)$$

where $\langle\mu\rangle_\mathrm{e}$ is the average surface brightness within R_e, measured in mag/arcsec^2. Equation (3.25) defines a plane in this three-dimensional parameter space that is known as the *fundamental plane* (*FP*). Different projections of the fundamental plane are displayed in Fig. 3.23.

How can this be Explained? The mass within R_e can be derived from the virial theorem, $M \propto \sigma_0^2 R_\mathrm{e}$. Combining this with (3.22) yields

$$R_\mathrm{e} \propto \frac{L}{M}\,\frac{\sigma_0^2}{\langle I\rangle_\mathrm{e}}\,, \tag{3.26}$$

which agrees with the FP in the form of (3.24) if

$$\frac{L}{M}\,\frac{\sigma_0^2}{\langle I\rangle_\mathrm{e}} \propto \frac{\sigma_0^{1.4}}{\langle I\rangle_\mathrm{e}^{0.85}}\,,$$

or

$$\frac{M}{L} \propto \frac{\sigma_0^{0.6}}{\langle I\rangle_\mathrm{e}^{0.15}} \propto \frac{M^{0.3}}{R_\mathrm{e}^{0.3}}\,\frac{R_\mathrm{e}^{0.3}}{L^{0.15}}\,.$$

Hence, the FP follows from the virial theorem provided

$$\left(\frac{M}{L}\right) \propto M^{0.2} \qquad \text{or}$$
$$\left(\frac{M}{L}\right) \propto L^{0.25}\,, \qquad \text{respectively}\,, \tag{3.27}$$

i.e., if the mass-to-light ratio of galaxies increases slightly with mass. Like the Tully–Fisher relation, the fundamental plane is an important tool for distance estimations. It will be discussed more thoroughly later.

3.4.4 The D_n–σ Relation

Another scaling relation for ellipticals which is of substantial importance in practical applications is the D_n–σ relation. D_n is defined as that diameter of an ellipse within which the average surface brightness I_n corresponds to a value of 20.75 mag/arcsec^2 in the B-band. If we now assume that all ellipticals have a self-similar brightness profile, $I(R) = I_\mathrm{e}\,f(R/R_\mathrm{e})$, with $f(1) = 1$, then the luminosity within D_n can be written as

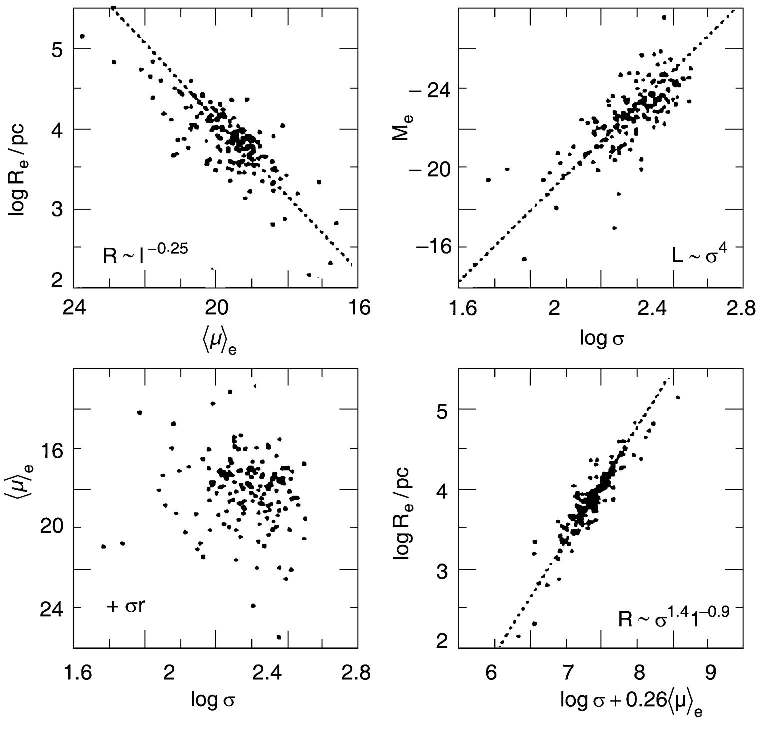

Fig. 3.23. Projections of the fundamental plane onto different two-parameter planes. Upper left: the relation between radius and mean surface brightness within the effective radius. Upper right: Faber–Jackson relation. Lower left: the relation between mean surface brightness and velocity dispersion shows the fundamental plane viewed from above. Lower right: the fundamental plane viewed from the side – the linear relation between radius and a combination of surface brightness and velocity dispersion

$$I_n \left(\frac{D_n}{2}\right)^2 \pi = 2\pi I_e \int_0^{D_n/2} \mathrm{d}R\, R\, f(R/R_e)$$
$$= 2\pi I_e R_e^2 \int_0^{D_n/(2R_e)} \mathrm{d}x\, x\, f(x)\,.$$

For a de Vaucouleurs profile we have approximately $f(x) \propto x^{-1.2}$ in the relevant range of radius. Computing the integral with this expression, we obtain

$$D_n \propto R_e\, I_e^{0.8}\,. \tag{3.28}$$

Replacing R_e by the fundamental plane (3.24) then results in

$$D_n \propto \sigma_0^{1.4}\, \langle I \rangle_e^{-0.85}\, I_e^{0.8}\,.$$

Since $\langle I \rangle_e \propto I_e$ due to the assumed self-similar brightness profile, we finally find

$$\boxed{D_n \propto \sigma_0^{1.4}\, I_e^{0.05}}\,. \tag{3.29}$$

This implies that D_n is nearly independent of I_e and only depends on σ_0. The D_n–σ relation (3.29) describes the properties of ellipticals considerably better than the Faber–Jackson relation and, in contrast to the fundamental plane, it is a relation between only two observables. Empirically, we find that ellipticals follow the normalized D_n–σ relation

$$\frac{D_n}{\mathrm{kpc}} = 2.05 \left(\frac{\sigma_0}{100\,\mathrm{km/s}}\right)^{1.33}\,, \tag{3.30}$$

and they scatter around this relation with a relative width of about 15%.

3.5 Black Holes in the Centers of Galaxies

As we have seen in Sect. 2.6.3, the Milky Way harbors a black hole in its center. Furthermore, it is generally accepted that the energy for the activity of AGNs is generated by accretion onto a black hole (see Sect. 5.3). Thus, the question arises as to whether all (or most) galaxies contain a supermassive black hole (SMBH) in their nuclei. We will pursue this question in this section and show that SMBHs are very abundant indeed. This result then instigates further questions: what distinguishes a "normal" galaxy from an AGN if both have a SMBH in the nucleus? Is it the mass of the black hole, the rate at which material is accreted onto it, or the efficiency of the mechanism which is generating the energy?

We will start with a concise discussion of how to search for SMBHs in galaxies, then present some examples for the discovery of such SMBHs. Finally, we will discuss the very tight relationship between the mass of the SMBH and the properties of the stellar component of a galaxy.

3.5.1 The Search for Supermassive Black Holes

We will start with the question of what a black hole actually is. A technical answer is that a black hole is the simplest solution of Einstein's theory of General Relativity which describes the gravitational field of a point mass. Less technically – though sufficient for our needs – we may say that a black hole is a point mass, or a compact mass concentration, with an extent smaller than its Schwarzschild radius r_S (see below).

The Schwarzschild Radius. The first discussion of black holes can be traced back to Laplace in 1795, who considered the following: if one reduces the radius r of a celestial body of mass M, the escape velocity v_{esc} at its surface will change,

$$v_{esc} = \sqrt{\frac{2GM}{r}}\,.$$

As a thought experiment, one can now see that for a sufficiently small radius v_{esc} will be equal to the speed of light, c. This happens when the radius decreases to

$$\boxed{r_S := \frac{2GM}{c^2} = 2.95 \times 10^5\,\mathrm{cm}\left(\frac{M}{M_\odot}\right)}\,. \tag{3.31}$$

The radius r_S is named the *Schwarzschild radius*, after Karl Schwarzschild who, in 1916, discovered the point-mass solution for Einstein's field equations. For our purpose we will define a black hole as a mass concentration with a radius smaller than r_S. As we can see, r_S is very small: about 3 km for the Sun, and $r_S \sim 10^{12}$ cm for the SMBH in the Galactic center. At a distance of $D = R_0 \approx 8$ kpc, this corresponds to an angular radius of $\sim 6 \times 10^{-6}$ arcsec. Current observing capabilities are still far from resolving scales of order $r_{\rm S}$, but in the near future VLBI observations at very short radio wavelengths may achieve sufficient angular resolution to resolve the Schwarzschild radius for the Galactic black hole. The largest observed velocities of stars in the Galactic center, $\sim 5000\,\mathrm{km/s} \ll c$, indicate that they are still well away from the Schwarzschild radius. However, in the case of the SMBH in our Galactic center we can "look" much closer to the Schwarzschild radius: with VLBI observations at wavelengths of 3 mm the angular size of the compact radio source Sgr A* can be constrained to be less than 0.3 mas, corresponding to about $20 r_{\rm S}$. We will show in Sect. 5.3.3 that relativistic effects are directly observed in AGNs and that velocities close to c do in fact occur there – which again is a very direct indication of the existence of a SMBH.

If even for the closest SMBH, the one in the GC, the Schwarzschild radius is significantly smaller than the achievable angular resolution, how can we hope to prove that SMBHs exist in other galaxies? Like in the GC, this proof has to be found indirectly by detecting a compact mass concentration incompatible with the mass concentration of the stars observed.

The Radius of Influence. We consider a mass concentration of mass $M_\bullet$ in the center of a galaxy where the characteristic velocity dispersion of stars (or gas) is σ. We compare this velocity dispersion with the characteristic velocity (e.g., the Kepler rotational velocity) around a SMBH at a distance r, given by $\sqrt{GM_\bullet/r}$. From this it follows that, for distances smaller than

$$r_{\rm BH} = \frac{GM_\bullet}{\sigma^2} \sim 0.4 \left(\frac{M_\bullet}{10^6 M_\odot}\right) \left(\frac{\sigma}{100\,\mathrm{km/s}}\right)^{-2} \mathrm{pc}\,, \tag{3.32}$$

the SMBH will significantly affect the kinematics of stars and gas in the galaxy. The corresponding angular scale is

$$\begin{aligned}\theta_{\rm BH} &= \frac{r_{\rm BH}}{D} \\ &\sim 0\overset{\prime\prime}{.}1 \left(\frac{M_\bullet}{10^6 M_\odot}\right) \left(\frac{\sigma}{100\,\mathrm{km/s}}\right)^{-2} \left(\frac{D}{1\,\mathrm{Mpc}}\right)^{-1}\,,\end{aligned} \tag{3.33}$$

where D is the distance of the galaxy. From this we immediately conclude that our success in finding SMBHs will depend heavily on the achievable angular resolution. The HST enabled scientists to make huge progress in this field. The search for SMBHs promises to be successful only in relatively nearby galaxies. In addition, from (3.33) we can see that for increasing distance D the mass $M_\bullet$ has to increase for a SMBH to be detectable at a given angular resolution.

Kinematic Evidence. The presence of a SMBH inside $r_{\rm BH}$ is revealed by an increase in the velocity dispersion for $r \lesssim r_{\rm BH}$, which should then behave as $\sigma \propto r^{-1/2}$ for $r \lesssim r_{\rm BH}$. If the inner region of the galaxy rotates, one expects, in addition, that the rotational velocity $v_{\rm rot}$ should also increase inwards $\propto r^{-1/2}$.

Problems in Detecting These Signatures. The practical problems in observing a SMBH have already been mentioned above. One problem is the angular resolution. To measure an increase in the velocities for small radii, the angular resolution needs to be better than $\theta_{\rm BH}$. Furthermore, projection effects play a role because only the velocity dispersion of the projected stellar distribution, weighted by the luminosity of the stars, is measured. Added to this, the kinematics of stars can be rather complicated, so that the observed values for σ and $v_{\rm rot}$ depend on the distribution of orbits and on the geometry of the distribution.

Despite these difficulties, the detection of SMBHs has been achieved in recent years, largely due to the much improved angular resolution of optical telescopes (like the HST) and to improved kinematic models.

3.5.2 Examples for SMBHs in Galaxies

Figure 3.24 shows an example for the kinematical method discussed in the previous section. A long-slit spectrum across the nucleus of the galaxy M84 clearly shows that, near the nucleus, both the rotational velocity and the velocity dispersion change; both increase dramatically towards the center. Figure 3.25 illustrates how strongly the measurability of the kinematical evidence for a SMBH depends on the achievable angular resolution of the observation. For this example of NGC 3115, observing with the resolution offered by space-based spectroscopy yields much higher measured velocities than is possible from the ground. Particularly interesting is the observation of the rotation curve very close to the center. Another impressive example is the central region of M87, the central galaxy of the Virgo Cluster. The increase of the rotation curve and the broadening of the [OII]-line (a spectral line of singly-ionized oxygen) at $\lambda = 3727$ Å towards the center are displayed in Fig. 3.26 and argue very convincingly for a SMBH with $M_\bullet \approx 3 \times 10^9 M_\odot$.

The mapping of the Kepler rotation in the center of the Seyfert galaxy NGC 4258 is especially spectacular. This galaxy contains water masers – very compact sources whose position can be observed with very high precision using VLBI techniques (Fig. 3.27). In this case, the deviation from a Kepler rotation in the gravitational field of a point mass of $M_\bullet \sim 3.5 \times 10^7 M_\odot$ is much less than 1%. The maser sources are embedded in an accretion disk having a thickness of less than 0.3% of its radius, of which also a warping is detected. Changes in the radial velocities and the proper motions of these maser sources have already been measured, so that the model of a Kepler accretion disk has been confirmed in detail.

All these observations are of course no proof of the existence of a SMBH in these galaxies because the sources from which we obtain the kinematic evidence are still too far away from the Schwarzschild radius. The conclusion of the presence of SMBHs is rather that of a missing alternative, as was already explained for the case of the GC (Sect. 2.6.3). We have no other plausible model for the mass concentrations detected. As for the case of the SMBH in the Milky Way, an ultra-compact star cluster might be postulated, but such a cluster would not be stable over a long period of time. Based on the existence of a SMBH in our Galaxy and in AGNs, the SMBH hypothesis is the only plausible explanation for these mass concentrations.

3.5.3 Correlation Between SMBH Mass and Galaxy Properties

Currently, strong indications of SMBHs have been found in about 35 normal galaxies, and their masses have been estimated. This permits us to examine whether, and in what way, $M_\bullet$ is related to the properties of the host galaxy. This leads us to the discovery of a remarkable correlation; it is found that $M_\bullet$ is correlated with the absolute magnitude of the bulge component (or the spheroidal component) of the galaxy in which

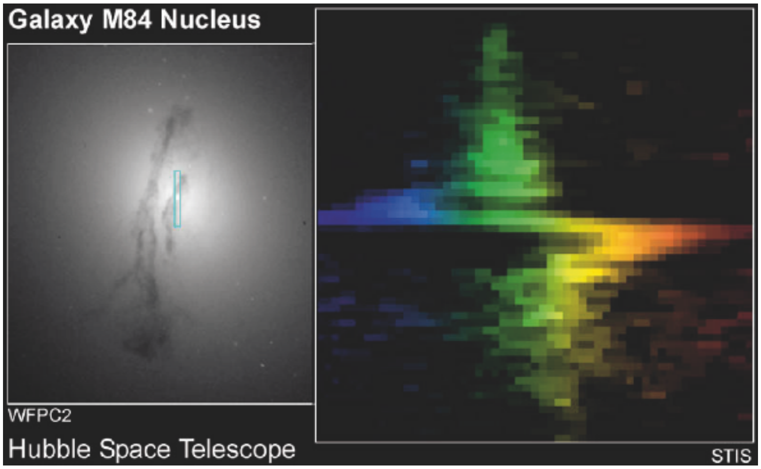

Fig. 3.24. An HST image of the nucleus of the galaxy M84 is shown in the left-hand panel. M84 is a member of the Virgo Cluster, about 15 Mpc away from us. The small rectangle depicts the position of the slit used by the STIS (Space Telescope Imaging Spectrograph) instrument on-board the HST to obtain a spectrum of the central region. This long-slit spectrum is shown in the right-hand panel; the position along the slit is plotted vertically, the wavelength of the light horizontally, also illustrated by colors. Near the center of the galaxy the wavelength suddenly changes because the rotational velocity steeply increases inwards and then changes sign on the other side of the center. This shows the Kepler rotation in the central gravitational field of a SMBH, whose mass can be estimated as $M_\bullet \sim 3 \times 10^8 M_\odot$

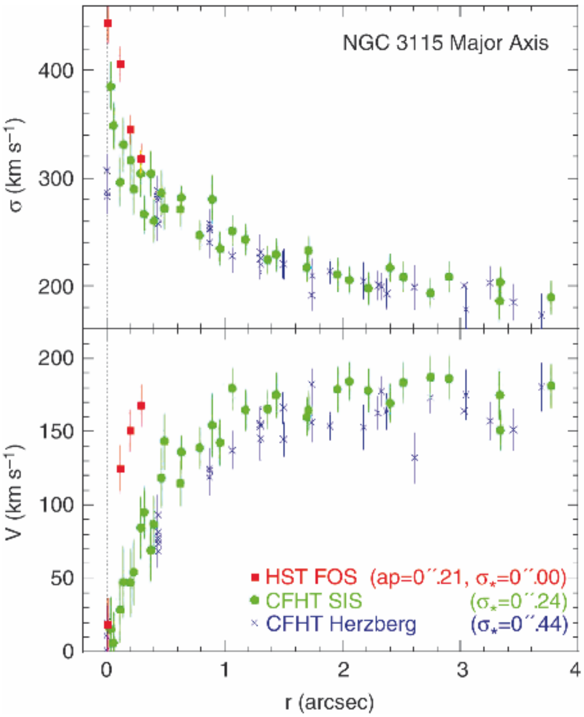

Fig. 3.25. Rotational velocity (bottom) and velocity dispersion (top), as functions of the distance from the center along the major axis of the galaxy NGC 3115. Colors of the symbols mark observations with different instruments. Results from CFHT data which have an angular resolution of 0.″44 are shown in blue. The SIS instrument at the CFHT uses active optics to achieve roughly twice this angular resolution; corresponding results are plotted in green. Finally, the red symbols show the result from HST observations using the Faint Object Spectrograph (FOS). As expected, with improved angular resolution an increase in the velocity dispersion is seen towards the center. Even more dramatic is the impact of resolution on measurements of the rotational velocity. Due to projection effects, the measured central velocity dispersion is smaller than the real one; this effect can be corrected for. After correction, a central value of $\sigma \sim 600\,\mathrm{km/s}$ is found. This value is much higher than the escape velocity from the central star cluster if it were to consist solely of stars – it would dissolve within $\sim 2 \times 10^4$ years. Therefore, an additional compact mass component of $M_\bullet \sim 10^9 M_\odot$ must exist

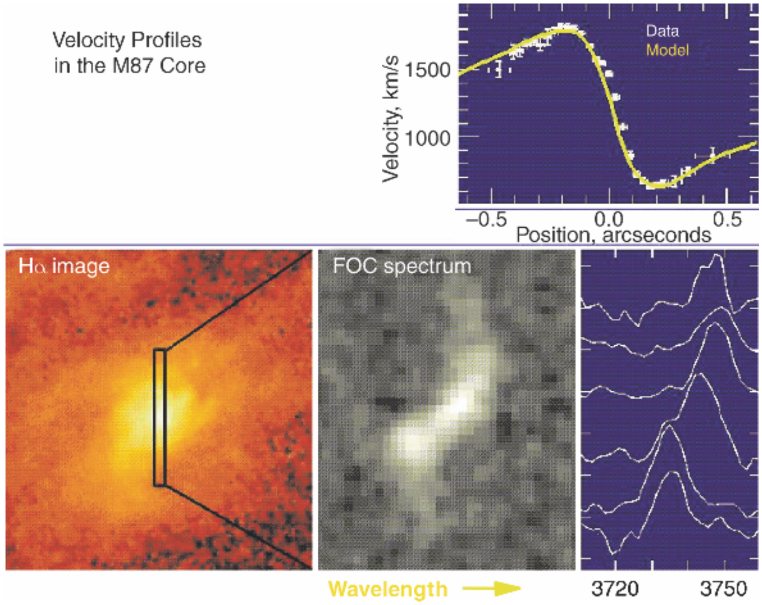

Fig. 3.26. M87 has long been one of the most promising candidates for harboring an SMBH in its center. In this figure, the position of the slit is shown superimposed on an Hα image of the galaxy (lower left) together with the spectrum of the [OII] line along this slit (bottom, center), and six spectra corresponding to six different positions along the slit, separated by 0.″14 each (lower right). In the upper right panel the rotation curve extracted from the data using a kinematical model is displayed. These results show that a central mass concentration with $\sim 3 \times 10^9 M_\odot$ must be present, confined to a region less than 3 pc across – indeed leaving basically no alternative but a SMBH

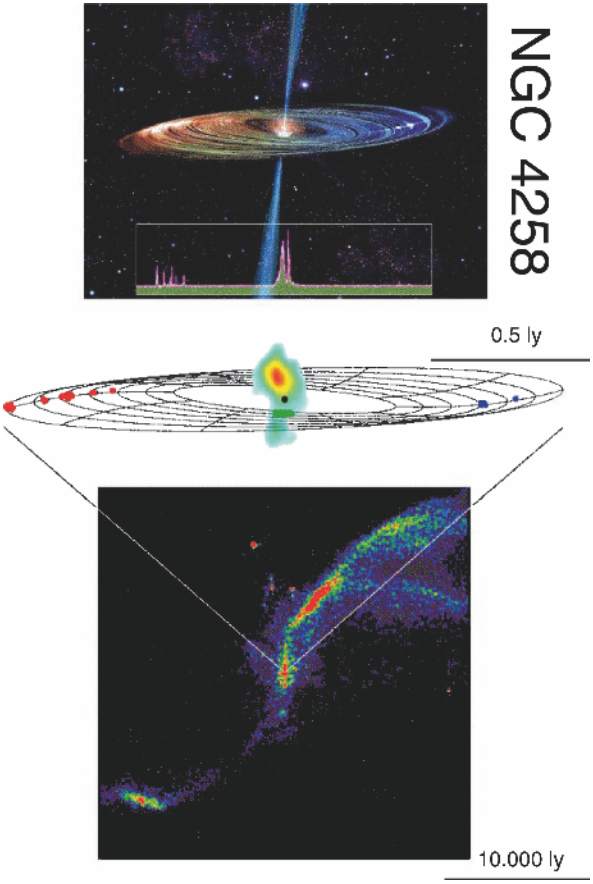

Fig. 3.27. The Seyfert galaxy NGC 4258 contains an accretion disk in its center in which several water masers are embedded. In the top image, an artist's impression of the hidden disk and the jet is displayed, together with the line spectrum of the maser sources. Their positions (center image) and velocities have been mapped by VLBI observations. From these measurements, the Kepler law for rotation in the gravitational field of a point mass of $M_\bullet = 25 \times 10^6 M_\odot$ in the center of this galaxy was verified. The best-fitting model of the central disk is also plotted. The bottom image is a 20-cm map showing the large-scale radio structure of the Seyfert galaxy

the SMBH is located (see Fig. 3.28, left). Here, the bulge component is either the bulge of a spiral galaxy or an elliptical galaxy as a whole. This correlation is described by

$$M_\bullet = 0.93 \times 10^8\, M_\odot \left(\frac{L_{\rm B,bulge}}{10^{10} L_{\rm B\odot}}\right)^{1.11} ; \qquad (3.34)$$

it is statistically highly significant, but the deviations of the data points from this power law are considerably larger than their error bars. An alternative way to express this correlation is provided by the relation $M/L \propto L^{0.25}$ found previously – see (3.27) – by which we can also write $M_\bullet \propto M_{\rm bulge}^{0.9}$.

An even better correlation exists between $M_\bullet$ and the velocity dispersion in the bulge component, as can be seen in the right-hand panel of Fig. 3.28. This relation is best described by

$$\boxed{M_\bullet = 1.35 \times 10^8\, M_\odot \left(\frac{\sigma_{\rm e}}{200\,{\rm km/s}}\right)^4} , \qquad (3.35)$$

where the exact value of the exponent is still subject to discussion, and where a slightly higher value $M_\bullet \propto \sigma^{4.5}$ might better describe the data. The difference in the results obtained by different groups can partially be traced back to different definitions of the velocity dispersion, especially concerning the choice of the spatial region across which it is measured. It is remarkable that the deviations of the data points from the correlation (3.35) are compatible with the error bars for the measurements of $M_\bullet$. Thus, we have at present no indication of an intrinsic dispersion of the $M_\bullet$-σ relation.

In fact, there have been claims in the literature that even globular clusters contain a black hole; however, these claims are not undisputed. In addition, there may be objects that appear like globular clusters, but are in fact the stripped nucleus of a former dwarf galaxy. In this case, the presence of a central black hole is not unexpected, provided the scaling relation (3.35) holds down to very low velocity dispersion.

To date, the physical origin of this very close relation has not been understood in detail. The most obvious apparent explanation – that in the vicinity of a SMBH with a very large mass the stars are moving faster than around a smaller-mass SMBH – is not conclusive: the mass of the SMBH is significantly less than one percent of the mass of the bulge component. We can therefore disregard its contribution to the gravitational field in which the stars are orbiting. Instead, this correlation has to be linked to the fact that the spheroidal component of a galaxy evolves together with the SMBH. A better understanding of this relation can only be found from models of galaxy evolution. We will continue with this topic in Sect. 9.6.

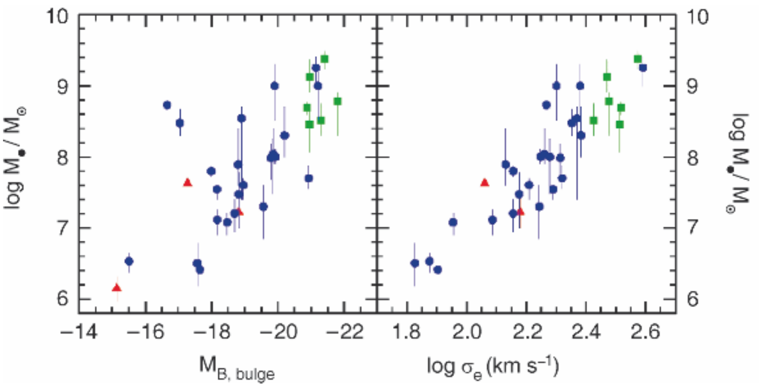

Fig. 3.28. Correlation of SMBH mass $M_\bullet$ with the absolute magnitude $M_{\rm B,bulge}$ (left) and the velocity dispersion $\sigma_{\rm e}$ (right) in the bulge component of the host galaxy. Circles (squares, triangles) indicate measurements that are based on stellar kinematics (gas kinematics, maser disks)

3.6 Extragalactic Distance Determination

In Sect. 2.2 we discussed methods for distance determination within our own Galaxy. We will now proceed with the determination of distances to other galaxies. It should be noted that the Hubble law (1.2) specifies a relation between the redshift of an extragalactic object and its distance. The redshift z is easily measured from the shift in spectral lines. For this reason, the Hubble law (and its generalization – see Sect. 4.3.3) provides a simple method for determining distance. However, to apply this law, the Hubble constant H_0 must first be known, i.e., the Hubble law must be calibrated. Therefore, in order to determine the Hubble constant, distances have to be measured independently from redshift.

Furthermore, it has to be kept in mind that besides the general cosmic expansion, which is expressed in the Hubble law, objects also show *peculiar motion*, like the velocities of galaxies in clusters of galaxies or the motion of the Magellanic Clouds around our Milky Way. These peculiar velocities are induced by gravitational acceleration resulting from the locally inhomogeneous mass distribution in the Universe. For instance, our Galaxy is moving towards the Virgo Cluster of galaxies, a dense accumulation of galaxies, due to the gravitational attraction caused by the cluster mass. The measured redshift, and therefore the Doppler shift, is always a superposition of the cosmic expansion velocity and peculiar velocities.

CMB Dipole Anisotropy. The peculiar velocity of the Galaxy is very precisely known. The radiation of the cosmic microwave background is not completely isotropic but instead shows a dipole component. This component originates in the velocity of the Solar System relative to the rest-frame in which the CMB appears isotropic (see Fig. 1.17). Due to the Doppler effect, the CMB appears hotter than average in the direction of our motion and cooler in the opposite direction. Analyzing this CMB dipole allows us to determine our peculiar velocity, which yields the result that the Sun moves at a velocity of (368 ± 2) km/s relative to the CMB rest-frame. Furthermore, the Local Group of galaxies (see Sect. 6.1) is moving at $v_{\rm LG} \approx 600$ km/s relative to the CMB rest-frame.

Distance Ladder. For the redshift of a source to be dominated by the Hubble expansion, the cosmic expansion velocity $v = cz = H_0 D$ has to be much larger than typical peculiar velocities. This means that in order to determine H_0 we have to consider sources at large distances for the peculiar velocities to be negligible compared to $H_0\,D$.

Making a direct estimate of the distances of distant galaxies is very difficult. Traditionally one uses a *distance ladder*: at first, the *absolute distances* to nearby galaxies are measured directly. If methods to measure *relative distances* (that is, distance ratios) with sufficient precision are utilized, the distances to galaxies further away are then determined relative to those nearby. In this way, by means of relative methods, distances are estimated for galaxies that are sufficiently far away for their redshift to be dominated by the Hubble flow.

3.6.1 Distance of the LMC

The distance of the Large Magellanic Cloud (LMC) can be estimated using various methods. For example, we can resolve and observe individual stars in the LMC, which forms the basis of the MACHO experiments (see Sect. 2.5.2). Because the metallicity of the LMC is significantly lower than that of the Milky Way, some of the methods discussed in Sect. 2.2 are only applicable after correcting for metallicity effects, e.g., the photometric distance determination or the period–luminosity relation for pulsating stars.

Perhaps the most precise method of determining the distance to the LMC is a purely geometrical one. The supernova SN 1987A that exploded in 1987 in the LMC illuminates a nearly perfectly elliptical ring (see Fig. 3.29). This ring consists of material that was once ejected by the stellar winds of the progenitor star of the supernova and that is now radiatively excited by energetic photons from the supernova explosion. The corresponding recombination radiation is thus emitted only when photons from the SN hit this gas. Because the observed ring is almost certainly intrinsically circular and the observed ellipticity is caused only by its inclination with respect to the line-of-sight, the distance to SN 1987A can be derived from observations of the ring. First, the inclination angle is determined from its observed ellipticity. The gas in the ring is excited by photons from the SN a time R/c after the original explosion, where R is the radius of the ring. We do not observe the illumination of the ring instantaneously because light from the section of the ring closer to us reaches us earlier than light from the more distant part. Thus, its illumination was seen sequentially along the ring. Combining the time delay in the illumination between the nearest and farthest part of the ring with its inclination angle, we then obtain the physical diameter of the ring. When this is compared to the measured angular diameter of $\sim 1\overset{''}{.}7$, the ratio yields the distance to SN 1987A,

$$D_{\rm SN1987A} \approx 51.8\,{\rm kpc} \pm 6\% \;.$$

If we now assume the extent of the LMC along the line-of-sight to be small, this distance can be identified with the distance to the LMC. The value is also compatible with other distance estimates (e.g., as derived by using photometric methods based on the properties of main-sequence stars – see Sect. 2.2.4).

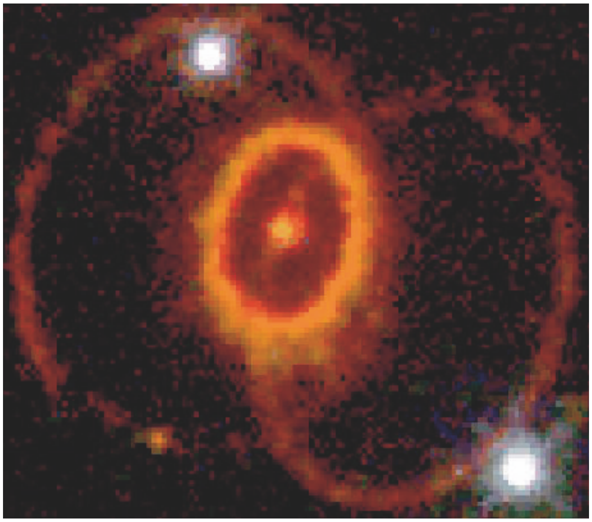

Fig. 3.29. The ring around supernova 1987A in the LMC is illuminated by photons from the explosion which induce the radiation from the gas in the ring. It is inclined towards the line-of-sight; thus it appears to be elliptical. Lighting up of the ring was not instantaneous, due to the finite speed of light: those sections of the ring closer to us lit up earlier than the more distant parts. From the time shift in the onset of radiation across the ring, its diameter can be derived. Combining this with the measured angular diameter of the ring, the distance to SN 1987A – and thus the distance to the LMC – can be determined

3.6.2 The Cepheid Distance

In Sect. 2.2.7, we discussed the period–luminosity relation of pulsating stars. Due to their high luminosity, Cepheids turn out to be particularly useful since they can be observed out to large distances.

For the period–luminosity relation of the Cepheids to be a good distance measure, it must first be calibrated. This calibration has to be done with as large a sample of Cepheids as possible at a known distance. Cepheids in the LMC are well-suited for this purpose because we believe we know the distance to the LMC quite precisely, see above. Also, due to the relatively small extent of the LMC along the line-of-sight, all Cepheids in the

LMC should be located at approximately the same distance. For this reason, the period–luminosity relation is calibrated in the LMC. Due to the large number of Cepheids available for this purpose (many of them have been found in the microlensing surveys), the resulting statistical errors are small. Uncertainties remain in the form of systematic errors related to the metallicity dependence of the period–luminosity relation; however, these can be corrected for since the color of Cepheids depends on the metallicity as well.

With the high angular resolution of the HST, individual Cepheids in galaxies are visible at distances up to that of the Virgo cluster of galaxies. In fact, determining the distance to Virgo as a central step in the determination of the Hubble constant was one of the major scientific aims of the HST. In the *Hubble Key Project*, the distances to numerous spiral galaxies in the Virgo Cluster were determined by identifying Cepheids and measuring their periods.

3.6.3 Secondary Distance Indicators

The Virgo Cluster, at a measured distance of about 16 Mpc, is not sufficiently far away from us to directly determine the Hubble constant from its distance and redshift, because peculiar velocities still contribute considerably to the measured redshift at this distance. To get to larger distances, a number of relative distance indicators are used. They are all based on measuring the distance *ratio* of galaxies. If the distance to one of the two is known, the distance to the other is then obtained from the ratio. By this procedure, distances to more remote galaxies can be measured. Below, we will review some of the most important secondary distance indicators.

SN Ia. Supernovae of Type Ia are to good approximation standard candles, as will be discussed more thoroughly in Sect. 8.3.1. This means that the absolute magnitudes of SNe Ia are all within a very narrow range. To measure the value of this absolute magnitude, distances must be known for galaxies in which SN Ia explosions have been observed and accurately measured. Therefore, the Cepheid method was applied especially to such galaxies, in this way calibrating the brightness of SNe Ia. SNe Ia are visible over very large distances, so that they also permit distance estimates at such large redshifts that the simple Hubble law (1.6) is no longer valid, but needs to be generalized based on a cosmological model (Sect. 4.3.3). As we will see later, these measurements belong to the most important pillars on which our standard model of cosmology rests.

Surface Brightness Fluctuations of Galaxies. Another method of estimating distance ratios is surface brightness fluctuations. It is based on the fact that the number of bright stars per area element in a galaxy fluctuates – purely by Poisson noise: If N stars are expected in an area element, relative fluctuations of $\sqrt{N}/N = 1/\sqrt{N}$ of the number of stars will occur. These are observed in fluctuations of the local surface brightness. To demonstrate that this effect can be used to estimate distances, we consider a solid angle $\mathrm{d}\omega$. The corresponding area element $\mathrm{d}A = D^2\,\mathrm{d}\omega$ depends quadratically on the distance D of the galaxy; the larger the distance, the larger the number of stars N in this solid angle, and the smaller the relative fluctuations of the surface brightness. By comparing the surface brightness fluctuations of different galaxies, one can then estimate relative distances. This method also has to be calibrated on the galaxies for which Cepheid distances are available.

Planetary Nebulae. The brightness distribution of planetary nebulae in a galaxy seems to have an upper limit which is the nearly the same for each galaxy (see Fig. 3.30). If a sufficient number of planetary nebulae are observed and their brightnesses measured, it enables us to determine their luminosity function from which the maximum apparent magnitude is then derived. By calibration on galaxies of known Cepheid distance, the corresponding maximum absolute magnitude can be determined, which then allows the determination of the distance modulus for other galaxies, thus their distances.

Scaling Relations. The scaling relations for galaxies – fundamental plane for ellipticals, Tully–Fisher relation for spirals (see Sect. 3.4) – can be calibrated on local groups of galaxies or on the Virgo Cluster, the distances of which have been determined from Cepheids. Although the scatter of these scaling relations can be 15% for individual galaxies, the statistical fluctuations are reduced when observing several galaxies at about the same distance (such as in clusters and groups). This

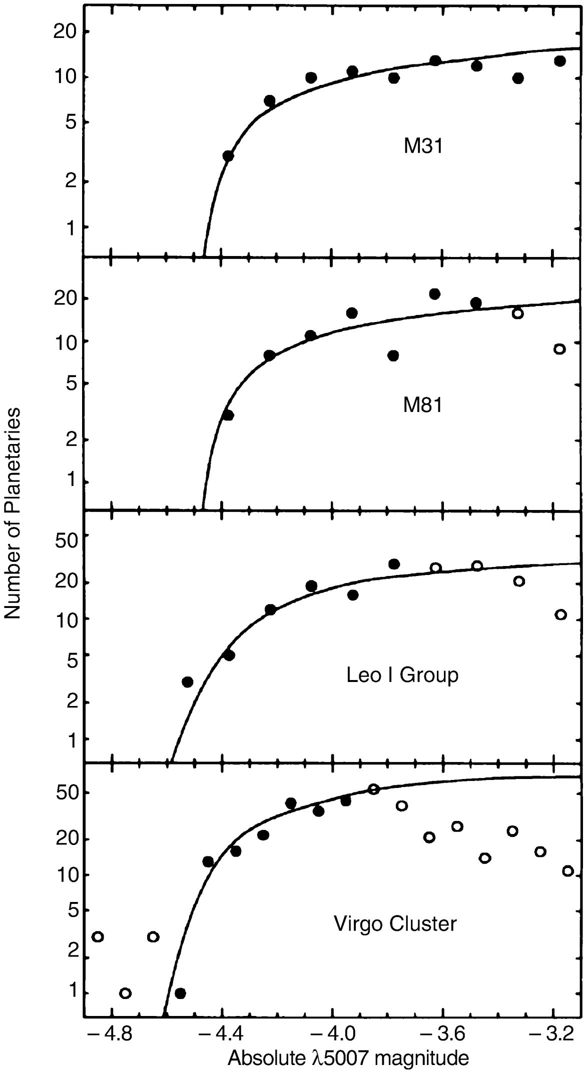

Fig. 3.30. Brightness distribution of planetary nebulae in Andromeda (M31), M81, three galaxies in the Leo I group, and six galaxies in the Virgo Cluster. The plotted absolute magnitude was measured in the emission line of double-ionized oxygen at $\lambda = 5007$ Å in which a large fraction of the luminosity of a planetary nebula is emitted. This characteristic property is also used in the identification of such objects in other galaxies. In all cases, the distribution is described by a nearly identical luminosity function; it seems to be a universal function in galaxies. Therefore, the brightness distribution of planetary nebulae can be used to estimate the distance of a galaxy. In the fits shown, the data points marked by open symbols were disregarded: at these magnitudes, the distribution function is probably not complete

enables us to estimate the distance ratio of two clusters of galaxies.

The Hubble Constant. In particular, the ratio of distances to the Virgo and the Coma clusters of galaxies is estimated by means of these various secondary distance measures. Together with the distance to the Virgo Cluster as determined from Cepheids, we can then derive the distance to Coma. Its redshift ($z \approx 0.023$) is large enough for its peculiar velocity to make no significant contribution to its redshift, so that it is dominated by the Hubble expansion. By combining the various methods we obtain a distance to the Coma cluster of about 90 Mpc, resulting in a Hubble constant of

$$\boxed{H_0 = 72 \pm 8\ \mathrm{km/s/Mpc}}\ . \tag{3.36}$$

The error given here denotes the statistical uncertainty in the determination of H_0. Besides this uncertainty, possible systematic errors of the same order of magnitude may exist. In particular, the distance to the LMC plays a crucial role. As the lowest rung in the distance latter, it has an effect on all further distance estimates. We will see later (Sect. 8.7.1) that the Hubble constant can also be measured by a completely different method, based on tiny small-scale anisotropies of the cosmic microwave background, and that this method results in a value which is in impressively good agreement with the one in (3.36).

3.7 Luminosity Function of Galaxies

Definition of the Luminosity Function. The luminosity function specifies the way in which the members of a class of objects are distributed with respect to their luminosity. More precisely, the luminosity function is the number density of objects (here galaxies) of a specific luminosity. $\Phi(M)\,\mathrm{d}M$ is defined as the number density of galaxies with absolute magnitude in the interval $[M, M + \mathrm{d}M]$. The total density of galaxies is then

$$\nu = \int_{-\infty}^{\infty} \mathrm{d}M\ \Phi(M)\ . \tag{3.37}$$

Accordingly, $\Phi(L)\,\mathrm{d}L$ is defined as the number density of galaxies with a luminosity between L and $L + \mathrm{d}L$. It should be noted here explicitly that both definitions of the luminosity function are denoted by the same symbol, although they represent different mathematical functions, i.e., they describe different functional relations. It is therefore important (and in most cases not difficult) to deduce from the context which of these two functions is being referred to.

Problems in Determining the Luminosity Function. At first sight, the task of determining the luminosity function of galaxies does not seem very difficult. The history of this topic shows, however, that we encounter a number of problems in practice. As a first step, the determination of galaxy luminosities is required, for which, besides measuring the flux, distance estimates are also necessary. For very distant galaxies redshift is a sufficiently reliable measure of distance, whereas for nearby galaxies the methods discussed in Sect. 3.6 have to be applied.

Another problem occurs for nearby galaxies, namely the large-scale structure of the galaxy distribution. To obtain a representative sample of galaxies, a sufficiently large volume has to be surveyed because the galaxy distribution is heavily structured on scales of $\sim 100\,h^{-1}$ Mpc. On the other hand, galaxies of particularly low luminosity can only be observed locally, so the determination of $\Phi(L)$ for small L always needs to refer to local galaxies. Finally, one has to deal with the so-called *Malmquist bias*; in a flux-limited sample luminous galaxies will always be overrepresented because they are visible at larger distances (and therefore are selected from a larger volume). A correction for this effect is always necessary.

3.7.1 The Schechter Luminosity Function

The global galaxy distribution is well approximated by the *Schechter luminosity function*

$$\boxed{\Phi(L) = \left(\frac{\Phi^*}{L^*}\right) \left(\frac{L}{L^*}\right)^{\alpha} \exp\left(-L/L^*\right)} \;, \qquad (3.38)$$

where L^* is a characteristic luminosity above which the distribution decreases exponentially, α is the slope of the luminosity function for small L, and Φ^* specifies the normalization of the distribution. A schematic plot of this function is shown in Fig. 3.31.

Expressed in magnitudes, this function appears much more complicated. Considering that an interval $\mathrm{d}L$ in luminosity corresponds to an interval $\mathrm{d}M$ in absolute magnitude, with $\mathrm{d}L/L = -0.4\,\ln 10\,\mathrm{d}M$, and using $\Phi(L)\,\mathrm{d}L = \Phi(M)\,\mathrm{d}M$, i.e., the number of sources in these intervals are of course the same, we obtain

$$\Phi(M) = \Phi(L)\left|\frac{\mathrm{d}L}{\mathrm{d}M}\right| = \Phi(L)\,0.4\,\ln 10\,L \qquad (3.39)$$

$$= (0.4\,\ln 10)\Phi^* 10^{0.4(\alpha+1)(M^*-M)} \times \exp\left(-10^{0.4(M^*-M)}\right) \;. \qquad (3.40)$$

As mentioned above, the determination of the parameters entering the Schechter function is difficult; a set of parameters in the blue band is

$$\begin{aligned} \Phi^* &= 1.6\times 10^{-2}\,h^3\,\mathrm{Mpc}^{-3}\;, \\ M_B^* &= -19.7 + 5\log h\;, \qquad \text{or} \\ L_B^* &= 1.2\times 10^{10}\,h^{-2}\,L_\odot\;, \qquad (3.41) \\ \alpha &= -1.07\;. \end{aligned}$$

While the blue light of galaxies is strongly affected by star formation, the luminosity function in the red bands measures the typical stellar distribution. In the K-band, we have

$$\begin{aligned} \Phi^* &= 1.6\times 10^{-2}\,h^3\,\mathrm{Mpc}^{-3}\;, \\ M_K^* &= -23.1 + 5\log h\;, \qquad (3.42) \\ \alpha &= -0.9\;. \end{aligned}$$

The total number density of galaxies is formally infinite if $\alpha \le -1$, but the validity of the Schechter function does of course not extend to arbitrarily small L. The luminosity density

$$l_{\mathrm{tot}} = \int_0^\infty \mathrm{d}L\;L\,\Phi(L) = \Phi^*\,L^*\,\Gamma(2+\alpha) \qquad (3.43)$$

is finite for $\alpha \ge -2$.[6] The integral in (3.43), for $\alpha \sim -1$, is dominated by $L \sim L^*$, and $n = \Phi^*$ is thus a good estimate for the mean density of L^*-galaxies.

[6] $\Gamma(x)$ is the Gamma function, defined by

$$\Gamma(x) = \int_0^\infty \mathrm{d}y\;y^{(x-1)}\,\mathrm{e}^{-y}\;. \qquad (3.44)$$

For positive integers, $\Gamma(n+1) = n!$. We have $\Gamma(0.7) \approx 1.30$, $\Gamma(1) = 1$, $\Gamma(1.3) \approx 0.90$. Since these values are all close to unity, $l_{\mathrm{tot}} \sim \Phi^* L^*$ is a good approximation for the luminosity density.

Deviations of the galaxy luminosity function from the Schechter form are common. There is also no obvious reason why such a simple relation for describing the luminosity distribution of galaxies should exist. Although the Schechter function seems to be a good representation of the total distribution, each type of galaxy has its own luminosity function, with each function having a form that strongly deviates from the Schechter function – see Fig. 3.32. For instance, spirals are relatively narrowly distributed in L, whereas the distribution of ellipticals is much broader if we account for the full L-range, from giant ellipticals to dwarf ellipticals. E's dominate in particular at large L; the low end of the luminosity function is likewise dominated by dwarf ellipticals and Irr's. In addition, the luminosity distribution of cluster and group galaxies differs from that of field galaxies. The fact that the total luminosity function can be described by an equation as simple as (3.38) is, at least partly, a coincidence ("cosmic conspiracy") and cannot be modeled easily.

3.7.2 The Bimodal Color Distribution of Galaxies

The classification of galaxies by morphology, given by the Hubble classification scheme (Fig. 3.2), has the disadvantage that morphologies of galaxies are not easy to quantify. Traditionally, this was done by visual inspection but of course this method bears some subjectivity of the researcher doing it. Furthermore, this visual inspection is time consuming and cannot be performed on large samples of galaxies. Various techniques were developed to perform such a classification automatically, including brightness profile fitting – a de Vaucouleurs profile indicates an elliptical galaxy whereas an exponential brightness profile corresponds to a spiral.

Even these methods cannot be applied to galaxy samples for which the angular resolution of the imaging is not much better than the angular size of galaxies – since then, no brightness profiles can be fitted. An alternative to classify galaxies is provided by their color. We expect that early-type galaxies are red, whereas late-type galaxies are considerably bluer. Colors are much eas-

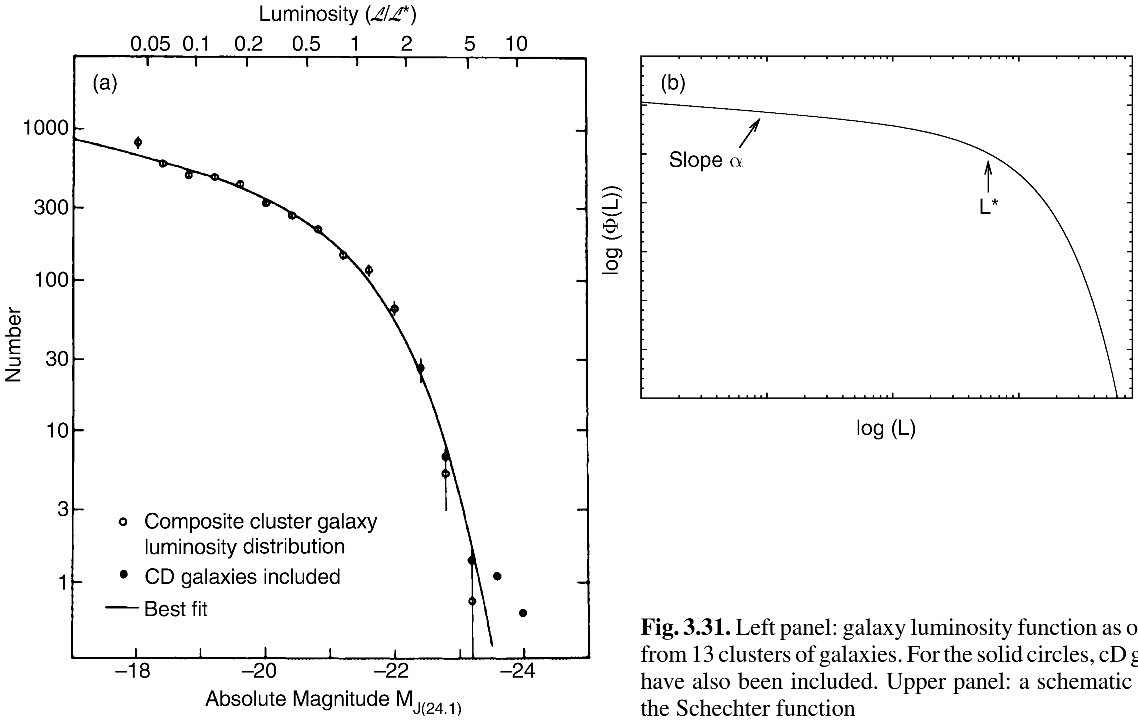

Fig. 3.31. Left panel: galaxy luminosity function as obtained from 13 clusters of galaxies. For the solid circles, cD galaxies have also been included. Upper panel: a schematic plot of the Schechter function

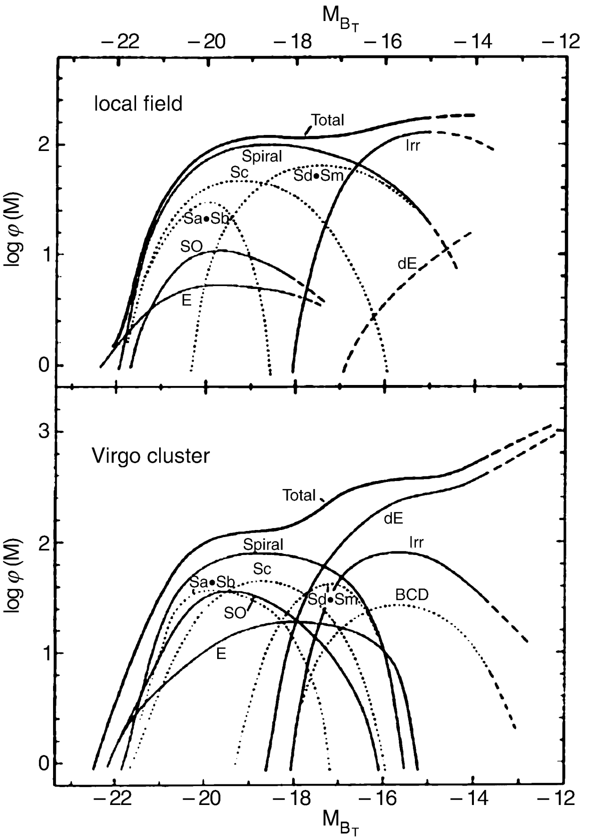

Fig. 3.32. The luminosity function for different Hubble types of field galaxies (top) and galaxies in the Virgo Cluster of galaxies (bottom). Dashed curves denote extrapolations. In contrast to Fig. 3.31, the more luminous galaxies are plotted towards the left. The Schechter luminosity function of the total galaxy distribution is compiled from the sum of the luminosity distributions of individual galaxy types that all deviate significantly from the Schechter function. One can see that in clusters the major contribution at faint magnitudes comes from the dwarf ellipticals (dEs), and that at the bright end ellipticals and S0's contribute much more strongly to the luminosity function than they do in the field. This trend is even more prominent in regular clusters of galaxies

ier to measure than morphology, in particular for very small galaxies. Therefore, one can study the luminosity function of galaxies, classifying them by their color.

Using photometric measurements and spectroscopy from the Sloan Digital Sky Survey (see Sect. 8.1.2), the colors and absolute magnitudes of $\sim 70\,000$ low-redshift galaxies has been studied; their density distribution in a color–magnitude diagram are plotted in the left-hand side of Fig. 3.33. From this figure we see immediately that there are two density peaks of the galaxy distribution in this diagram: one at high luminosities and red color, the other at significantly fainter absolute magnitudes and much bluer color. It appears that the galaxies are distributed at and around these two density peaks, hence galaxies tend to be either luminous and red, or less luminous and blue. We can also easily see from this diagram that the luminosity function of red galaxies is quite different from that of blue galaxies, which is another indication for the fact that the simple Schechter luminosity function (3.38) for the whole galaxy population most likely is a coincidence.

We can next consider the color distribution of galaxies at a fixed absolute magnitude M_r. This is obtained by plotting the galaxy number density along vertical cuts through the left-hand side of Fig. 3.33. When this is done for different M_r, it turns out that the color distribution of galaxies is bimodal: over a broad range in absolute magnitude, the color distribution has two peaks, one at red, the other at blue $u-r$. Again, this fact can be seen directly from Fig. 3.33. For each value of M_r, the color distribution of galaxies can be very well fitted by the sum of two Gaussian functions. The central colors of the two Gaussians is shown by the two dashed curves in the left panel of Fig. 3.33. They become redder the more luminous the galaxies are. This luminosity-dependent reddening is considerably more pronounced for the blue population than for the red galaxies.

To see how good this fit indeed is, the right-hand side of Fig. 3.33 shows the galaxy density as obtained from the two-Gaussian fits, with solid contours corresponding to the red galaxies and dashed contours to the blue ones. We thus conclude that the local galaxy population can be described as a bimodal distribution in $u-r$ color, where the characteristic color depends slightly on absolute magnitude. The galaxy distribution at bright absolute magnitudes is dominated by red galaxies, whereas for less luminous galaxies the blue population dominates. The luminosity function of both populations can be described by Schechter functions; however these two are quite different. The characteristic luminosity is about one magnitude brighter for the red galaxies than for the blue ones, whereas the faint-end slope α is significantly steeper for the blue galaxies. This

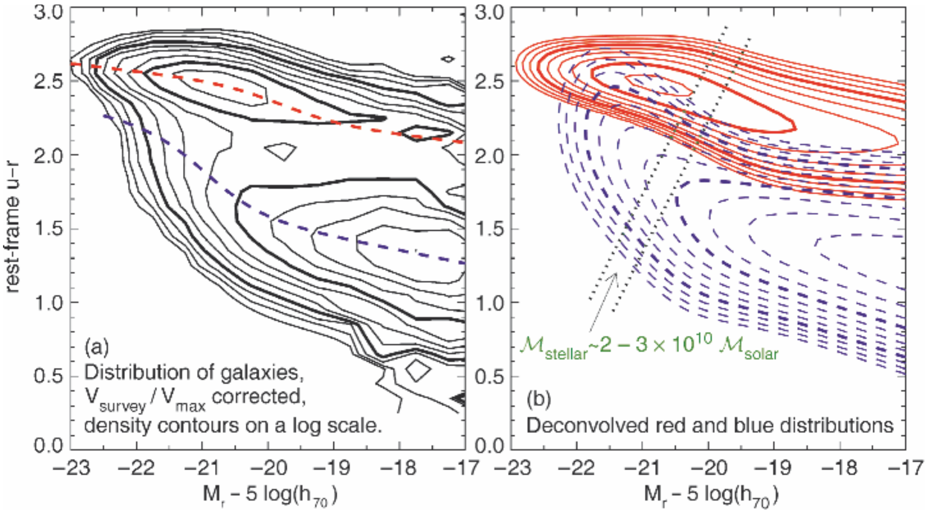

Fig. 3.33. The density of galaxies in color–magnitude space. The color of $\sim 70\,000$ galaxies with redshifts $0.01 \le z \le 0.08$ from the Sloan Digital Sky Survey is measured by the rest-frame $u-r$, i.e., after a (small) correction for their redshift was applied. The density contours, which were corrected for selection effects like the Malmquist bias, are logarithmically spaced, with a factor of $\sqrt{2}$ between consecutive contours. In the left-hand panel, the measured distribution is shown. Obviously, two peaks of the galaxy density are clearly visible, one at a red color of $u-r \sim 2.5$ and an absolute magnitude of $M_r \sim -21$, the other at a bluer color of $u-r \sim 1.3$ and significantly fainter magnitudes. The right-hand panel corresponds to the modeled galaxy density, as is described in the text

again is in agreement of what we just learned: for high luminosities, the red galaxies clearly dominate, whereas at small luminosities, the blue galaxies are much more abundant.

The mass-to-light ratio of a red stellar population is larger than that of a blue population, since the former no longer contains massive luminous stars. The difference in the peak absolute magnitude between the red and blue galaxies therefore corresponds to an even larger difference in the stellar mass of these two populations. Red galaxies in the local Universe have on average a much higher stellar mass than blue galaxies. This fact is illustrated by the two dotted lines in the right-hand panel of Fig. 3.33 which correspond to lines of constant stellar mass of $\sim 2\text{–}3 \times 10^{10}\,M_\odot$. This seems to indicate a very characteristic mass scale for the galaxy distribution: most galaxies with a stellar mass larger than this characteristic mass scale are red, whereas most of those with a lower stellar mass are blue.

Obviously, these statistical properties of the galaxy distribution must have an explanation in terms of the evolution of galaxies; we will come back to this issue in Chap. 9.

3.8 Galaxies as Gravitational Lenses

In Sect. 2.5 the gravitational lens effect was discussed, where we concentrated on the deflection of light by point masses. The lensing effect by stars leads to image separations too small to be resolved by any existing telescope. Since the separation angle is proportional to the square root of the lens mass (2.79), the angular separation of the images will be about a million times larger if a galaxy acts as a gravitational lens. In this case it should be observable, as was predicted in 1937 by Fritz Zwicky. Indeed, multiple images of very distant sources have been found, together with the galaxy responsible for the image splitting. In this section we will first describe this effect by continuing the discussion we began in Sect. 2.5.1. Examples of the lens effect and its various applications will then be discussed.

3.8.1 The Gravitational Lensing Effect – Part II

The geometry of a typical gravitational lens system is sketched in Fig. 2.21 and again in Fig. 3.34. The phys-

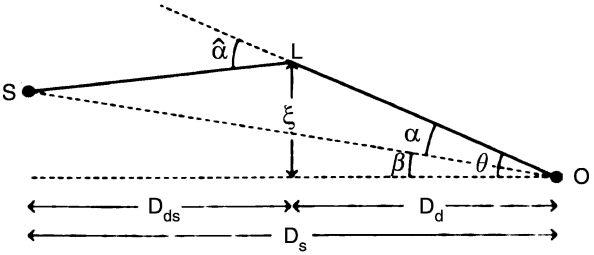

Fig. 3.34. As a reminder, another sketch of the lens geometry

ical description of such a lens system for an arbitrary mass distribution of the deflector is obtained from the following considerations.

If the gravitational field is weak (which is the case in all situations considered here), the gravitational effects can be linearized.[7] Hence, the deflection angle of a lens that consists of several mass components can be described by a linear superposition of the deflection angles of the individual components,

$$\hat{\boldsymbol{\alpha}} = \sum_i \hat{\boldsymbol{\alpha}}_i \; . \tag{3.45}$$

We assume that the deflecting mass has a small extent along the line-of-sight, as compared to the distances between observer and lens (D_d) and between lens and source (D_ds), $L \ll D_\mathrm{d}$ and $L \ll D_\mathrm{ds}$. All mass elements can then be assumed to be located at the same distance D_d. This physical situation is called a *geometrically thin lens*. If a galaxy acts as the lens, this condition is certainly fulfilled – the extent of galaxies is typically $\sim 100\,h^{-1}$ kpc while the distances of lens and source are typically $\sim$ Gpc. We can therefore write (3.45) as a superposition of Einstein angles of the form (2.71),

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \sum_i \frac{4Gm_i}{c^2} \frac{\boldsymbol{\xi} - \boldsymbol{\xi}_i}{|\boldsymbol{\xi} - \boldsymbol{\xi}_i|^2} \; , \tag{3.46}$$

where $\boldsymbol{\xi}_i$ is the projected position vector of the mass element m_i, and $\boldsymbol{\xi}$ describes the position of the light ray in the lens plane, also called the impact vector.

For a continuous mass distribution we can imagine subdividing the lens into mass elements of mass $\mathrm{d}m = \Sigma(\boldsymbol{\xi})\mathrm{d}^2\xi$, where $\Sigma(\boldsymbol{\xi})$ describes the *surface mass density* of the lens at the position $\boldsymbol{\xi}$, obtained by projecting the spatial (three-dimensional) mass density ρ along the line-of-sight to the lens. With this definition the deflection angle (3.46) can be transformed into an integral,

$$\boxed{\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \frac{4G}{c^2} \int \mathrm{d}^2\xi' \; \Sigma(\boldsymbol{\xi}') \frac{\boldsymbol{\xi} - \boldsymbol{\xi}'}{|\boldsymbol{\xi} - \boldsymbol{\xi}'|^2}} \; . \tag{3.47}$$

This deflection angle is then inserted into the lens equation (2.75),

$$\boxed{\boldsymbol{\beta} = \boldsymbol{\theta} - \frac{D_\mathrm{ds}}{D_\mathrm{s}} \hat{\boldsymbol{\alpha}}(D_\mathrm{d}\boldsymbol{\theta})} \; , \tag{3.48}$$

where $\boldsymbol{\xi} = D_\mathrm{d}\boldsymbol{\theta}$ describes the relation between the position $\boldsymbol{\xi}$ of the light ray in the lens plane and its apparent direction $\boldsymbol{\theta}$. We define the scaled deflection angle as in (2.76),

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \frac{D_\mathrm{ds}}{D_\mathrm{s}} \hat{\boldsymbol{\alpha}}(D_\mathrm{d}\boldsymbol{\theta}) \; ,$$

so that the lens equation (3.48) can be written in the simple form (see Fig. 3.34)

$$\boxed{\boldsymbol{\beta} = \boldsymbol{\theta} - \boldsymbol{\alpha}(\boldsymbol{\theta})} \; . \tag{3.49}$$

A more convenient way to write the scaled deflection is as follows,

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \frac{1}{\pi} \int \mathrm{d}^2\theta' \; \kappa(\boldsymbol{\theta}') \frac{\boldsymbol{\theta} - \boldsymbol{\theta}'}{|\boldsymbol{\theta} - \boldsymbol{\theta}'|^2} \; , \tag{3.50}$$

where

$$\kappa(\boldsymbol{\theta}) = \frac{\Sigma(D_\mathrm{d}\boldsymbol{\theta})}{\Sigma_\mathrm{cr}} \tag{3.51}$$

is the *dimensionless surface mass density*, and the so-called *critical surface mass density*

$$\Sigma_\mathrm{cr} = \frac{c^2 \, D_\mathrm{s}}{4\pi G \, D_\mathrm{d} \, D_\mathrm{ds}} \tag{3.52}$$

[7] To characterize the strength of a gravitational field, we refer to the gravitational potential Φ. The ratio Φ/c^2 is dimensionless and therefore well suited to distinguishing between strong and weak gravitational fields. For weak fields, $\Phi/c^2 \ll 1$. Another possible way to quantify the field strength is to apply the virial theorem: if a mass distribution is in virial equilibrium, then $v^2 \sim \Phi$, and weak fields are therefore characterized by $v^2/c^2 \ll 1$. Because the typical velocities in galaxies are ~ 200 km/s, for galaxies $\Phi/c^2 \lesssim 10^{-6}$. The typical velocities of galaxies in a cluster of galaxies are ~ 1000 km/s, so that in clusters $\Phi/c^2 \lesssim 10^{-5}$. Thus the gravitational fields occurring are weak in both cases.

depends only on the distances to the lens and to the source. Although $\Sigma_{\rm cr}$ incorporates a combination of cosmological distances, it is of a rather "human" order of magnitude,

$$\Sigma_{\rm cr} \approx 0.35 \left(\frac{D_{\rm d}\, D_{\rm ds}}{D_{\rm s}\; 1\,{\rm Gpc}} \right)^{-1} \; {\rm g\, cm^{-2}} \; .$$

A source is visible at several positions $\boldsymbol{\theta}$ on the sphere, or multiply imaged, if the lens equation (3.49) has several solutions $\boldsymbol{\theta}$ for a given source position $\boldsymbol{\beta}$. A more detailed analysis of the properties of this lens equation yields the following general result:

> If $\Sigma \geq \Sigma_{\rm cr}$ in at least one point of the lens, then source positions $\boldsymbol{\beta}$ exist such that a source at $\boldsymbol{\beta}$ has multiple images. It immediately follows that κ is a good measure for the strength of the lens. A mass distribution with $\kappa \ll 1$ at all points is a weak lens, unable to produce multiple images, whereas one with $\kappa \gtrsim 1$ for certain regions of $\boldsymbol{\theta}$ is a strong lens.

For sources that are small compared to the characteristic scales of the lens, the magnification μ of an image, caused by the differential light deflection, is given by (2.83), i.e.,

$$\boxed{\mu = \left| \det \left(\frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\theta}} \right) \right|^{-1}} \; . \tag{3.53}$$

The importance of the gravitational lens effect for extragalactic astronomy stems from the fact that gravitational light deflection is independent of the nature and the state of the deflecting matter. Therefore, it is equally sensitive to both dark and baryonic matter and independent of whether or not the matter distribution is in a state of equilibrium. The lens effect is thus particularly suitable for probing matter distributions, without requiring any further assumptions about the state of equilibrium or the relation between dark and luminous matter.

3.8.2 Simple Models

Axially Symmetric Mass Distributions. The simplest models for gravitational lenses are those which are axially symmetric, for which $\Sigma(\boldsymbol{\xi}) = \Sigma(\xi)$, where $\xi = |\boldsymbol{\xi}|$ denotes the distance of a point from the center of the lens. In this case, the deflection angle is directed radially inwards, and we obtain

$$\hat{\alpha} = \frac{4GM(\xi)}{c^2\, \xi} \; , \tag{3.54}$$

where $M(\xi)$ is the mass within radius ξ. Accordingly, for the scaled deflection angle we have

$$\alpha(\theta) = \frac{m(\theta)}{\theta} := \frac{1}{\theta}\, 2 \int_0^\theta {\rm d}\theta'\, \theta'\, \kappa(\theta') \; , \tag{3.55}$$

where, in the last step, $m(\theta)$ was defined as the dimensionless mass within θ. Since $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$ are collinear, the lens equation becomes one-dimensional because only the radial coordinate needs to be considered,

$$\beta = \theta - \alpha(\theta) = \theta - \frac{m(\theta)}{\theta} \; . \tag{3.56}$$

An illustration of this one-dimensional lens mapping is shown in Fig. 3.35.

Example: Point-Mass Lens. For a point mass M, the dimensionless mass becomes

$$m(\theta) = \frac{4GM}{c^2} \frac{D_{\rm ds}}{D_{\rm d}\, D_{\rm s}} \; ,$$

reproducing the lens equation from Sect. 2.5.1 for a point-mass lens.

Example: Isothermal Sphere. We saw in Sect. 2.4.2 that the rotation curve of our Milky Way is flat for large radii, and we know from Sect. 3.3.3 that the rotation curves of other spiral galaxies are flat as well. This indicates that the mass of a galaxy increases proportional to r, thus $\rho(r) \propto r^{-2}$, or more precisely,

$$\rho(r) = \frac{\sigma_v^2}{2\pi G r^2} \; . \tag{3.57}$$

Here, σ_v is the one-dimensional velocity dispersion of stars in the potential of the mass distribution if the distribution of stellar orbits is isotropic. In principle, σ_v is therefore measurable spectroscopically from the line width. The mass distribution described by (3.57) is called a *singular isothermal sphere* (SIS). Because this mass model is of significant importance not only for the analysis of the lens effect, we will discuss its properties in a bit more detail.

The density (3.57) diverges for $r \to 0$ as $\rho \propto r^{-2}$, so that the mass model cannot be applied up to the very center of a galaxy. However, the steep central increase of the rotation curve shows that the core region of the mass distribution, in which the density function will deviate considerably from the r^{-2}-law, must be small for galaxies. Furthermore, the mass diverges for large r such that $M(r) \propto r$. The mass profile thus has to be cut off at some radius in order to get a finite total mass. This cut-off radius is probably very large ($\gtrsim 100$ kpc for L^*-galaxies) because the rotation curves are flat to at least the outermost point at which they are observable.

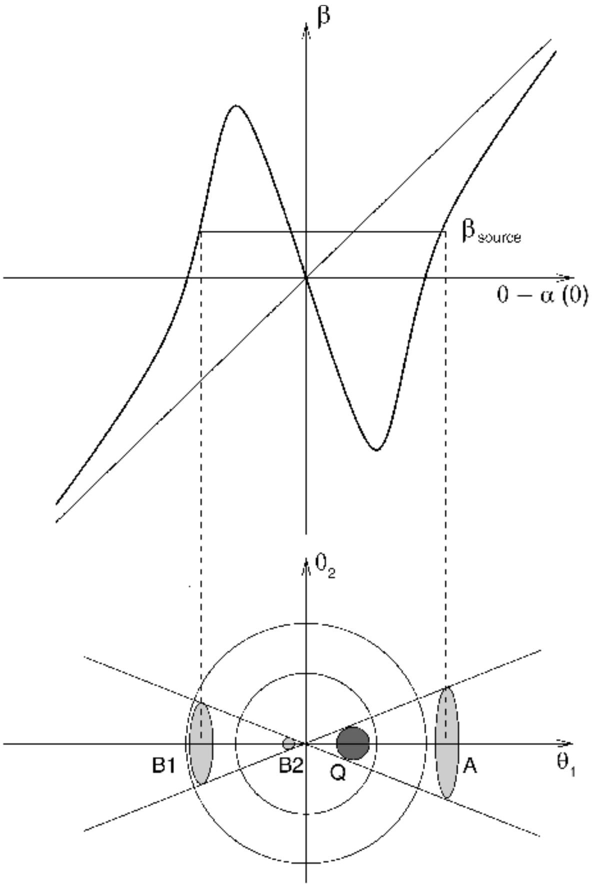

Fig. 3.35. Sketch of an axially symmetric lens. In the top panel, $\theta - \alpha(\theta)$ is plotted as a function of the angular separation θ from the center of the lens, together with the straight line $\beta = \theta$. The three intersection points of the horizontal line at fixed β with the curve $\theta - \alpha(\theta)$ are the three solutions of the lens equation. The bottom image indicates the positions and sizes of the images on the observer's sky. Here, Q is the unlensed source (which is not visible itself in the case of light deflection, of course!), and A, B1, B2 are the observed images of the source. The sizes of the images, and thus their fluxes, differ considerably; the inner image B2 is particularly weak in the case depicted here. The flux of B2 relative to that of image A depends strongly on the core radius of the lens; it can be so low as to render the third image unobservable. In the special case of a singular isothermal sphere, the innermost image is in fact absent

The SIS is an appropriate simple model for gravitational lenses over a wide range in radius since it seems to reproduce the basic properties of lens systems (such as image separation) quite well. The surface mass density is obtained from the projection of (3.57) along the line-of-sight,

$$\Sigma(\xi) = \frac{\sigma_v^2}{2G\xi} \,, \tag{3.58}$$

which yields the projected mass $M(\xi)$ within radius ξ

$$M(\xi) = 2\pi \int_0^{\xi} \mathrm{d}\xi' \, \xi' \, \Sigma(\xi') = \frac{\pi \sigma_v^2 \xi}{G} \,. \tag{3.59}$$

With (3.54) the deflection angle can be obtained,

$$\hat{\alpha}(\xi) = 4\pi \left(\frac{\sigma_v}{c}\right)^2 \,,$$

$$\boxed{\alpha(\theta) = 4\pi \left(\frac{\sigma_v}{c}\right)^2 \left(\frac{D_{\rm ds}}{D_{\rm s}}\right) \equiv \theta_{\rm E}} \,. \tag{3.60}$$

Thus the deflection angle for an SIS is constant and equals $\theta_{\rm E}$, and it depends quadratically on σ_v. $\theta_{\rm E}$ is called the *Einstein angle* of the SIS. The characteristic scale of the Einstein angle is

$$\theta_{\rm E} = 1\overset{\prime\prime}{.}15 \left(\frac{\sigma_v}{200\,\mathrm{km/s}}\right)^2 \left(\frac{D_{\rm ds}}{D_{\rm s}}\right) \,, \tag{3.61}$$

from which we conclude that the angular scale of the lens effect in galaxies is about an arcsecond for massive galaxies. The lens equation (3.56) for an SIS is

$$\beta = \theta - \theta_{\rm E} \, \frac{\theta}{|\theta|} \; , \tag{3.62}$$

where we took into account the fact that the deflection angle is negative for $\theta < 0$ since it is always directed inwards.

Solution of the Lens Equation for the Singular Isothermal Sphere. If $|\beta| < \theta_{\rm E}$, two solutions of the lens equation exist,

$$\theta_1 = \beta + \theta_{\rm E} \; , \quad \theta_2 = \beta - \theta_{\rm E} \; . \tag{3.63}$$

Without loss of generality, we assume $\beta \geq 0$; then $\theta_1 > \theta_{\rm E} > 0$ and $0 > \theta_2 > -\theta_{\rm E}$: one image of the source is located on either side of the lens center, and the separation of the images is

$$\boxed{\Delta\theta = \theta_1 - \theta_2 = 2\theta_{\rm E} = 2\overset{\prime\prime}{.}3 \left(\frac{\sigma_v}{200\,{\rm km/s}}\right)^2 \left(\frac{D_{\rm ds}}{D_{\rm s}}\right)} \; . \tag{3.64}$$

Thus, the angular separation of the images does not depend on the position of the source. For massive galaxies acting as lenses it is of the order of somewhat more than one arcsecond. For $\beta > \theta_{\rm E}$ only one image of the source exists, at θ_1, meaning that it is located on the same side of the center of the lens as the unlensed source.

For the magnification, we find

$$\mu(\theta) = \frac{|\theta/\theta_{\rm E}|}{||\theta/\theta_{\rm E}| - 1|} \; . \tag{3.65}$$

If $\theta \approx \theta_{\rm E}$, μ is very large. Such solutions of the lens equation exist for $|\beta| \ll \theta_{\rm E}$, so that sources close to the center of the source plane may be highly magnified. If $\beta = 0$, the image of the source will be a ring of radius $\theta = \theta_{\rm E}$, a so-called *Einstein ring*.

More Realistic Models. Mass distributions occurring in nature are not expected to be truly symmetric. The ellipticity of the mass distribution or external shear forces (caused, for example, by the tidal gravitational field of neighboring galaxies) will disturb the symmetry. The lensing properties of the galaxy will change by this symmetry breaking. For example, more than two images may be generated. Figure 3.36 illustrates the lens properties of such elliptical mass distributions. One can see, for example, that pairs of images, which are both heavily magnified, may be observed with a separation significantly smaller than the Einstein radius of the lens. Nevertheless, the characteristic image separation is still of the order of magnitude given by (3.64).

3.8.3 Examples for Gravitational Lenses

Currently, about 70 gravitational lens systems are known in which a galaxy acts as the lens. Some of them were discovered serendipitously, but most were found in systematic searches for lens systems. Amongst the most important lens surveys are: (1) *The HST Snapshot Survey.* The ~ 500 most luminous quasars have been imaged with the HST, and six lens systems have been identified. (2) *JVAS.* About 2000 bright radio sources with a flat radio spectrum (these often contain compact radio components, see Sect. 5.1.3) were scanned for multiple components with the VLA. Six lens systems have been found. (3) *CLASS.* Like in JVAS, radio sources with a flat spectrum were searched with the VLA for multiple components, but the flux limit was lower than in JVAS, which form a subset of the CLASS sources. The survey contains 15 000 sources, of which, to data, 22 have been identified as lenses. In this section we will discuss some examples of identified lens systems.

QSO 0957+561: The First Double Quasar. The first lens system was discovered in 1979 by Walsh, Carswell & Weymann when the optical identification of a radio source showed two point-like optical sources (see Fig. 3.37). Both could be identified as quasars located at the same redshift of $z_{\rm s} = 1.41$ and having very similar spectra (see Fig. 3.38). Deep optical images of the field show an elliptical galaxy situated between the two quasar images, at a redshift of $z_{\rm d} = 0.36$. The galaxy is so massive and so close to image B of the source that it *has to* produce a lens effect. However, the observed image separation of $\Delta\theta = 6\overset{\prime\prime}{.}1$ is considerably larger than expected from the lens effect by a single galaxy (3.64). The explanation for this is that the lens galaxy is located in a cluster of galaxies; the additional lens effect of the cluster adds to that of the galaxy,

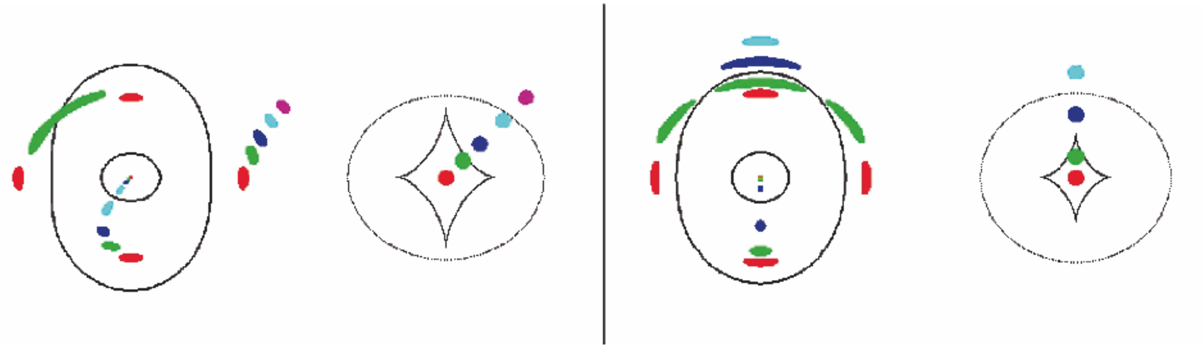

Fig. 3.36. Geometry of an "elliptical" lens, whereby it is of little importance whether the surface mass density Σ is constant on ellipses (i.e., the mass distribution has elliptical isodensity contours) or whether an originally spherical mass distribution is distorted by an external tidal field. On the right-hand side in both panels, several different source positions in the source plane are displayed, each corresponding to a different color. The origin in the source plane is chosen as the intersection point of the line connecting the center of symmetry in the lens and the observer with the source plane (see also Fig. 2.22). Depending on the position of the source, one, three, or five images may appear in the lens plane (i.e., the observer's sky); they are shown on the left-hand side of each panel. The curves in the lens plane are the *critical curves*, the location of all points for which $\mu \to \infty$. The curves in the source plane (i.e., on the right-hand side of each panel) are *caustics*, obtained by mapping the critical curves onto the source plane using the lens equation. Obviously, the number of images of a source depends on the source location relative to the location of the caustics. Strongly elongated images of a source occur close to the critical curves

boosting the image separation to a large value. The lens system QSO 0957+561 was observed in all wavelength ranges, from the radio to the X-ray. The two images of the quasar are very similar at all λ, including the VLBI structure (Fig. 3.38) – as would be expected since the lens effect is independent of the wavelength, i.e., achromatic.

QSO PG1115+080. In 1980, the so-called triple quasar was discovered, composed of three optical quasars at a maximum angular separation of just below $3''$. Component (A) is significantly brighter than the other two images (B, C; see Fig. 3.39, left). In high-resolution images it was found that the brightest image is in fact a double image: A is split into A1 and A2. The angular separation of the two roughly equally bright images is $\sim 0\overset{''}{.}5$, which is considerably smaller than all other angular separations in this system. The four quasar images have a redshift of $z_{\rm s} = 1.72$, and the lens is located at $z_{\rm d} = 0.31$. The image configuration is one of those that are expected for an elliptical lens, see Fig. 3.36.

With the NIR camera NICMOS on-board HST, not only were the quasar images and the lens galaxy observed, but also a nearly complete Einstein ring (Fig. 3.39, right). The source of this ring is the host galaxy of the quasar (see Sect. 5.4.5) which is substantially redder than the active galactic nucleus itself.

From the image configuration in such a quadruple system, the mass of the lens within the images can be estimated very accurately. The four images of the lens system trace a circle around the center of the lens galaxy, the radius of which can be identified with the Einstein radius of the lens. From this, the mass of the lens within the Einstein radius follows immediately because the Einstein radius is obtained from the lens equation (3.56) by setting $\beta = 0$. Therefore, the Einstein radius is the solution of the equation

$$\theta = \alpha(\theta) = \frac{m(\theta)}{\theta} \; ,$$

or

$$m(\theta_{\rm E}) = \frac{4GM(\theta_{\rm E})}{c^2} \frac{D_{\rm ds}}{D_{\rm d}\, D_{\rm s}} = \theta_{\rm E}^2 \; .$$

This equation is best written as

$$\boxed{M(\theta_{\rm E}) = \pi (D_{\rm d}\theta_{\rm E})^2 \, \Sigma_{\rm cr}} \; , \qquad (3.66)$$

which is readily interpreted:

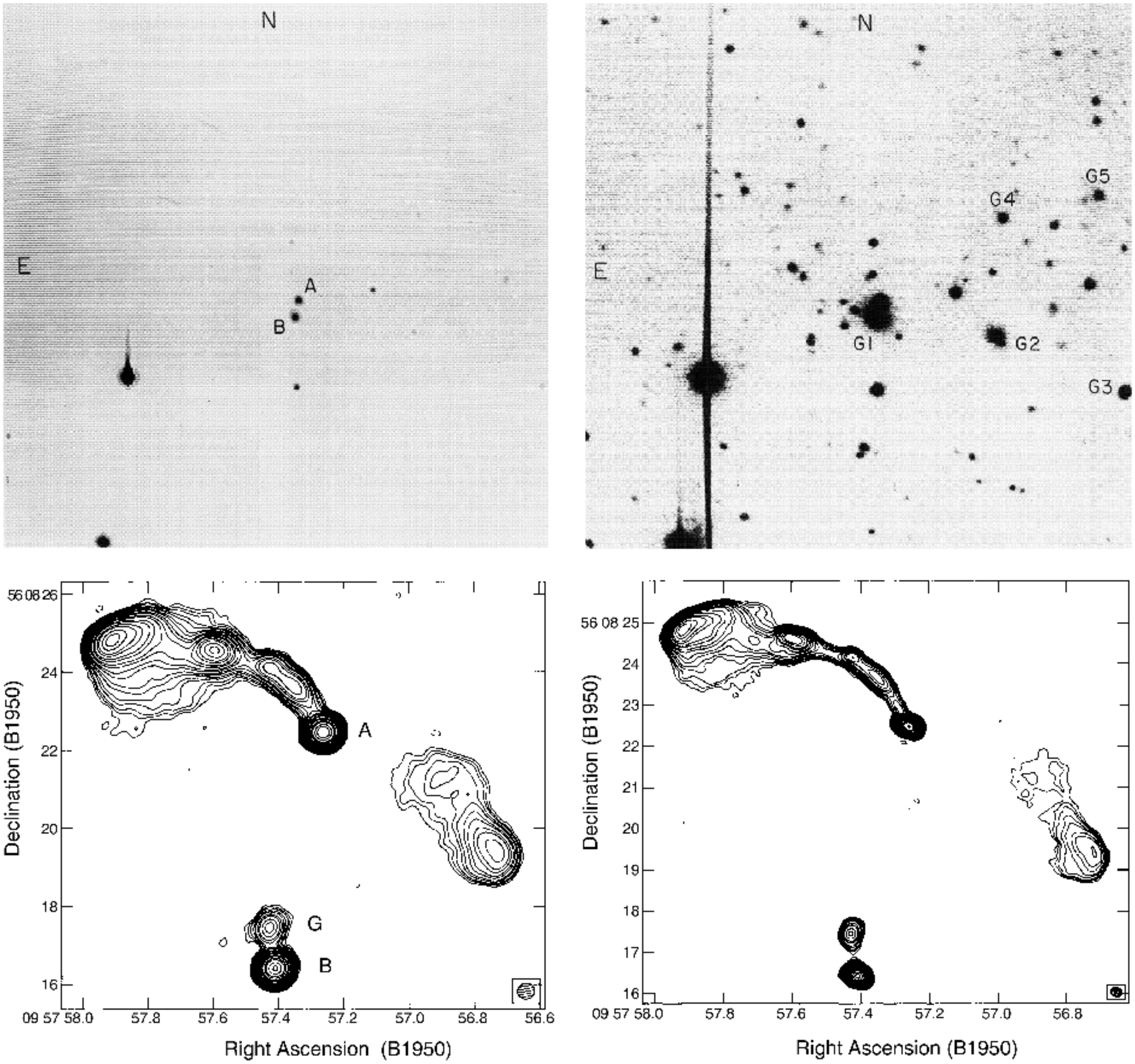

Fig. 3.37. Top: optical images of the double quasar QSO 0957+561. The image on the left has a short exposure time; here, the two point-like images A,B of the quasar are clearly visible. In contrast, the image on the right has a longer exposure time, showing the lens galaxy G1 between the two quasar images. Several other galaxies (G2-G5) are visible as well. The lens galaxy is a member of a cluster of galaxies at $z_\mathrm{d} = 0.36$. Bottom: two radio maps of QSO 0957+561, observed with the VLA at 6 cm (left) and 3.6 cm (right), respectively. The two images of the quasar are denoted by A,B; G is the radio emission of the lens galaxy. The quasar has a radio jet, which is a common property of many quasars (see Sect. 5.3.1). On small angular scales, the jet can be observed by VLBI techniques in both images (see Fig. 3.38). On large scales only a single image of the jet exists, seen in image A; this property should be compared with Fig. 3.36 where it was demonstrated that the number of images of a source (component) depends on its position in the source plane

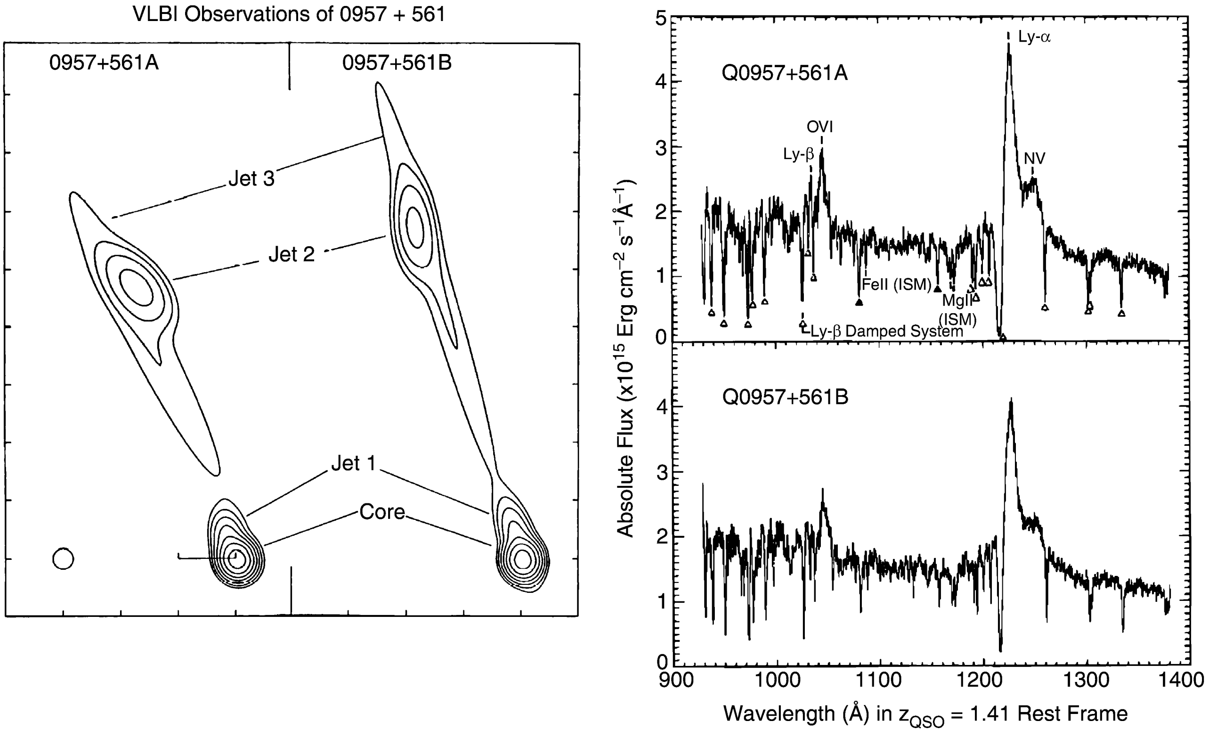

Fig. 3.38. Left: milliarcsecond structure of the two images of the quasar QSO 0957+561, a VLBI map at 13 cm wavelength by Gorenstein et al. Both quasar images show a core-jet structure, and it is clearly seen that they are mirror-symmetric, as predicted by lens models. right: spectra of the two quasar images QSO 0957+561A,B, observed by the Faint Object Camera (FOC) on-board HST. The similarity of the spectra, in particular the identical redshift, is a clear indicator of a common source of the two quasar images. The broad Lyα line, in the wings of which an Nv line is visible, is virtually always the strongest emission line in quasars

> The mass within $\theta_{\rm E}$ of a lens follows from the fact that the mean surface mass density within $\theta_{\rm E}$ equals the critical surface mass density $\Sigma_{\rm cr}$. A more accurate determination of lens masses is possible by means of detailed lens models. For quadruple image systems, the masses can be derived with a precision of a few percent – these are the most precise mass determinations in (extragalactic) astronomy.

QSO 2237+0305: The Einstein Cross. A spectroscopic survey of galaxies found several unusual emission lines in the nucleus of a nearby spiral galaxy which cannot originate from this galaxy itself. Instead, they are emitted by a background quasar at redshift $z_{\rm s} = 1.7$ situated exactly behind this spiral. High-resolution images show four point sources situated around the nucleus of this galaxy, with an image separation of $\Delta\theta \approx 1\farcs8$ (Fig. 3.40). The spectroscopic analysis of these point sources revealed that all four are images of the same quasar (Fig. 3.41).

The images in this system are positioned nearly symmetrically around the lens center; this is also a typical lens configuration which may be caused by an elliptical lens (see Fig. 3.36). The Einstein radius of this lens is $\theta_{\rm E} \approx 0\farcs9$, and we can determine the mass within this radius with a precision of $\sim 3\%$.

Einstein Rings. More examples of Einstein rings are displayed in Figs. 3.42 and 3.43. The first of these is a radio galaxy, with its two radio components be-

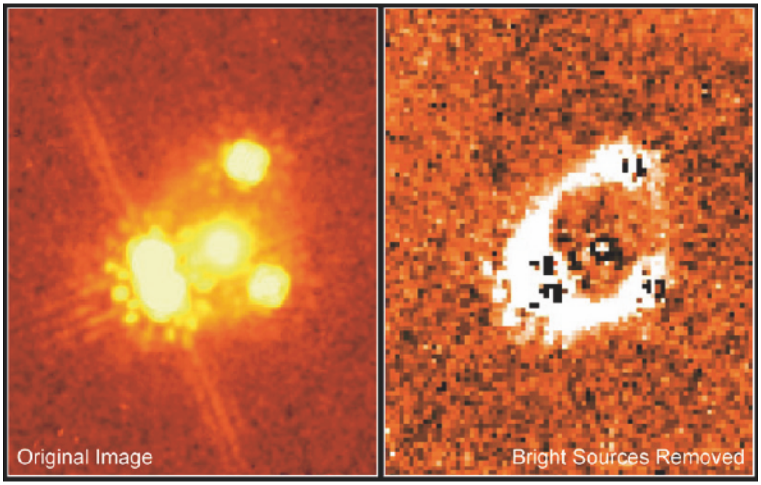

Fig. 3.39. A NIR image of QSO 1115+080 is shown on the left, as observed with the NICMOS camera on-board HST. The double structure of image A (the left of the QSO images) is clearly visible, although the image separation of the two A components is less than 0.″5. The lens galaxy, located in the "middle" of the QSO images, has a much redder spectral energy distribution than the quasar images. In the right-hand panel, the quasar images and the lens galaxy have been subtracted. What remains is a nearly closed ring; the light of the galaxy which hosts the active galactic nucleus is imaged into an Einstein ring

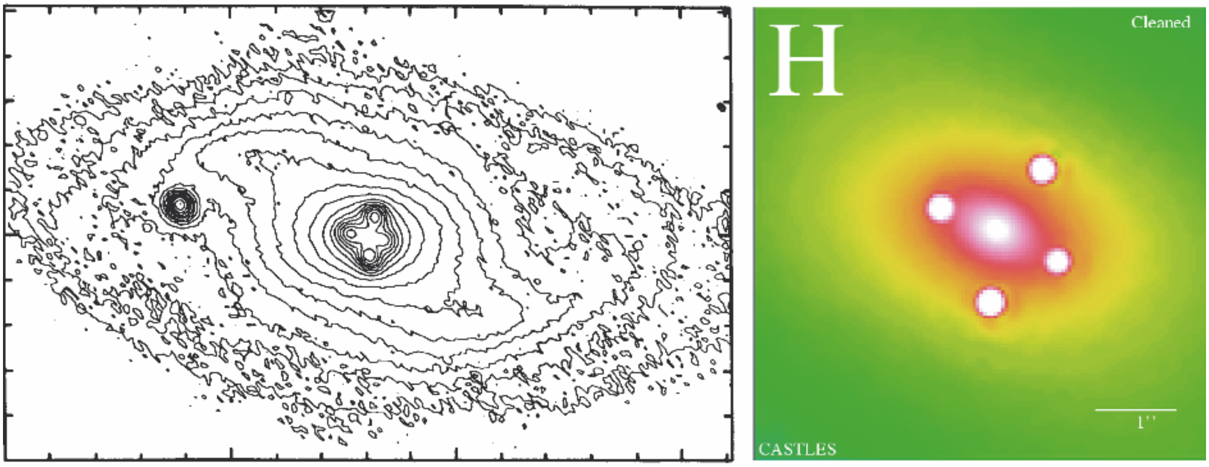

Fig. 3.40. Left: in the center of a nearby spiral galaxy, four point-like sources were found whose spectra show strong emission lines. This image from the CFHT clearly shows the bar structure in the core of the lens galaxy. An HST/NICMOS image of the center of QSO 2237+0305 is shown on the right. The central source is not a fifth quasar image but rather the bright nucleus of the lens galaxy

ing multiply imaged by a lens galaxy – one of the two radio sources is imaged into four components, the other mapped into a double image. In the NIR the radio galaxy is visible as a complete Einstein ring. This example shows very clearly that the appearance of the images of a source depends on the source size: to obtain an Einstein ring a sufficiently extended source is needed.

At radio wavelengths, the quasar MG 1654+13 consists of a compact central source and two radio lobes. As we will discuss in Sect. 5.1.3, this is a very typical radio morphology for quasars. One of the two lobes has a ring-shaped structure, which prior to this observation had never been observed before. An optical image of the field shows the optical quasar at the position of the compact radio component and, in addition, a bright elliptical galaxy right in the center of the ring-shaped radio lobe. This galaxy has a significantly lower redshift than the quasar and hence is the gravitational lens responsible for imaging the lobe into an Einstein ring.

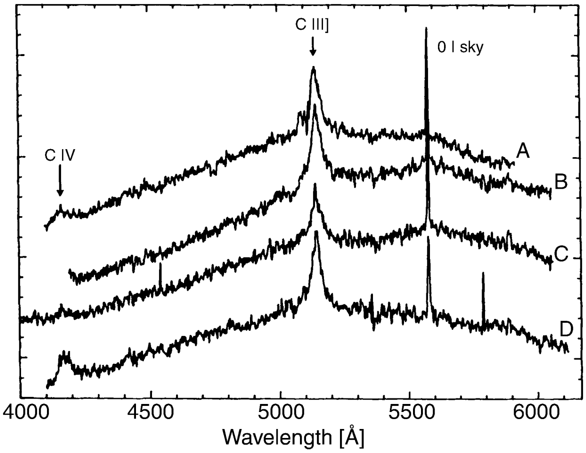

Fig. 3.41. Spectra of the four images of the quasar 2237+0305, observed with the CFHT. As is clearly visible, the spectral properties of these four images are very similar; this is the final proof that we are dealing with a lens system here. Measuring the individual spectra of these four very closely spaced sources is extremely difficult and can only be performed under optimum observing conditions

3.8.4 Applications of the Lens Effect

Mass Determination. As mentioned previously, the mass within a system of multiple images can be determined directly, sometimes very precisely. Since the length-scale in the lens plane (at given angular scale) and $\Sigma_{\rm cr}$ depend on H_0, these mass estimates scale with H_0. For instance, for QSO 2237+0305, a mass within $0\farcs9$ of $(1.08 \pm 0.02) h^{-1} \times 10^{10} M_\odot$ is derived. An even more precise determination of the mass was obtained for the lens galaxy of the Einstein ring in the system MG 1654+13 (Fig. 3.43). The dependence on the other cosmological parameters is comparatively weak, especially at low redshifts of the source and the lens. Most lens galaxies are early-type galaxies (ellipticals), and from the determination of their mass it can be concluded that ellipticals also contain dark matter.

Environmental Effects. Detailed lens models show that the light deflection of most gravitational lenses is affected by an external tidal field. This is due to the fact that lens galaxies are often members of galaxy groups which contribute to the light deflection as well. In some cases the members of the group have been identified. Mass properties of the corresponding group can be derived from the strength of this external influence.

Determination of the Hubble Constant. The light travel times along the different paths (according to the multiple images) are not the same. On the one hand the paths have different geometrical lengths, and on the other hand the light rays traverse different depths of the gravitational potential of the lens, resulting in a (general relativistic) time dilation effect. The difference in the light travel times Δt is measurable because luminosity variations of the source are observed at different times in the individual images. Δt can be measured from this difference in arrival time, called the time delay.

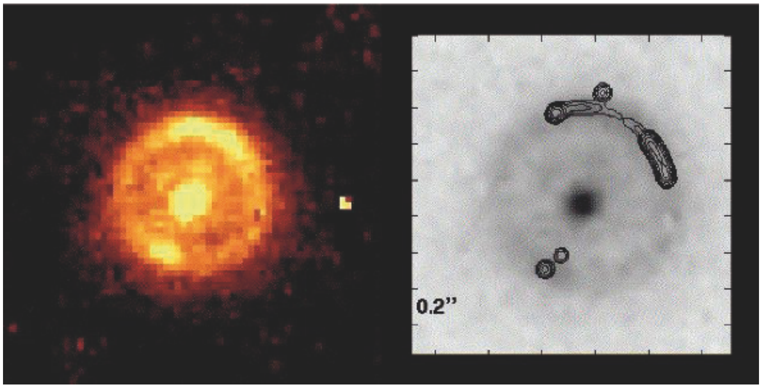

Fig. 3.42. The radio source 1938+666 is seen to be multiply imaged (contours in the right-hand figure); here, the radio source consists of two components, one of which is imaged four-fold, the other two-fold. A NIR image taken with the NICMOS camera onboard the HST (left-hand figure, also shown on the right in gray-scale) shows the lens galaxy in the center of an Einstein ring that originates from the stellar light of the host galaxy of the active galactic nucleus

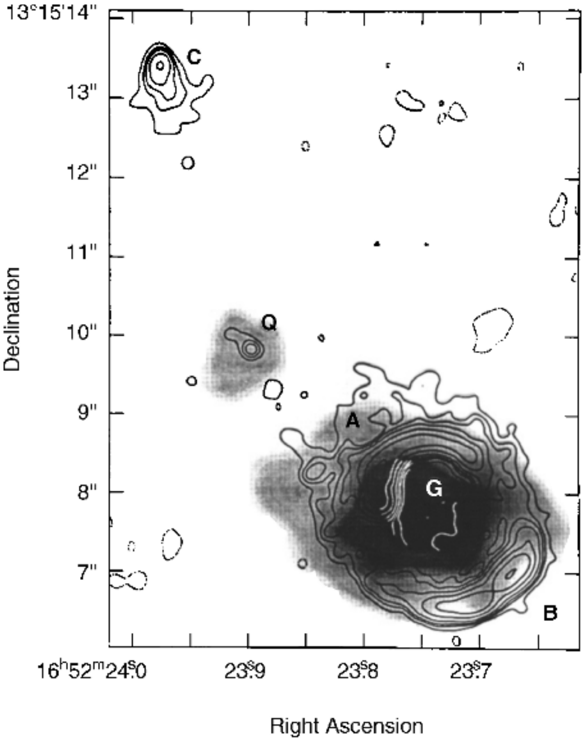

Fig. 3.43. The quasar MG1654+13 shows, in addition to the compact radio core (Q), two radio lobes; the northern lobe is denoted by C, whereas the southern lobe is imaged into a ring. An optical image is displayed in gray-scales, showing not only the quasar at Q ($z_s = 1.72$) but also a massive foreground galaxy at $z_d = 0.25$ that is responsible for the lensing of the lobe into an Einstein ring. The mass of this galaxy within the ring can be derived with a precision of $\sim 1\%$

It is easy to see that Δt depends on the Hubble constant, or in other words, on the size of the Universe. If a universe is twice the size of our own, Δt would be twice as large as well – see Fig. 3.44. Thus if the mass distribution of the lens can be modeled sufficiently well, by modeling the geometry of the image configuration, then the Hubble constant can be derived from measuring the difference in the light travel time. To date, Δt has been measured in about 10 lens systems (see Fig. 3.45 for an example). Based on "plausible" lens models we can derive values for the Hubble constant that are compatible with other measurements (see Sect. 3.6), but which tend towards slightly smaller values of H_0 than that determined from the HST Key Project (3.36). The main difficulty here is that the mass distribution in lens galaxies cannot unambiguously be derived from the positions of the multiple images. Therefore, these determinations of H_0 are currently not considered to be precision measurements. On the other hand, we can draw interesting conclusions about the radial mass profile of lens galaxies from Δt if we assume H_0 is known. In Sect. 6.3.4 we will discuss the value of H_0 determinations from lens time delays in a slightly different context.

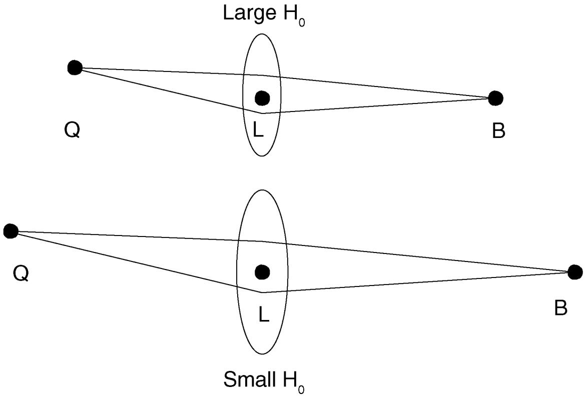

Fig. 3.44. Lens geometry in two universes with different Hubble constant. All observables are dimensionless – angular separations, flux ratios, redshifts – except for the difference in the light travel time. This is larger in the universe at the bottom than in the one at the top; hence, $\Delta t \propto H_0^{-1}$. If the time delay Δt can be measured, and if one has a good model for the mass distribution of the lens, then the Hubble constant can be derived from measuring Δt

The ISM in Lens Galaxies. Since the same source is seen along different sight lines passing through the lens galaxy, the comparison of the colors and spectra of the individual images provides information on reddening and on dust extinction in the ISM of the lens galaxy. From such investigations it was shown that the extinction in ellipticals is in fact very low, as is to be expected from the small amount of interstellar medium they contain, whereas the extinction is considerably higher for spirals. These analyses also enable us to study the relation between extinction and reddening, and from this to search for deviations from the Galactic reddening law (2.21). In fact, the constant of proportionality R_V is different in other galaxies, indicating a different composition of the dust, e.g., with respect to the chemical composition and to the size distribution of the dust grains.

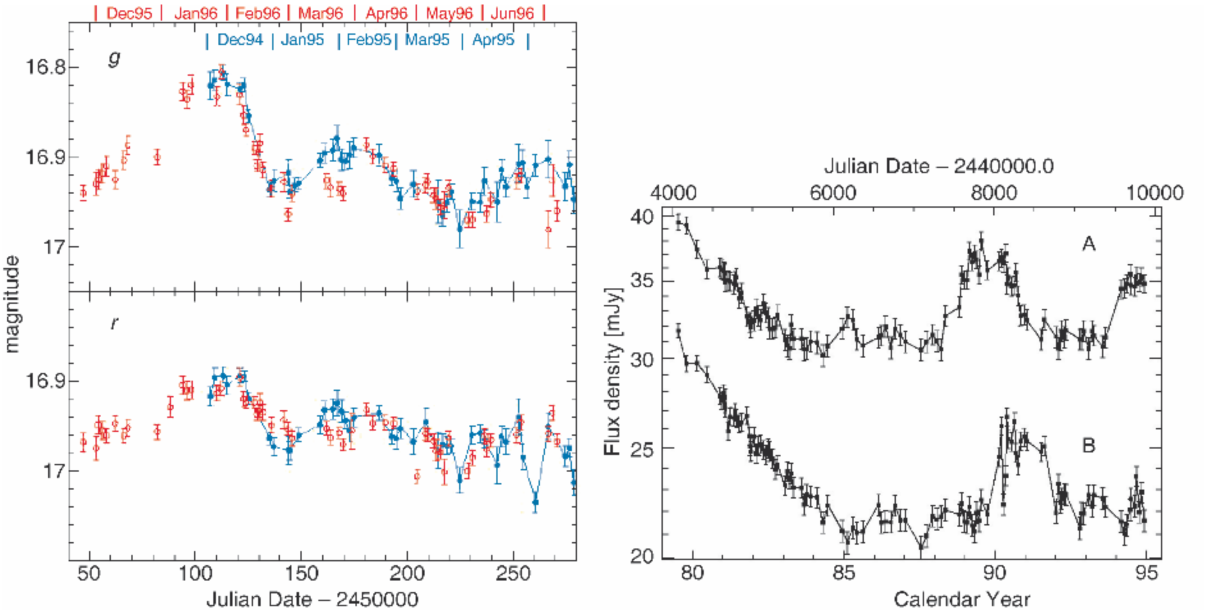

Fig. 3.45. Left: optical light curves of the double quasar 0957+561 in two broad-band filters. The light curve of image A is displayed in red and that of image B in blue, where the latter is shifted in time by 417 days. With this shift, the two light curves are made to coincide – this light travel time difference of 417 days is determined with an accuracy of $\sim \pm 3$ days. Right: radio light curves of QSO 0957+561A,B at 6 cm. From these radio measurements Δt can also be measured, and the corresponding value is compatible with that obtained from optical data

3.9 Population Synthesis

The light of normal galaxies originates from stars. Stellar evolution is largely understood, and the spectral radiation of stars can be calculated from the theory of stellar atmospheres. If the distribution of the number density of stars is known as a function of their mass, chemical composition, and evolutionary stage, we can compute the light emitted by them. The *theory of population synthesis* aims at interpreting the spectrum of galaxies as a superposition of stellar spectra. We have to take into account the fact that the distribution of stars changes over time; e.g., massive stars leave the main sequence after several 10^6 years, the number of luminous blue stars thus decreases, which means that the spectral distribution of the population also changes in time. The spectral energy distribution of a galaxy thus reflects its history of star formation and stellar evolution. For this reason, simulating different star-formation histories and comparing them with observed galaxy spectra provides important clues to understanding the evolution of galaxies. In this section, we will discuss some aspects of the theory of population synthesis; this subject is of tremendous importance for our understanding of galaxy spectra.

3.9.1 Model Assumptions

The processes of star formation are not understood in detail; for instance, it is currently impossible to compute the mass spectrum of a group of stars that jointly formed in a molecular cloud. Obviously, high-mass and low-mass stars are born together and form young (open) star clusters. The mass spectra of these stars are determined empirically from observations.

The *initial mass function* (IMF) is defined as the initial mass distribution at the time of birth of the stars, such that $\phi(m)\,\mathrm{d}m$ specifies the fraction of stars in the mass interval of width $\mathrm{d}m$ around m, where the distribution is normalized,

$$\boxed{\int_{m_L}^{m_U} \mathrm{d}m\; m\,\phi(m) = 1 M_\odot} \; .$$

The integration limits are not well defined. Typically, one puts $m_L \sim 0.1 M_\odot$ because less massive stars do not ignite their hydrogen (and are thus brown dwarfs), and $m_U \sim 100 M_\odot$, because more massive stars have not been observed. Such very massive stars would be difficult to observe because of their very short lifetime; furthermore, the theory of stellar structure tells us that more massive stars can probably not form a stable configuration due to excessive radiation pressure. The shape of the IMF is also subject to uncertainties; in most cases, the *Salpeter-IMF* is used,

$$\boxed{\phi(m) \propto m^{-2.35}} \;, \tag{3.67}$$

as obtained from investigating the stellar mass spectrum in young star clusters. It is by no means clear whether a universal IMF exists, or whether it depends on specific conditions like metallicity, the mass of the galaxy, or other parameters. The Salpeter IMF seems to be a good description for stars with $M \gtrsim 1 M_\odot$, whereas the IMF for less massive stars is less steep.

The *star-formation rate* is the gas mass that is converted into stars per unit time,

$$\psi(t) = -\frac{\mathrm{d}M_\mathrm{gas}}{\mathrm{d}t} \;.$$

The metallicity Z of the ISM defines the metallicity of the newborn stars, and the stellar properties in turn depend on Z. During stellar evolution, metal-enriched matter is ejected into the ISM by stellar winds, planetary nebulae, and SNe, so that $Z(t)$ is an increasing function of time. This chemical enrichment must be taken into account in population synthesis studies in a self-consistent form.

Let $S_{\lambda,Z}(t')$ be the emitted energy per wavelength and time interval, normalized to an initial total mass of $1 M_\odot$, emitted by a group of stars of initial metallicity Z and age t'. The function $S_{\lambda,Z(t-t')}(t')$, which describes this emission at any point t in time, accounts for the different evolutionary tracks of the stars in the Hertzsprung–Russell diagram (HRD) – see Appendix B.2. It also accounts for their initial metallicity (i.e., at time $t - t'$), where the latter follows from the chemical evolution of the ISM of the corresponding galaxy. Then the total spectral luminosity of this galaxy at a time t is given by

$$F_\lambda(t) = \int_0^t \mathrm{d}t' \; \psi(t-t') \; S_{\lambda,Z(t-t')}(t') \;, \tag{3.68}$$

thus by the convolution of the star-formation rate with the spectral energy distribution of the stellar population. In particular, $F_\lambda(t)$ depends on the star-formation history.

3.9.2 Evolutionary Tracks in the HRD; Integrated Spectrum

In order to compute $S_{\lambda,Z(t-t')}(t')$, models for stellar evolution and stellar atmospheres are needed. As a reminder, Fig. 3.46(a) displays the evolutionary tracks in the HRD. Each track shows the position of a star with specified mass in the HRD and is parametrized by the time since its formation. Positions of equal time in the HRD are called *isochrones* and are shown in Fig. 3.46(b). As time proceeds, fewer and fewer massive stars exist because they quickly leave the main sequence and end up as supernovae or white dwarfs. The number density of stars along the isochrones depends on the IMF. The spectrum $S_{\lambda,Z(t-t')}(t')$ is then the sum over all spectra of the stars on an isochrone – see Fig. 3.47(b).

In the beginning, the spectrum and luminosity of a stellar population are dominated by the most massive stars, which emit intense UV radiation. But after $\sim 10^7$ years, the flux below 1000 Å is diminished significantly, and after $\sim 10^8$ years, it hardly exists any more. At the same time, the flux in the NIR increases because the massive stars evolve into red supergiants.

For $10^8\,\mathrm{yr} \lesssim t \lesssim 10^9\,\mathrm{yr}$, the emission in the NIR remains high, whereas short-wavelength radiation is more and more diminished. After $\sim 10^9$ yr, red giant stars (RGB stars) account for most of the NIR production. After $\sim 3 \times 10^9$ yr, the UV radiation increases again due to blue stars on the horizontal branch into which stars evolve after the AGB phase, and due to white dwarfs which are hot when they are born. Between an age of 4 and 13 billion years, the spectrum of a stellar population evolves fairly little.

Of particular importance is the spectral break located at about 4000 Å which becomes visible in the spectrum after a few 10^7 years. This break is caused by a strongly changing opacity of stellar atmospheres at this wavelength, mainly due to strong transitions of singly ionized calcium and the Balmer lines of hydrogen. This 4000 Å-*break* is one of the most important spectral properties of galaxies; as we will discuss in Sect. 9.1.2, it allows us to estimate the redshifts of early-type galaxies from their

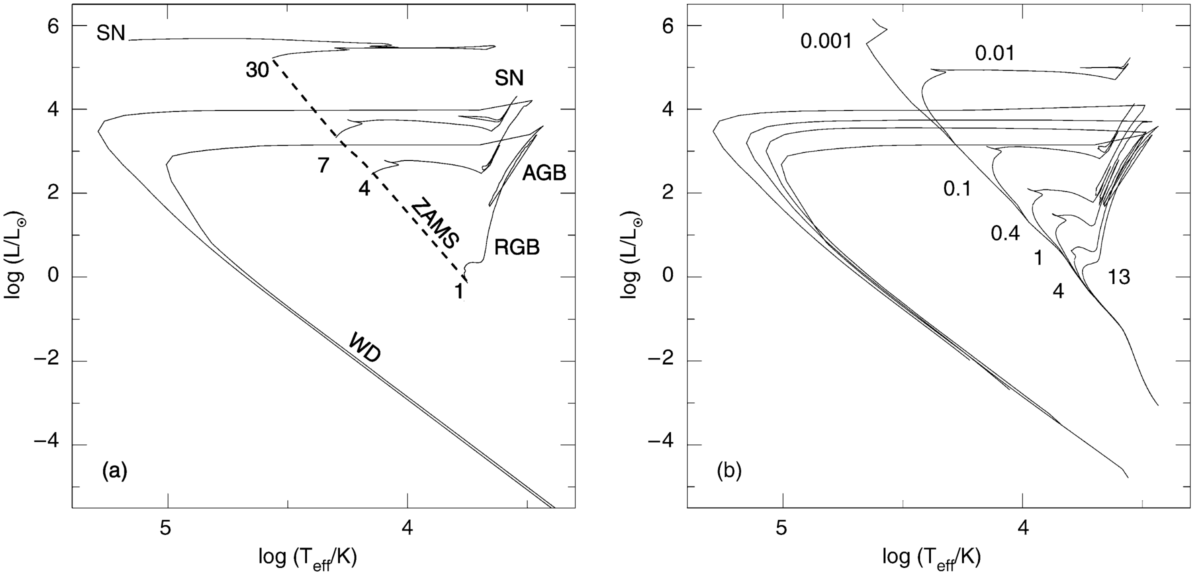

Fig. 3.46. a) Evolutionary tracks in the HRD for stars of different masses, as indicated by the numbers near the tracks (in units of $M_\odot$). The ZAMS (zero age main sequence) is the place of birth in the HRD; evolution moves stars away from the main sequence. Depending on the mass, they explode as a core-collapse SN (for $M \geq 8M_\odot$) or end as a white dwarf (WD). Prior to this, they move along the red giant branch (RGB) and the asymptotic giant branch (AGB). **b**) Isochrones at different times, indicated in units of 10^9 years. The upper main sequence is quickly depopulated by the rapid evolution of massive stars, whereas the red giant branch is populated over time

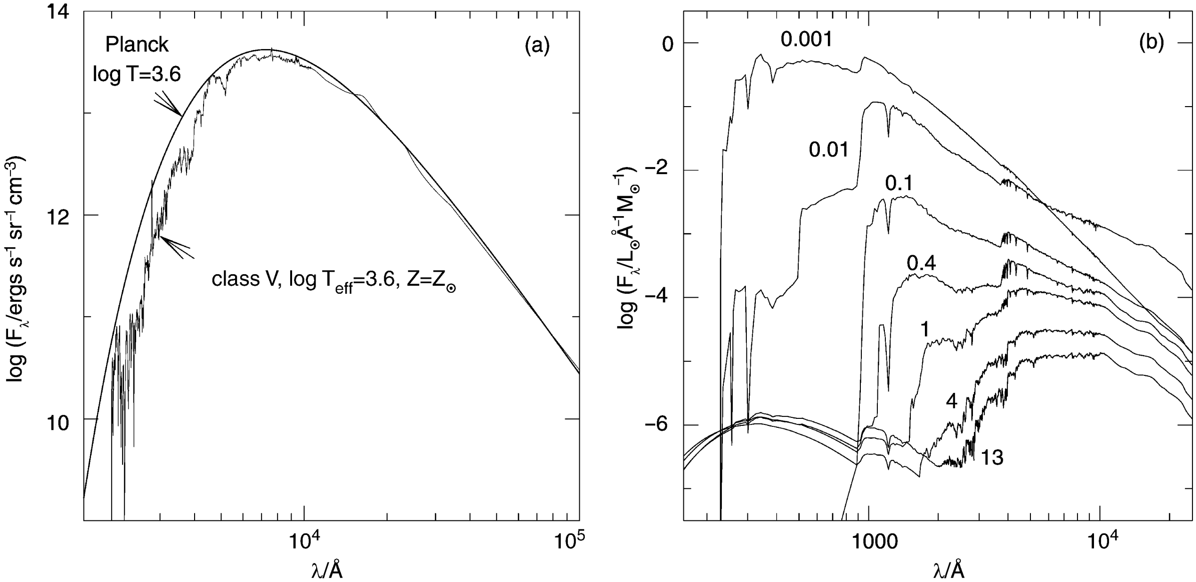

Fig. 3.47. a) Comparison of the spectrum of a main-sequence star with a blackbody spectrum of equal effective temperature. The opacity of the stellar atmosphere causes clear deviations from the Planck spectrum in the UV/optical. **b**) Spectrum of a stellar population with solar metallicity that was instantaneously born a time *t* ago; *t* is given in units of 10^9 years

photometric properties – so-called photometric redshift estimates.

3.9.3 Color Evolution

Detailed spectra of galaxies are often not available. Instead we have photometric images in different broadband filters, since the observing time required for spectroscopy is substantially larger than for photometry. In addition, modern wide-field cameras can obtain photometric data of numerous galaxies simultaneously. From the theory of population synthesis we can derive photometric magnitudes by multiplying model spectra with the filter functions, i.e., the transmission curves of the color filters used in observations, and then integrating over wavelength (A.25). Hence the spectral evolution implies a color evolution, as is illustrated in Fig. 3.48(a).

For a young stellar population the color evolution is rapid and the population becomes redder, again because the hot blue stars have a higher mass and thus evolve quickly in the HRD. For the same reason, the evolution is faster in $B-V$ than in $V-K$. It should be mentioned that this color evolution is also observed in star clusters of different ages. The mass-to-light ratio M/L also increases with time because M remains constant while L decreases.

As shown in Fig. 3.48(b), the blue light of a stellar population is always dominated by main-sequence stars, although at later stages a noticeable contribution also comes from horizontal branch stars. The NIR radiation is first dominated by stars burning helium in their center (this class includes the supergiant phase of massive stars), later by AGB stars, and after $\sim 10^9$ yr by red giants. Main sequence stars never contribute more than 20% of the light in the K-band. The fact that M/L_K varies only little with time implies that the NIR luminosity is a good indicator for the total stellar mass: the NIR mass-to-light ratio is much less dependent on the age of the stellar population than that for bluer filters.

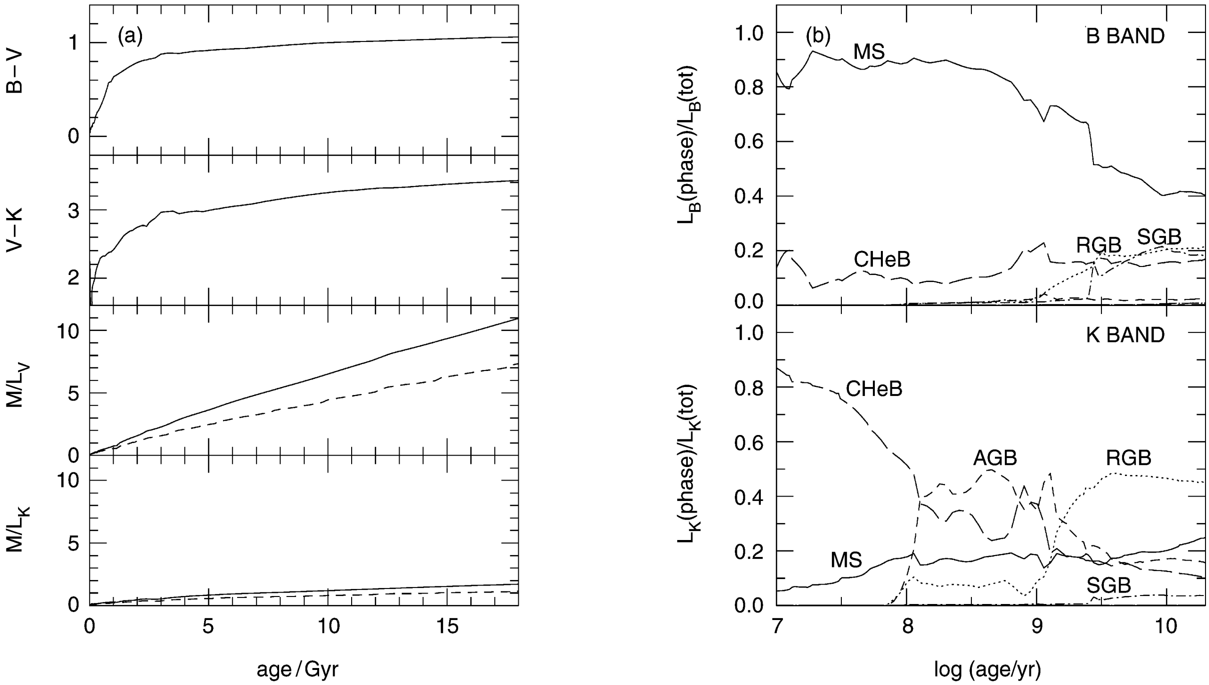

Fig. 3.48. a) For the same stellar population as in Fig. 3.47(b), the upper two graphs show the colors $B-V$ and $V-K$ as a function of age. The lower two graphs show the mass-to-light ratio M/L in two color bands in Solar units. The solid curves show the total M/L (i.e., including the mass that is later returned into the ISM), whereas the dashed curves show the M/L of the stars itself. **b**) The fraction of B- (top) and K-luminosity (bottom) contributed by stars in their different phases of stellar evolution (CHeB: core helium burning stars; SGB: subgiant branch)

3.9.4 Star Formation History and Galaxy Colors

Up to now, we have considered the evolution of a stellar population of a common age (called an *instantaneous burst of star formation*). However, star formation in a galaxy takes place over a finite period of time. We expect that the star-formation rate decreases over time because more and more matter is bound in stars and thus no longer available to form new stars. Since the star-formation history of a galaxy is a priori unknown, it needs to be parametrized in a suitable manner. A "standard model" of an exponentially decreasing star-formation rate was established for this,

$$\boxed{\psi(t) = \tau^{-1} \exp\left[-(t-t_{\rm f})/\tau\right] \, \mathrm{H}(t-t_{\rm f})} \; , \qquad (3.69)$$

where τ is the characteristic duration and $t_{\rm f}$ the onset of star formation. The last factor in (3.69) is the Heaviside step function, $\mathrm{H}(x) = 1$ for $x \geq 0$, $\mathrm{H}(x) = 0$ for $x < 0$. This Heaviside step function accounts for the fact that $\psi(t) = 0$ for $t < t_{\rm f}$. We may hope that this simple model describes the basic aspects of a stellar population. Results of this model are plotted in Fig. 3.49(a) in a color–color diagram.

From the diagram we find that the colors of the population depend strongly on τ. Specifically, galaxies do not become very red if τ is large because their star-formation rate, and thus the fraction of massive blue stars, does not decrease sufficiently. The colors of Sc spirals, for example, are not compatible with a constant star-formation rate – except if the total light of spirals is strongly reddened by dust absorption (but there are good reasons why this is not the case). To explain the colors of early-type galaxies we need $\tau \lesssim 4 \times 10^9$ yr. In general, one deduces from these models that a substantial evolution to redder colors occurs for $t \gtrsim \tau$. Since the luminosity of a stellar population in the blue spectral range decreases quickly with the age of the population, whereas increasing age affects the red luminosity much less, we conclude:

> The spectral distribution of galaxies is mainly determined by the ratio of the star-formation rate today to the mean star-formation rate in the past, $\psi(\text{today}) / \langle \psi \rangle$.

One of the achievements of this standard model is that it explains the colors of present day galaxies, which have an age of $\gtrsim$ 10 billion years. However, this model is not unambiguous because other star-formation histories $\psi(t)$ can be constructed with which the colors of galaxies can be modeled as well.

3.9.5 Metallicity, Dust, and HII Regions

Predictions of the model depend on the metallicity Z – see Fig. 3.49(b). A small value of Z results in a bluer color and a smaller M/L ratio. The age and metallicity of a stellar population are degenerate in the sense that an increase in the age by a factor X is nearly equivalent to an increase of the metallicity by a factor $0.65X$ with respect to the color of a population. The age estimate of a population from color will therefore strongly depend on the assumed value for Z. However, this degeneracy can be broken by taking several colors, or information from spectroscopy, into account.

Intrinsic dust absorption will also change the colors of a population. This effect cannot be easily accounted for in the models because it depends not only on the properties of the dust but also on the geometric distribution of dust and stars. For example, it makes a difference whether the dust in a galaxy is homogeneously distributed or concentrated in a thin disk. Empirically, it is found that galaxies show strong extinction during their active phase of star formation, whereas normal galaxies are presumably not much affected by extinction, with early-type galaxies (E/S0) affected the least.

Besides stellar light, the emission by HII regions also contributes to the light of galaxies. It is found, though, that after $\sim 10^7$ yr the emission from gas nebulae only marginally contributes to the broad-band colors of galaxies. However, this nebular radiation is the origin of emission lines in the spectra of galaxies. Therefore, emission lines are used as diagnostics for the star-formation rate and the metallicity in a stellar population.

3.9.6 Summary

After this somewhat lengthy section, we shall summarize the most important results of population synthesis here:

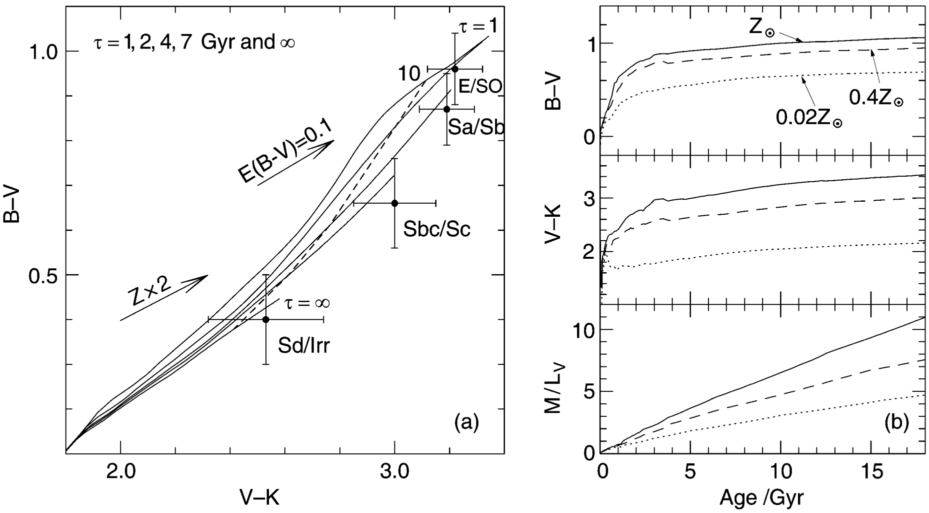

Fig. 3.49. a) Evolution of colors between $0 \leq t \leq 17 \times 10^9$ yr for a stellar population with star-formation rate given by (3.69), for five different values of the characteristic time-scale τ ($\tau = \infty$ is the limiting case for a constant star-formation rate) –Galactic center see solid curves. The typical colors for four different morphological types of galaxies are plotted. For each τ, the evolution begins at the lower left, i.e., as a blue population in both color indices. In the case of constant star formation, the population never becomes redder than Irr's; to achieve redder colors, τ has to be smaller. The dashed line connects points of $t = 10^{10}$ yr on the different curves. Here, a Salpeter IMF and Solar metallicity was assumed. The shift in color obtained by doubling the metallicity is indicated by an arrow, as well as that due to an extinction coefficient of $E(B-V) = 0.1$; both effects will make galaxies appear redder. **b**) The dependence of colors and M/L on the metallicity of the population

- A simple model of star-formation history reproduces the colors of today's galaxies fairly well.
- (Most of) the stars in elliptical and S0 galaxies are old – the earlier the Hubble type, the older the stellar population.
- Detailed models of population synthesis provide information about the star-formation history, and predictions by the models can be compared with observations of galaxies at high redshift (and thus smaller age).

We will frequently refer to results from population synthesis in the following chapters. For example, we will use them to interpret the colors of galaxies at high redshifts and the different spatial distributions of early-type and late-type galaxies (see Chap. 6). Also, we will present a method of estimating the redshift of galaxies from their broad-band colors (photometric redshifts). As a special case of this method, we will discuss the efficient selection of galaxies at very high redshift (Lyman-break galaxies, LBGs, see Chap. 9). Because the color and luminosity of a galaxy are changing even when no star formation is taking place, tracing back such a *passive evolution* allows us to distinguish this passive aging process from episodes of star formation and other processes.

3.9.7 The Spectra of Galaxies

At the end of this section we shall consider the typical spectra of different galaxy types. They are displayed for six galaxies of different Hubble types in Fig. 3.50. To make it easier to compare them, they are all plotted in a single diagram where the logarithmic flux scale is arbitrarily normalized (since this normalization does not affect the shape of the spectra).

It is easy to recognize the general trends in these spectra: the later the Hubble type, (1) the bluer the overall spectral distribution, (2) the stronger the emission lines, (3) the weaker the absorption lines, and (4) the smaller the 4000-Å break in the spectra. From the above discussion, we would also expect these trends if the Hubble sequence is considered an ordering of galaxy types

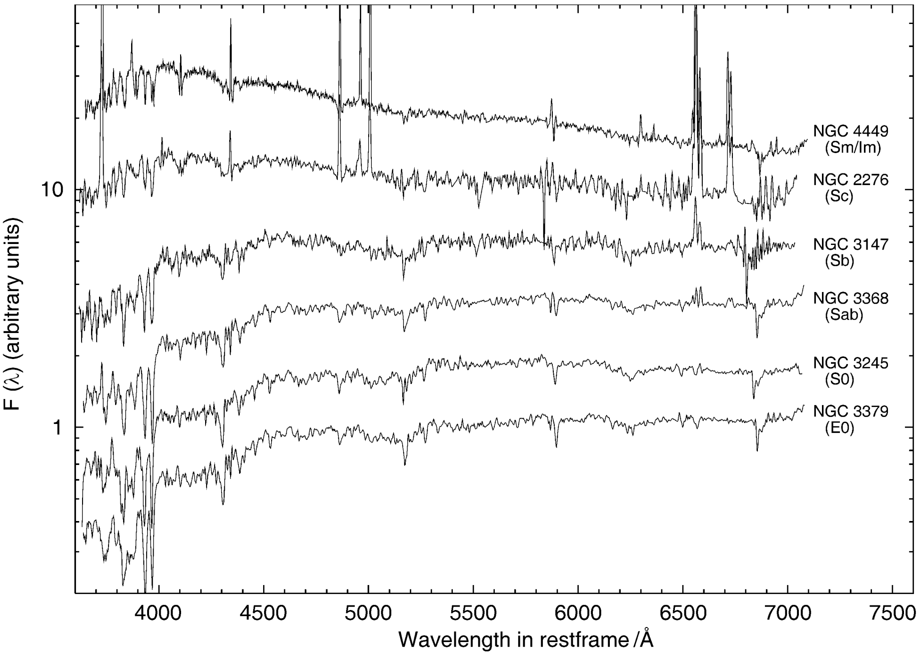

Fig. 3.50. Spectra of galaxies of different types, where the spectral flux is plotted logarithmically in arbitrary units. The spectra are ordered according to the Hubble sequence, with early types at the bottom and late-type spectra at the top

according to the characteristic age of their stellar population or according to their star-formation rate. Elliptical and S0 galaxies essentially have no star-formation activity, which renders their spectral energy distribution dominated by red stars. Furthermore, in these galaxies there are no HII regions where emission lines could be generated. The old stellar population produces a pronounced 4000-Å break, which corresponds to a jump by a factor of ~ 2 in the spectra of early-type galaxies. It should be noted that the spectra of ellipticals and S0 galaxies are quite similar.

By contrast, Sc spirals and irregular galaxies have a spectrum which is dominated by emission lines, where the Balmer lines of hydrogen as well as nitrogen and oxygen lines are most pronounced. The relative strength of these emission lines are characteristic for HII regions, implying that most of this line emission is produced in the ionized regions surrounding young stars. For irregular galaxies, the spectrum is nearly totally dominated by the stellar continuum light of hot stars and the emission lines from HII regions, whereas clear contributions by cooler stars can be identified in the spectra of Sc spiral galaxies.

The spectra of Sa and Sb galaxies form a kind of transition between those of early-type galaxies and Sc galaxies. Their spectra can be described as a superposition of an old stellar population generating a red continuum and a young population with its blue continuum and its emission lines. This can be seen in connection with the decreasing contribution of the bulge to the galaxy luminosity towards later spiral types.

The properties of the spectral light distribution of different galaxy types, as briefly discussed here, is described and interpreted in the framework of population synthesis. This gives us a detailed understanding of stellar populations as a function of the galaxy type. Extending these studies to spectra of high-redshift galaxies allows us to draw conclusions about the evolutionary history of their stellar populations.

3.10 Chemical Evolution of Galaxies

During its evolution, the chemical composition of a galaxy changes. Thus the observed metallicity yields information about the galaxy's star-formation history. We expect the metallicity Z to increase with star-formation rate, integrated over the lifetime of the galaxy. We will now discuss a simple model of the chemical evolution of a galaxy, which will provide insight into some of the principal aspects.

We assume that at the formation epoch of the stellar population of a galaxy, at time $t = 0$, no metals were present; hence $Z(0) = 0$. Furthermore, the galaxy did not contain any stars at the time of its birth, so that all baryonic matter was in the form of gas. In addition, we consider the galaxy as a closed system out of which no matter can escape or be added later on by processes of accretion or merger. Finally, we assume that the time-scales of the stellar evolution processes that lead to the metal enrichment of the galaxy are small compared to the evolutionary time-scale of the galaxy. Under these assumptions, we can now derive a relation between the metallicity and the gas content of a galaxy.

Of the total mass of a newly formed stellar population, part of it is returned to the ISM by supernova explosions and stellar winds. We define this fraction as R, so that the fraction $\alpha = (1 - R)$ of a newly-formed stellar population remains enclosed in stars, i.e., it no longer takes part in the further chemical evolution of the ISM. The value of α depends on the IMF of the stellar population and can be computed from models of population synthesis. Furthermore, let q be the ratio of the mass in metals, which is produced by a stellar population and then returned into the ISM, and the initial total mass of the population. The *yield* $y = q/\alpha$ is defined as the ratio of the mass in metals that is produced by a stellar population and returned into the ISM, and the mass that stays enclosed in the stellar population. The yield can also be calculated from population synthesis models. If $\psi(t)$ is the star-formation rate as a function of time, then the mass of all stars formed in the history of the galaxy is given by

$$S(t) = \int_0^t \mathrm{d}t' \; \psi(t') \; ,$$

and the total mass that remains enclosed in stars is $s(t) = \alpha S(t)$. Since we have assumed a closed system for the baryons, the sum of gas mass $g(t)$ and stellar mass $s(t)$ is a constant, namely the baryon mass of the galaxy,

$$g(t) + s(t) = M_\mathrm{b} \;\; \Rightarrow \;\; \frac{\mathrm{d}g}{\mathrm{d}t} + \frac{\mathrm{d}s}{\mathrm{d}t} = 0 \; . \qquad (3.70)$$

The mass of the metals in the ISM is gZ; it changes when stars are formed. Through this formation, the mass of the ISM and thus also that of its metals decreases. On the other hand, metals are also returned into the ISM by processes of stellar evolution. Under the above assumption that the time-scales of stellar evolution are small, this return occurs virtually instantaneously. The metals returned to the ISM are composed of metals that were already present at the formation of the stellar population – a fraction R of these will be returned – and newly formed metals. Together, the total mass of the metals in the ISM obeys the evolution equation

$$\frac{\mathrm{d}(gZ)}{\mathrm{d}t} = \psi \, (RZ + q) - Z\psi \; ,$$

where the last term specifies the rate of the metals extracted from the ISM in the process of star formation and the first term describes the return of metals to the ISM by stellar evolution processes. Since $\mathrm{d}S/\mathrm{d}t = \psi$, this can also be written as

$$\frac{\mathrm{d}(gZ)}{\mathrm{d}S} = (R - 1)Z + q = q - \alpha Z \; .$$

Dividing this equation by α and using $s = \alpha S$ and the definition of the yield, $y = q/\alpha$, we obtain

$$\frac{\mathrm{d}(gZ)}{\mathrm{d}s} = \frac{\mathrm{d}g}{\mathrm{d}s} Z + g \frac{\mathrm{d}Z}{\mathrm{d}s} = y - Z \; . \qquad (3.71)$$

From (3.70) it follows that $\mathrm{d}g/\mathrm{d}s = -1$ and $\mathrm{d}Z/\mathrm{d}s = -\mathrm{d}Z/\mathrm{d}g$, and so we obtain a simple equation for the metallicity,

$$g \frac{\mathrm{d}Z}{\mathrm{d}g} = \frac{\mathrm{d}Z}{\mathrm{d}\ln g} = -y$$
$$\Rightarrow \;\; Z(t) = -y \, \ln\left(\frac{g(t)}{M_\mathrm{b}}\right) = -y \, \ln(\mu_\mathrm{g}) \; , \qquad (3.72)$$

where $\mu_\mathrm{g} = g/M_\mathrm{b}$ is the fraction of baryons in the ISM, and where we chose the integration constant such that at the beginning, when $\mu_\mathrm{g} = 1$, the metallicity was $Z = 0$. From this relation, we can now see that with decreasing gas content in a galaxy, the metallicity will increase; in our simple model this increase depends only on the yield y. Since y can be calculated from population synthesis models, (3.72) is a well-defined relation.

If (3.72) is compared with observations of galaxies, rather strong deviations from this relation are found which are particularly prominent for low-mass galaxies. While the assumption of an instantaneous evolution of the ISM is fairly well justified, we know from structure formation in the Universe (Chap. 7) that galaxies are by no means isolated systems: their mass continuously changes through accretion and merging processes. In addition, the kinetic energy transferred to the ISM by supernova explosions causes an outflow of the ISM, in particular in low-mass galaxies where the gas is not strongly gravitationally bound. Therefore, the observed deviations from relation (3.72) allow us to draw conclusions about these processes.

Also, from observations in our Milky Way we find indications that the model of the chemical evolution sketched above is too simplified. This is known as the *G-dwarf problem*. The model described above predicts that about half of the F- and G-main-sequence stars should have a metallicity of less than a quarter of the Solar value. These stars have a long lifetime on the main sequence, so that many of those observed today should have been formed in the early stages of the Galaxy. Thus, in accordance with our model they should have very low metallicity. However, a low metallicity is in fact observed in only very few of these stars. The discrepancy is far too large to be explained by selection effects. Rather, observations show that the chemical evolution of our Galaxy must have been substantially more complicated than described by our simple model.