Active galactic nuclei

5

The light of normal galaxies in the optical and near infrared part of the spectrum is dominated by stars, with small contributions by gas and dust. This is thermal radiation since the emitting plasma in stellar atmospheres is basically in thermodynamical equilibrium. To a first approximation, the spectral properties of a star can be described by a Planck spectrum whose temperature depends on the stellar mass and the evolutionary state of the star. As we have seen in Sect. 3.5, the spectrum of galaxies can be described quite well as a superposition of stellar spectra. The temperature of stars varies over a relatively narrow range: Only few stars are found with $T \gtrsim 40\,000\,\mathrm{K}$, and those with $T \lesssim 3000\,\mathrm{K}$ hardly contribute to the spectrum of a galaxy, due to their low luminosity. Therefore, as a rough approximation, the light distribution of a galaxy can be described by a superposition of Planck spectra from a temperature range that covers about one decade. Since the Planck spectrum has a very narrow energy distribution around its maximum at $h_\mathrm{P}\nu \sim 3k_\mathrm{B}T$, the spectrum of a galaxy is basically confined to a range between $\sim$ 4000 and $\sim$ 20 000 Å. If the galaxy is actively forming stars, young hot stars extend this frequency range to somewhat higher frequency, and the thermal radiation from dust, heated by these new-born stars, extends the emission to the far-infrared.

However, there are galaxies which have a much broader energy distribution. Some of these show significant emission in the full range from radio wavelengths to the X-ray and even gamma range (see Fig. 3.4). This emission originates mainly from a very small central region of such an *active galaxy* which is called the *active galactic nucleus (AGN)*. This small emission region is structured and consists of multiple components with different physical properties, as we will see below. Active galaxies form a family of many different types of AGNs which differ in their spectral properties, including a wide range of ratios of radio-to-optical emission strength, their total luminosities and their ratio of nuclear luminosity to that of the stellar light. The optical spectra of three AGNs are presented in Fig. 5.1.

Some classes of AGNs, in particular the quasars, belong to the most luminous sources in the Universe, and they have been observed out to the highest measured redshifts ($z \sim 7$). The luminosity of quasars can exceed the luminosity of normal galaxies by a factor of a thousand. This luminosity originates from a tiny region in space, $r \lesssim 1\,\mathrm{pc}$. The optical/UV spectra of quasars are dominated by numerous strong and very broad emission lines, some of them emitted by highly ionized atoms (see Figs. 5.2 and 5.3). The processes in AGNs are among the most energetic ones in astrophysics. The enormous bandwidth of AGN spectra suggests that the radiation is non-thermal, i.e., not a superposition of (approximately) thermally radiating sources. As we will discuss later, processes in AGNs can produce highly energetic particles which are the origin of the non-thermal radiation.

After an introduction in which we will briefly present the history of the discovery of AGNs and their basic properties, in Sect. 5.2 we will describe the most important subgroups of the AGN family. In Sect. 5.3, we will discuss several arguments which lead to the conclusion that the energy source of an AGN originates in accretion of matter onto a supermassive black hole (SMBH). In particular, we will learn about the phenomenon of superluminal motion, where *apparent* velocities of source components are larger than the speed of light. We will then consider the different components of an AGN where radiation in different wavelength regions is produced.

Of particular importance for understanding the phenomenon of active galaxies are the unified models of AGNs that will be discussed next. We will see that the seemingly quite different appearances of AGNs can all be explained by geometric or projection effects. Finally, we will consider AGNs as cosmological probes. Due to their enormous luminosity they are observable up to very high redshifts.

P. Schneider, *Extragalactic Astronomy and Cosmology*, DOI 10.1007/978-3-642-54083-7_5,

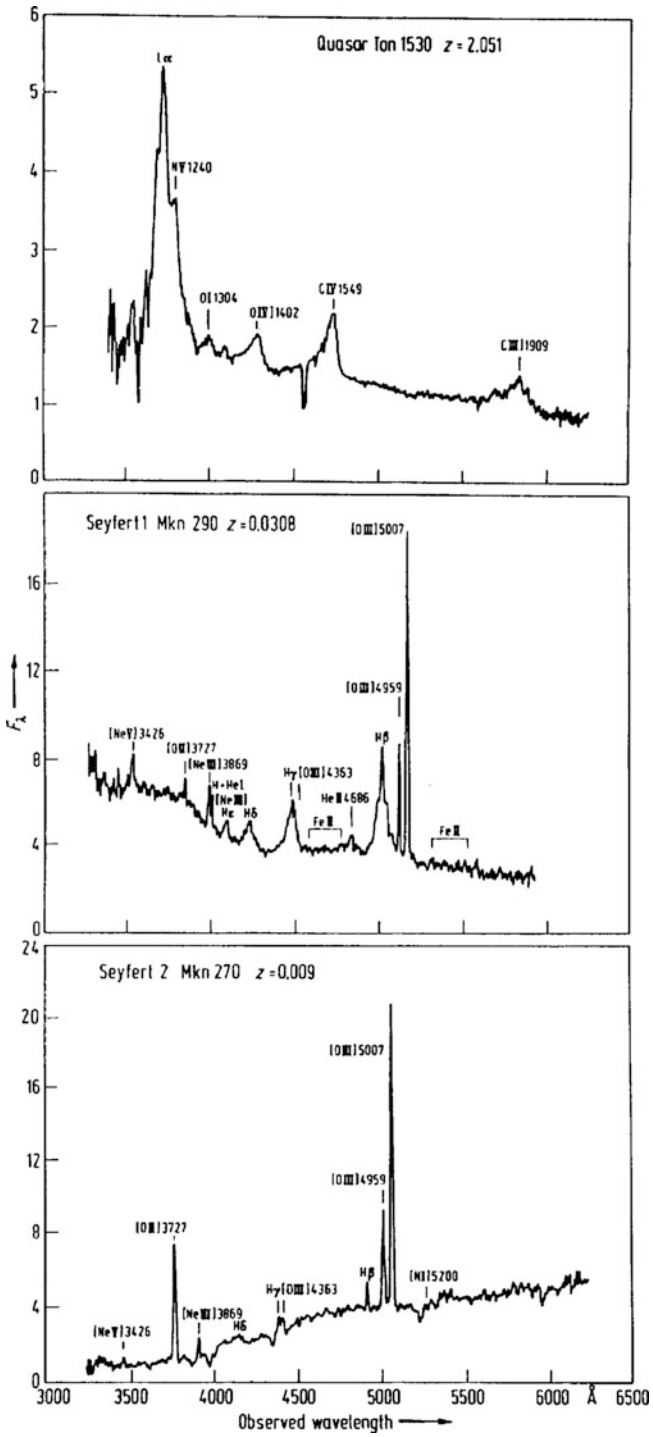

Fig. 5.1 Optical spectra of three AGNs. The *top panel* displays the spectrum of a quasar at redshift $z \sim 2$, which shows the characteristic broad emission lines. The strongest are Lyα of hydrogen, and the CIV-line and CIII]-line of triple and double ionized carbon, respectively (where the *squared bracket* means that this is a semi-forbidden transition, as will be explained in Sect. 5.4.2). The *middle panel* shows the spectrum of a nearby Seyfert galaxy of Type 1. Here both very broad emission lines and narrow lines, in particular of double ionized oxygen, are visible. In contrast, the spectrum in the *bottom panel*, of a Seyfert galaxy of Type 2, shows only relatively narrow emission lines. Source: H. Netzer 1990, in "Active Galactic Nuclei", eds. R.D. Blandford, H. Hetzer, L. Woltjer, Springer-Verlag 1990

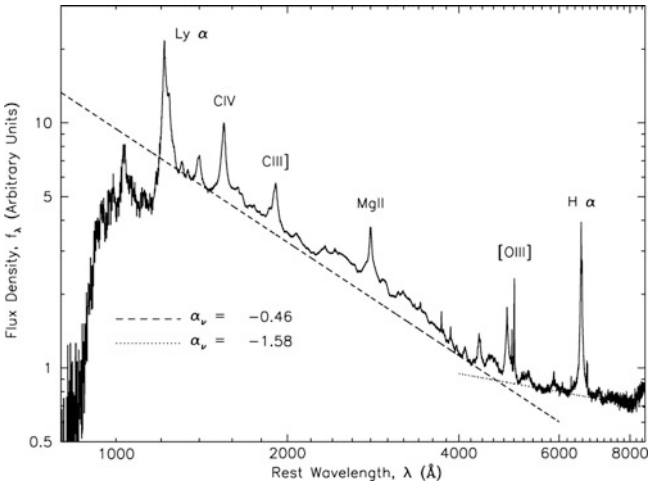

Fig. 5.2 The combined optical spectrum of more than 2200 QSOs from the SDSS. This 'mean' spectrum has a considerably better signal-to-noise ratio and a larger wavelength coverage than individual spectra. It was combined from the individual quasar spectra by transforming their wavelengths into the sources' rest-frames. The most prominent lines are marked. The *dashed* and *dotted lines* show power-law fits of the estimated underlying continuum emission of the QSOs. Source: D.E. Vanden Berk et al. 2001, *Composite Quasar Spectra from the Sloan Digital Sky Survey*, AJ 122, 549, p. 553, Fig. 3. ©AAS. Reproduced with permission

These observations allow us to draw conclusions about the properties of the early Universe.

5.1 Introduction

5.1.1 Brief history of AGNs

As long ago as 1908, strong emission lines were discovered in the galaxy NGC 1068 (see Fig. 5.4), which showed a line width of up to $\sim$ 3000 km/s. However, only the systematic analysis by Carl Seyfert in 1943 drew the focus of astronomers to this new class of galaxies. The cores of these *Seyfert galaxies* have an extremely high surface brightness, as demonstrated in Fig. 5.5, and the spectrum of their central region is dominated by emission lines of very high excitation. Some of these lines are extremely broad (see Fig. 5.1). The line width, when interpreted as Doppler broadening, $\Delta\lambda/\lambda = \Delta v/c$, yields values of up to $\Delta v \sim$ 8500 km/s for the full line width. The high excitation energy of some of the line-emitting atoms shows that they must have been excited by photons that are more energetic than photons from young stars that are responsible for the ionization of HII-regions. The hydrogen lines are often broader than other spectral lines. Most of the Seyfert galaxies are spirals, but one cD galaxy is also found in his original catalog.

In 1959, Lodewijk Woltjer argued that the extent of the cores of Seyfert galaxies cannot be larger than $r \lesssim 100$ pc because they appear point-like on optical images, i.e., they

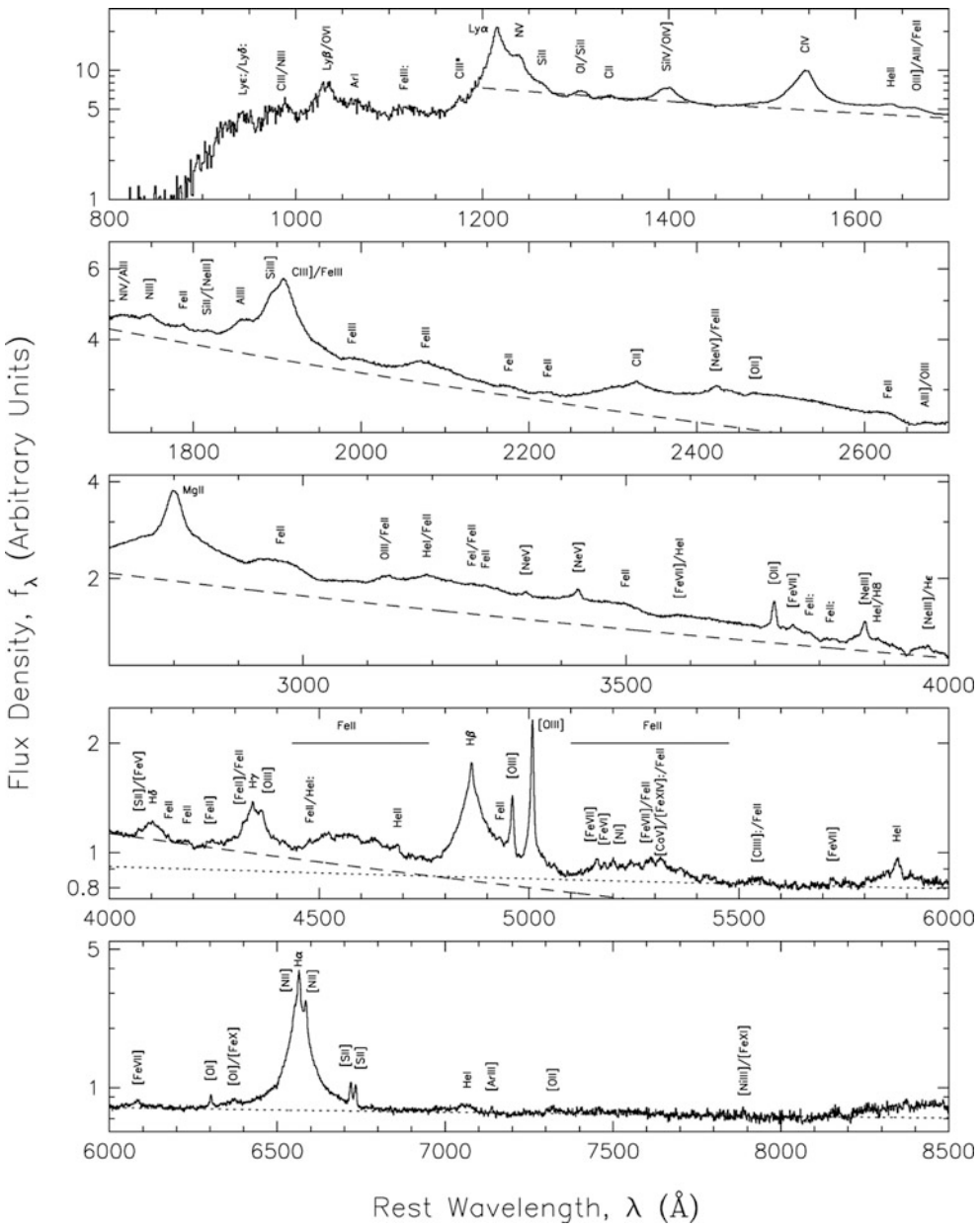

Fig. 5.3 An enlargement of the composite QSO spectrum shown in Fig. 5.2. Here, weaker spectral features are also visible. Also clearly seen is the break in the spectral flux bluewards of the Lyα line which is caused by the Lyα forest (Sect. 5.7), absorption by intergalactic hydrogen along the line-of-sight. The *dashed* and *dotted lines* indicate the average continuum. The substantial deviation of the spectrum from its estimated underlying continuum between 1600 and 3800 Å, even in spectral regions without obvious strong emission lines, is due to such a large number of overlapping iron lines that they blend into a quasi-continuum, and Balmer continuum (i.e., free-bound) radiation. Source: D.E. Vanden Berk et al. 2001, *Composite Quasar Spectra from the Sloan Digital Sky Survey*, AJ 122, 549, p. 555, Fig. 6. ©AAS. Reproduced with permission

are spatially not resolved. If the line-emitting gas is gravitationally bound, the relation

$$\frac{GM}{r} \simeq v^2$$

between the central mass $M(< r)$, the separation r of the gas from the center, and the typical velocity v must be satisfied. The latter is obtained from the line width: typically $v \sim 1000\,\mathrm{km/s}$. Therefore, with $r \lesssim 100\,\mathrm{pc}$, a mass estimate is immediately obtained,

$$M \gtrsim 10^{10} \left(\frac{r}{100\,\mathrm{pc}} \right) M_\odot \,. \tag{5.1}$$

Thus, either $r \sim 100\,\mathrm{pc}$, which implies an enormous mass concentration in the center of these galaxies, or r is much smaller than the estimated upper limit, which then implies an enormous energy density inside AGNs.

An important milestone in the history of AGNs was made with the 3C and 3CR radio catalogs which were completed around 1960. These are surveys of the northern ($\delta > -22°$) sky at 158 and 178 MHz, with a flux limit of $S_{\min} = 9\,\mathrm{Jy}$ (a Jansky is the flux unit used by radio astronomers, where $1\,\mathrm{Jy} = 10^{-23}\,\mathrm{erg\,s^{-1}\,cm^{-2}\,Hz^{-1}}$). Many of these 3C sources could be identified with relatively nearby galaxies, but the low angular resolution of radio telescopes at these low frequencies and the resulting large positional uncertainty of the respective sources rendered the identification with

Fig. 5.4 Optical image of the Seyfert galaxy NGC 1068, obtained with the Hubble Space Telescope. This spiral galaxy, located at a distance of ~ 15 Mpc from us, is the prototype of the Type 2 Seyfert galaxies. Its active nucleus is seen as the intense, high surface brightness center; it is powered by accretion onto a $\sim 15 \times 10^6\, M_\odot$ central supermassive black hole. Credit: NASA, ESA & A. van der Hoeven

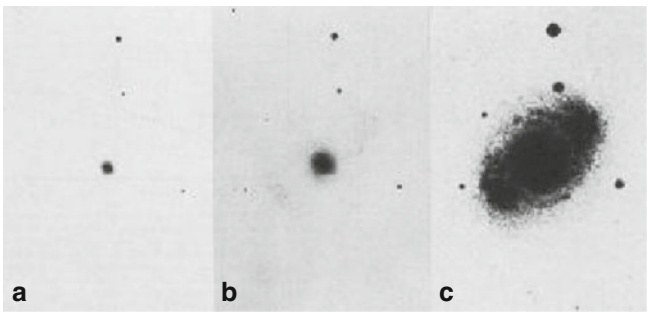

Fig. 5.5 Three images of the Seyfert galaxy NGC 4151, with the exposure time increasing to the right. In short exposures, the source appears point-like, with longer exposures displaying the galaxy. Source: W.W. Morgan 1968, *A Comparison of the Optical Forms of Certain Seyfert Galaxies with the N-Type Radio Galaxies*, ApJ 153, 27, PLATE 2, Fig. 2. ©AAS. Reproduced with permission

optical counterparts very difficult. If no striking nearby galaxy was found on optical photoplates within the positional uncertainty, the source was at first marked as unidentified.[1]

[1]The complete optical identification of the 3CR catalog, which was made possible by the enormously increased angular resolution of interferometric radio observations and thus by a considerably improved positional accuracy, was finalized only in the 1990s—some of these luminous radio sources are very faint optically.

In 1963, Thomas Matthews and Allan Sandage showed that 3C48 is a point-like ('stellar-like') source of $m = 16$ mag. It has a complex optical spectrum consisting of a blue continuum and strong, broad emission lines which could not be assigned to any atomic transition, and thus could not be identified. In the same year, Maarten Schmidt succeeded in identifying the radio source 3C273 with a point-like optical source which also showed strong and broad emission lines at unusual wavelengths. This was achieved by a lunar eclipse: the Moon passed in front of the radio source and eclipsed it. From the exact measurement of the time when the radio emission was blocked and became visible again, the position of the radio source was pinned down accurately. Schmidt could identify the emission lines of the source with those of the Balmer series of hydrogen, but at, for that time, an extremely high redshift of $z = 0.158$. Presuming the validity of the Hubble law and interpreting the redshift as cosmological redshift, 3C273 is located at the large distance of $D \sim 500h^{-1}$ Mpc. This huge distance of the source then implies an absolute magnitude of $M_B = -25.3 + 5 \log h$, i.e., it is about ~ 100 times brighter than normal (spiral) galaxies. Since the optical source had not been resolved but appeared point-like, this enormous luminosity must originate from a small spatial region. With the improving determination of radio source positions, many such *quasars* (quasi-stellar

radio sources = quasars) were identified in quick succession, the redshifts of some being significantly higher than that of 3C273.

5.1.2 Fundamental properties of quasars

In the following, we will review some of the most important properties of quasars. Although quasars are not the only class of AGNs, we will at first concentrate on them because they incorporate most of the properties of the other types of AGNs.

As already mentioned, quasars were discovered by identifying radio sources with point-like optical sources. Quasars emit at all wavelengths, from the radio to the X-ray, or even gamma-ray domain of the spectrum; see Fig. 3.4 for a sketch of the broad-band energy distribution of the quasar 3C 273. The radiation in the different frequency bands comes from various source components, as will be explained in the course of this chapter.

Interestingly, the flux of the source varies at nearly all frequencies, where the variability time-scale differs among the objects and also depends on the wavelength. As a rule, it is found that the variability time-scale of the observed radiation is smaller, and its amplitude larger, at higher frequencies. The optical spectrum is very blue; most quasars at redshifts $z \lesssim 2$ have $U - B < -0.3$ (for comparison: only hot white dwarfs have a similarly blue color index). Besides this blue continuum, very broad emission lines are characteristic of the optical and UV spectrum. Some of them correspond to transitions of very high ionization energy (see Fig. 5.3).

The continuum spectrum of a quasar can often be described, over a broad frequency range, by a power law of the form

$$S_\nu \propto \nu^{-\alpha} \,, \tag{5.2}$$

where α is the spectral index. $\alpha = 0$ corresponds to a flat spectrum, whereas $\alpha = 1$ describes a spectrum in which the same energy is emitted in every logarithmic frequency interval. Incidentally, the energy distribution of 3C 273 in Fig. 3.4 corresponds approximately to the latter case, over more than ten orders of magnitude in frequency, although over smaller frequency ranges, the spectral shape differs markedly from $\alpha = 1$.

5.1.3 AGNs as radio sources: synchrotron radiation

The morphology of quasars and other AGNs in the radio regime depends on the observed frequency and can often be very complex, consisting of several extended source components and one compact central one. In most cases, the extended component is observed as a double source in the form of two radio lobes situated more or less symmetrically around the optical position of the quasar. These lobes are frequently connected to the central core by jets, which are thin emission structures probably related to the energy transport from the core into the lobes (see Fig. 5.6 for an example). The observed length-scales are often impressive, in that the total extent of the radio source can reach values of up to 1 Mpc. The position of the optical quasar coincides with the compact radio source, which has an angular extent of $\ll 1''$ and is in some cases not resolvable even with VLBI methods. Thus the extent of these sources is $\lesssim 1$ mas, corresponding to $r \lesssim 1$ pc. This dynamical range in the extent of quasars is thus extremely large.

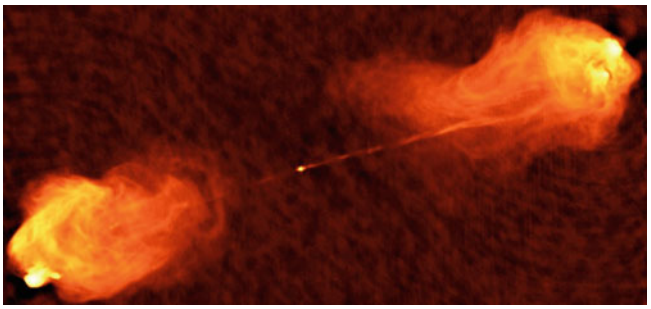

Fig. 5.6 The radio galaxy Cygnus A. Discovered by Grote Reber in 1939, it was the first very luminous active galaxy known. Cgynus A is one of the brightest radio sources in the sky, partly due to its low redshift ($z = 0.056$). This 6 cm map, covering a field of $2'.3 \times 1'.3$, impressively shows the narrow jet and counter-jet, and the two radio lobes on either side of the nucleus, in which the hot spots are embedded. Note how thin, straight, and highly collimated the jets are. Credit: Image courtesy of NRAO/AUI; Investigators: R. Perley, C. Carilli & J. Dreher

Classification of radio sources. Extended radio sources are often divided into two classes. *Fanaroff–Riley Type I* (FR I) are brightest close to the core, and the surface brightness decreases outwards. They typically have a luminosity of $L_\nu(1.4\,\mathrm{GHz}) \lesssim 10^{32}\,\mathrm{erg\,s^{-1}\,Hz^{-1}}$. In contrast, the surface brightness of *Fanaroff–Riley Type II* sources (FR II) increases outwards, and their luminosity is in general higher than that of FR I sources, $L_\nu(1.4\,\mathrm{GHz}) \gtrsim 10^{32}\,\mathrm{erg\,s^{-1}\,Hz^{-1}}$. One example for each of the two classes is shown in Fig. 5.7. FR II radio sources often have *jets*; they are extended linear structures that connect the compact core with a radio lobe. Jets often show internal structure such as knots and kinks. Their appearance indicates that they transport energy from the core out into the radio lobe. One of the most impressive examples of this is displayed in Fig. 5.8.

The jets are not symmetric. Often only one jet is observed, and in most sources where two jets are found one of them (the 'counter-jet') is much weaker than the other. The relative intensity of core, jet, and extended components varies with frequency, for sources as a whole and also within a source, because the components have different spectral indices. For

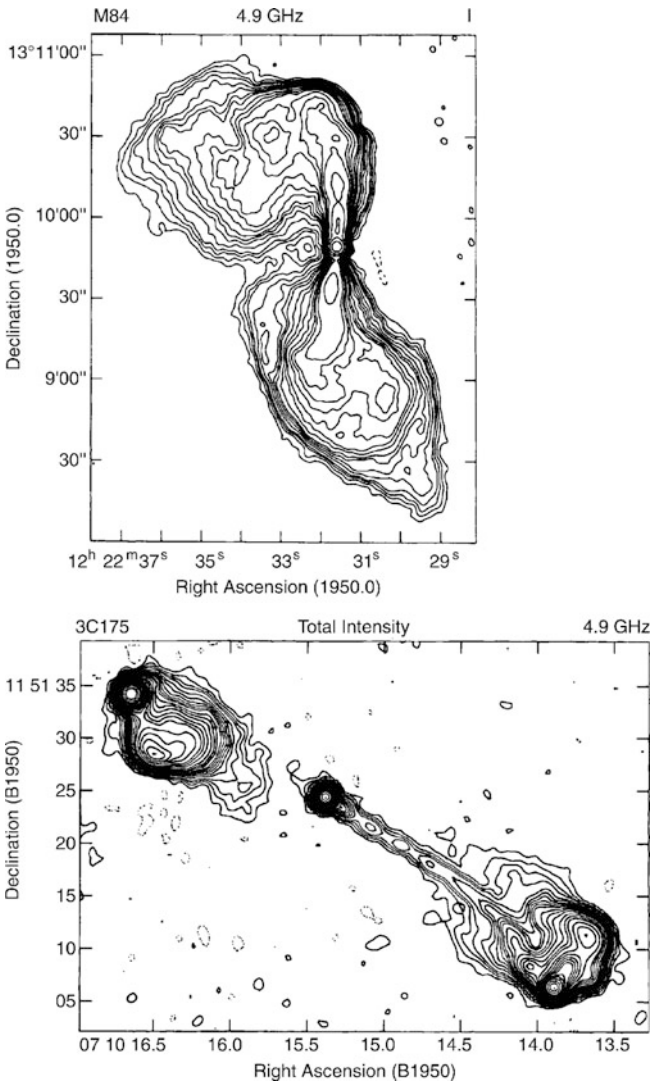

Fig. 5.7 Radio maps at $\lambda = 6$ cm of two radio galaxies: the *top* one is M84, an FR I-radio source, the *bottom* one is 3C175, an FR II-source. The radiation from M84 in the radio is strongest near the center and decreases outwards, whereas in 3C175 the most prominent components are the two radio lobes. The radio lobe on the right is connected to the compact core by a long and very thin jet, whereas on the opposite side no jet (counter-jet) is visible. Source: M84: R.A. Laing & A.H. Bridle 1987, *Rotation measure variation across M84*, MNRAS 228, 557, p. 559, Fig. 1. Reproduced by permission of Oxford University Press on behalf of the Royal Astronomical Society. 3C175: A.H. Bridle et al. 1994, *Deep VLA imaging of twelve extended 3CR quasars*, AJ 108, 766, p. 775, Fig. 7. ©AAS. Reproduced with permission

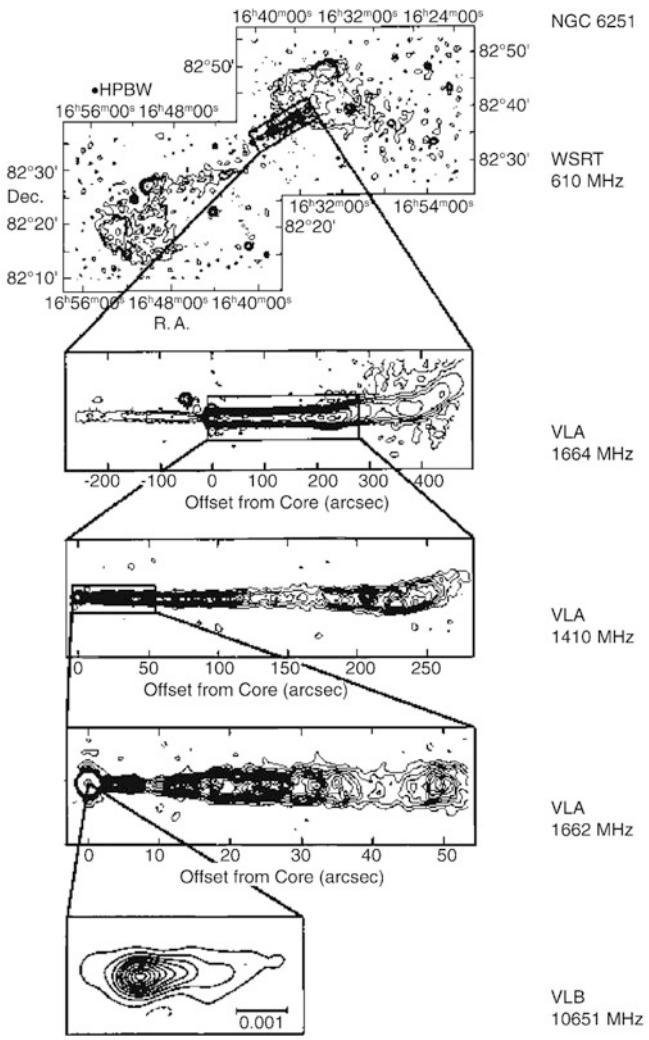

Fig. 5.8 The radio galaxy NGC 6251, with angular resolution increasing towards the bottom. On large scales (and at low frequencies), the two radio lobes dominate, while the core and the jets are clearly prominent at higher frequencies. NGC 6251 has a counter-jet, but with significantly lower luminosity than the main jet. Even at the highest resolution obtained by VLBI, structure can still be seen. The radio jets have a very small opening angle and are therefore strongly collimated. Source: A.H. Bridle & R. Perley 1984, *Extragalactic Radio Jets*, ARA&A 22, 319. Reprinted, with permission, from the *Annual Review of Astronomy & Astrophysics*, Volume 22 ©1984 by Annual Reviews www.annualreviews.org

this reason, radio catalogs of AGNs suffer from strong selection effects. Catalogs that are sampled at low frequencies will predominantly select sources that have a steep spectrum, i.e., in which the extended structures dominate, whereas high-frequency samples will preferentially contain core-dominated sources with a flat spectrum.[2]

Synchrotron radiation. Over a broad range in wavelengths, the radio spectrum of AGNs follows a power law of the form (5.2), with $\alpha \sim 0.7$ for the extended components and $\alpha \sim 0$ for the compact core components. Radiation in the radio is often linearly polarized, where the extended radio source may reach a degree of polarization up to 30 % or even more. The spectral form and the high degree of polarization are interpreted such that the radio emission is

[2]For this reason, radio surveys for gravitational lens systems, which were mentioned in Sect. 3.11.3, concentrate on sources with a flat spectral index because these are dominated by the compact nucleus. Multiple image systems are thus more easily recognized as such.

produced by *synchrotron radiation* of relativistic electrons. Electrons in a magnetic field propagate along a helical, i.e., corkscrew-shaped path, so that they are continually accelerated by the Lorentz force. Since accelerated charges emit electromagnetic radiation, this motion of the electrons leads to the emission of synchrotron radiation. Because of its importance for our understanding of the radio emission of AGNs, we will review some aspects of synchrotron radiation next.

The radiation can be characterized as follows. If an electron has energy $E = \gamma\, m_\mathrm{e}\, c^2$, the characteristic frequency of the emission is

$$\nu_\mathrm{c} = \frac{3\gamma^2 eB}{4\pi\, m_\mathrm{e} c} \sim 4.2 \times 10^6\, \gamma^2 \left(\frac{B}{1\mathrm{G}}\right) \mathrm{Hz}\,, \qquad (5.3)$$

where B denotes the magnetic field strength, e the electron charge, and $m_\mathrm{e} = 511\,\mathrm{keV}/c^2$ the mass of the electron. The *Lorentz factor* γ, and thus the energy of an electron, is related to its velocity v via

$$\gamma := \frac{1}{\sqrt{1-(v/c)^2}}\,. \qquad (5.4)$$

For frequencies considerably lower than ν_c, the spectrum of a single electron is $\propto \nu^{1/3}$, whereas it decreases exponentially for larger frequencies. To a first approximation, the spectrum of a single electron can be considered as quasi-monochromatic, i.e., the width of the spectral distribution is small compared to the characteristic emission frequency ν_c. The synchrotron radiation of a single electron is linearly polarized, where the observed polarization direction depends on the orientation of the magnetic field projected onto the sky. The degree of polarization of the radiation from an ensemble of electrons depends on the complexity of the magnetic field. If the magnetic field is uniform in the spatial region from which the radiation is measured, the observed polarization may reach values of up to 75 %. However, if the spatial region that lies within the telescope beam contains a complex magnetic field, with the direction changing strongly within this region, the polarizations partially cancel each other out and the observed degree of linear polarization is significantly reduced.

To produce radiation at cm wavelengths ($\nu \sim 10\,\mathrm{GHz}$) in a magnetic field of strength $B \sim 10^{-4}\,\mathrm{G}$, $\gamma \sim 10^5$ is required, i.e., the *electrons need to be highly relativistic!* To obtain particles at such high energies, very efficient processes of particle acceleration must occur in the inner regions of quasars. It should be mentioned in this context that cosmic ray particles of considerably higher energies are observed (see Sect. 2.3.4). The majority of cosmic rays are presumably produced in the shock fronts of supernova remnants. Thus, it is supposed that the energetic electrons in quasars (and other AGNs) are also produced by 'diffusive shock acceleration', where here the shock fronts are not caused by supernova explosions but rather by other hydrodynamical phenomena. As we will see later, we find clear indications in AGNs for outflow velocities that are considerably higher than the speed of sound in the plasma, so that the conditions for the formation of shock fronts are satisfied.

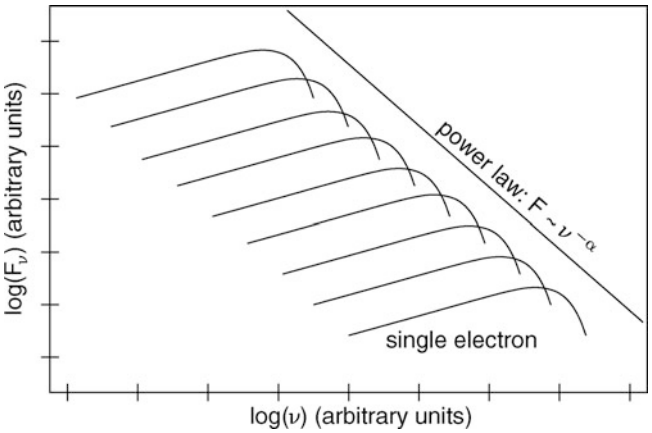

Fig. 5.9 Electrons at a given energy emit a synchrotron spectrum which is indicated by the *individual curves*; the maximum of the radiation is at ν_c (5.3), which depends on the electron energy. The superposition of many such spectra, corresponding to an energy distribution of the electrons, results in a power-law spectrum provided the energy distribution of the electrons follows a power law. Adapted from: B.W. Carroll & D.A. Ostlie 1996, *An introduction to Modern Astrophysics*, Reading

Spectral shape. Synchrotron radiation will follow a power law if the energy distribution of relativistic electrons also behaves like a power law (see Fig. 5.9). If $N(E)\,\mathrm{d}E \propto E^{-s}\,\mathrm{d}E$ represents the number density of electrons with energies between E and $E + \mathrm{d}E$, the power-law index of the resulting radiation will be $\alpha = (s-1)/2$, i.e., the slope in the power law of the electron distribution in energy defines the spectral shape of the resulting synchrotron emission (see Problem 5.1). In particular, an index of $\alpha = 0.7$ results for $s = 2.4$. An electron distribution with $N(E) \propto E^{-2.4}$ is very similar to the energy distribution of the cosmic rays in our Galaxy, which may be another indicator for the same or at least a similar mechanism being responsible for the generation of this energy spectrum.

The synchrotron spectrum is self-absorbed at low frequencies, i.e., the optical depth for absorption due to the synchrotron process is close to or larger than unity. In this case, the spectrum becomes flatter and, for small ν, it may even rise. In the limiting case of a high optical depth for self-absorption, we obtain $S_\nu \propto \nu^{2.5}$ for $\nu \to 0$. The extended radio components are optically thin at cm wavelength, so that $\alpha \sim 0.7$, whereas the compact core component is often

optically thick and thus self-absorbed, which yields $\alpha \sim 0$, or even inverted so that $\alpha < 0$.

Energy loss. Through emission, the electrons lose energy. Thus, the electrons cool and for only a limited time can they radiate at the frequency described by (5.3). The power emitted by an electron of Lorentz factor γ, integrated over all frequencies, is

$$P = -\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{4}{9}\frac{e^4 B^2 \gamma^2}{m_\mathrm{e}^2 c^3} \,. \tag{5.5}$$

The characteristic time in which an electron loses its energy is then obtained from its energy $E = \gamma m_\mathrm{e} c^2$ and its energy loss rate $\dot{E} = -P$ as

$$t_\mathrm{cool} = \frac{E}{P} = 2.4 \times 10^5 \left(\frac{\gamma}{10^4}\right)^{-1} \left(\frac{B}{10^{-4}\,\mathrm{G}}\right)^{-2} \mathrm{yr} \,. \tag{5.6}$$

For relatively low-frequency radio emission, this lifetime is longer than or comparable to the age of radio sources. But as we will see later, high-frequency synchrotron emission is also observed for which t_cool is considerably shorter than the age of a source component. The corresponding relativistic electrons can then only be generated locally. This means that the processes of particle acceleration are not confined to the inner core of an AGN, but also occur in the extended source components.

Since the characteristic frequency (5.3) of synchrotron radiation depends on a combination of the Lorentz factor γ and the magnetic field B, we cannot measure these two quantities independently. Therefore, it is difficult to estimate the magnetic field of a synchrotron source. In most cases, the (plausible) assumption of an equipartition of the energy density in the magnetic field and the relativistic particles is made, i.e., one assumes that the energy density $B^2/(8\pi)$ of the magnetic field roughly agrees with the energy density

$$\int \mathrm{d}\gamma \; n_\mathrm{e}(\gamma)\, \gamma\, m_\mathrm{e} c^2$$

of the relativistic electrons. Such approximate equipartition holds for the cosmic rays in our Galaxy and its magnetic field. Another approach is to estimate the magnetic field such that the total energy of relativistic electrons and magnetic field is minimized for a given source luminosity. The resulting value for B basically agrees with that derived from the assumption of equipartition.

5.1.4 Broad emission lines

The UV and optical spectra of quasars feature strong and very broad emission lines. Typically, lines of the Balmer series and Lyα of hydrogen, and metal lines of ions like MgII, CIII, CIV are observed[3]—these are found in virtually all quasar spectra. In addition, a large number of other emission lines occur which are not seen in every spectrum (Fig. 5.2).

To characterize the *strength* of an emission line, we define the *equivalent width* of a line W_λ as

$$W_\lambda = \int \mathrm{d}\lambda \, \frac{S_\mathrm{l}(\lambda) - S_\mathrm{c}(\lambda)}{S_\mathrm{c}(\lambda)} \approx \frac{F_\mathrm{line}}{S_\mathrm{c}(\lambda_0)} \,, \tag{5.7}$$

where $S_\mathrm{l}(\lambda)$ is the total spectral flux, and $S_\mathrm{c}(\lambda)$ is the spectral flux of the continuum radiation interpolated across the wavelength range of the line. F_line is the total flux in the line and λ_0 its wavelength. Hence, W_λ is the width of the wavelength interval over which the continuum needs to be integrated to obtain the same flux as measured in the line. Therefore, the equivalent width is a measure of the strength of a line relative to the continuum intensity.

The *width* of a line is characterized as follows: after subtracting the continuum, interpolated across the wavelength range of the line, the width is measured at half of the maximum line intensity. This width $\Delta\lambda$ is called the FWHM (full width at half maximum); it may be specified either in Å, or in km/s if the line width is interpreted as Doppler broadening, with $\Delta\lambda/\lambda_0 = \Delta v/c$. It should be noted that the width $\Delta\lambda$ and the equivalent width W_λ are very different quantities. For example, a strong narrow line can have a large W_λ, but a small $\Delta\lambda$. Conversely, a weak broad emission line can have $\Delta\lambda \gg W_\lambda$.

Broad emission lines in quasars often have a FWHM of $\sim$ 10 000 km/s, while narrower emission lines still have widths of several 100 km/s. Thus the 'narrow' emission lines are still broad compared to the typical velocities in normal galaxies.

5.1.5 Quasar demographics

Quasar surveys are always flux limited, i.e., one tries to find all quasars in a certain sky region with a flux above a predefined threshold. Only with such a selection criterion are the samples obtained of any statistical value. In addition, the selection of sources may include further criteria such as color, variability, radio or X-ray flux. For instance, radio surveys are defined by $S_\nu > S_\mathrm{lim}$ at a specific frequency. The optical identification of such radio sources reveals that quasars have a very broad redshift distribution. For decades, quasars have been the only sources known at $z > 3$. Below we will discuss different kinds of AGN surveys.

In the 1993 issue of the quasar catalog by Hewitt & Burbidge, 7236 sources are listed. This catalog contains a

[3] The ionization stages of an element are distinguished by Roman numbers. A neutral atom is denoted by 'I', a singly ionized atom by 'II', and so on. So, CIV is three times ionized carbon.

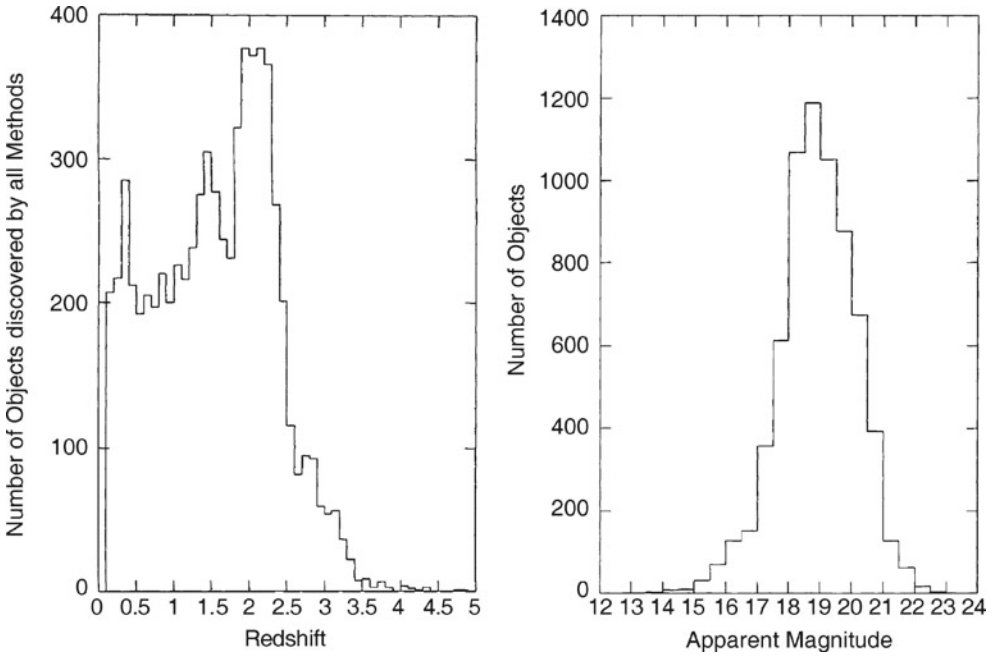

Fig. 5.10 The redshift (*left*) and brightness distribution (*right*) of QSOs in the 1993 Hewitt & Burbidge catalog. These distributions provide no proper statistical information, but they clearly show the width of the distributions. The decrease in abundances for $z \geq 2.3$ is a selection effect: many early QSO surveys started with an optical color selection, typically $U - B < -0.3$. If $z \geq 2.3$, the strong Lyα emission line moves into the B-filter and hence the quasar becomes redder in this color index and drops out of the color selection. Source: A. Hewitt & G. Burbidge 1993, *A revised and updated catalog of quasi-stellar objects*, ApJS 87, 451. ©AAS. Reproduced with permission

broad variety of different AGNs. Although it is statistically not well-defined, this catalog provides a good first indication of the width of the redshift and brightness distribution of AGNs (see Fig. 5.10).

The luminosity function of quasars extends over a very large range in luminosity, nearly three orders of magnitude in L (and over an even broader range if lower-luminosity AGNs are accounted for as well). It is steep at its bright end and has a significantly flatter slope at lower luminosities (see Sect. 5.6.2). We can compare this to the luminosity function of galaxies which is described by a Schechter function (see Sect. 3.10). While the faint end of the distribution is also described here by a relatively shallow power law, the Schechter function decreases exponentially for large L, whereas that of quasars decreases as a power law. For this reason, one finds quasars whose luminosity is much larger than the value of L^* where the break in the luminosity function occurs.

5.2 AGN zoology

Quasars are the most luminous members of the class of AGNs. Seyfert galaxies are another type of AGNs and were mentioned previously. In fact, a wide range of objects are subsumed under the name AGN, all of which have in common strong non-thermal emission in the core of a galaxy (*host galaxy*). We will mention the most important types of AGNs in this section. It is important to keep in mind that the frequency range in which sources are studied affects the source classification. We shall return to this point at the end of this section.

The classification of AGNs described below is very confusing at first glance. Different classes refer to different appearances of AGNs but do not necessarily correspond to the physical nature of these sources. A sample of optical spectra for different types of AGNs is displayed in Fig. 5.11 which illustrates a large variety of spectral properties. Similarly, the properties of the emission of AGNs in different wavebands (such as radio or gamma-rays) can differ most strongly. However, as we will discuss in Sect. 5.5, the large variety of appearances of AGNs can be understood, at least to a first approximation, by geometric considerations. The emission of an AGN is not isotropic; we will see that the flow of material which causes the energy release near the central black hole occurs in the form of a disk (the so-called accretion disk—see Sect. 5.3.2), which defines a pair of preferred directions, i.e., those perpendicular to the plane in which the disk lies. In the context of unified models, the way an AGN appears to us depends strongly on the angle between this disk axis and the line-of-sight to the source.

Outline of the unified model. In Fig. 5.12, this geometric picture of an AGN is sketched. The motivation for this model

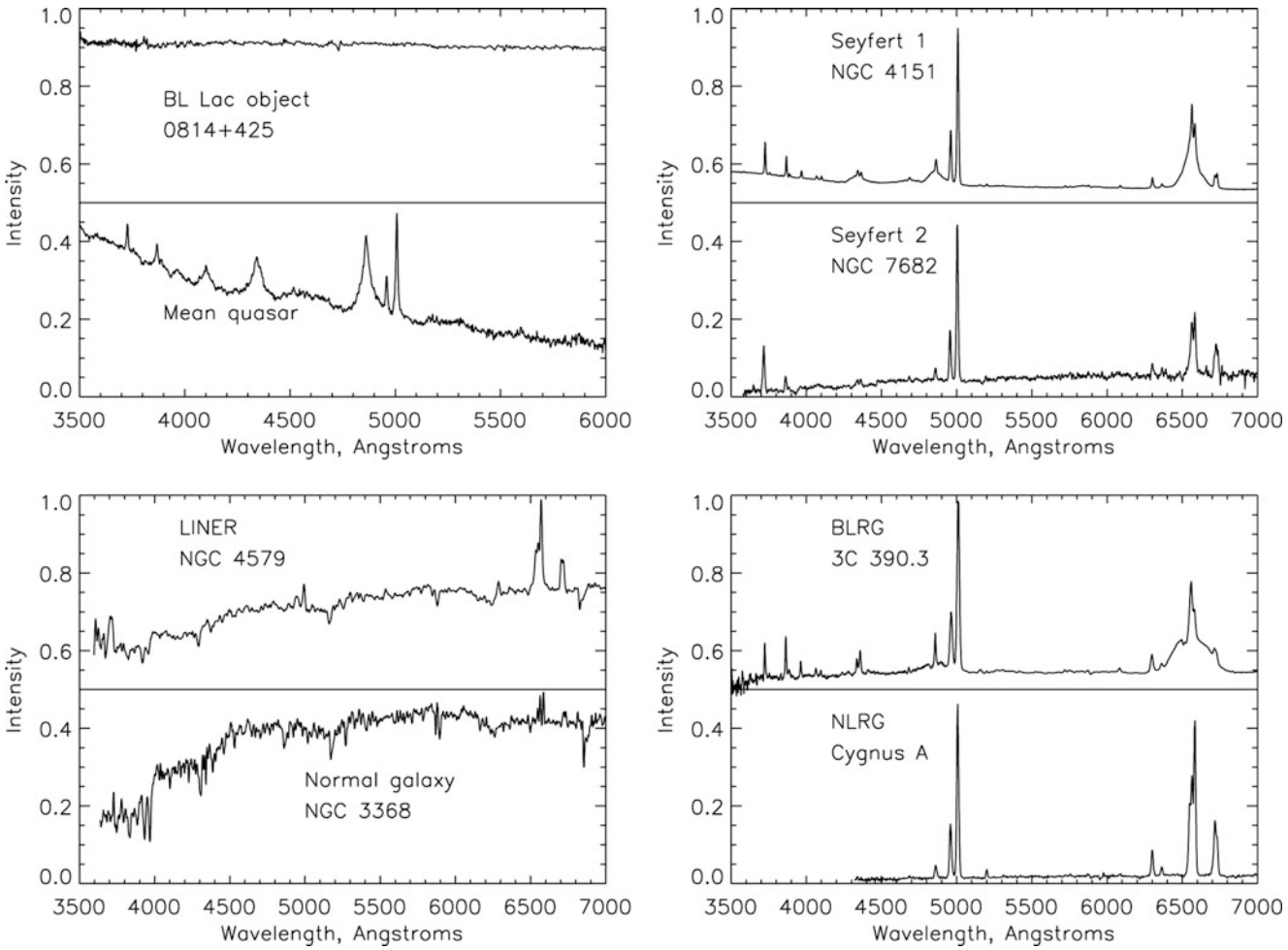

Fig. 5.11 Comparison of the optical spectra of various types of active galaxies with that of a 'normal', inactive galaxy (*bottom left*). *From top to bottom*, *left column*: a BL Lac object, the mean QSO spectrum, a LINER; in the *right column*: Seyfert 1 and 2, a Broad and a Narrow Line Radio Galaxy. Source: Bill Keel's WWW Gallery http://www.astr.ua.edu/keel/agn/spectra.html. Reproduced with permission of Bill Keel

will be explained in the course of this chapter, but we briefly summarize it here to provide a guide for the subsequent description of the various AGN classes. Surrounding the central supermassive black hole is an accretion disk which emits the bulk part of the optical and UV continuum emission. The central region around the accretion disk is the source of most of the X-ray radiation. Gas clouds above and below the accretion disk are responsible for the broad emission lines. In the plane of the disk, a distribution of gas and dust is present, which can absorb radiation from the inner region of the AGN; this obscuring material is sometimes depicted as a torus, though its geometry is probably more complicated. Nevertheless, the appearance of the AGN depends on whether the observer is located near the plane of the disk—where radiation is partly absorbed by the material in the torus—or placed in a direction closer to the axis of the disk. This concerns in particular the broad line emission, which may be fully obscured for an observer in the plane of the disk. In contrast, the gas responsible for the narrow emission lines is located at much larger distances from the black hole, so that it cannot be fully hidden by the obscuring torus.

The radio jets discussed before are launched very close to the central black hole along the direction of the disk axis. The emission from these jets is highly anisotropic, because the velocity in the inner part of the jets is close to the speed of light; then, according to the theory of Special Relativity, the jet emission is strongly beamed in the direction of jet motion. This implies that the appearance of the jet depends on how close the line-of-sight to an observer is to the jet axis. If the jet points almost directly at the observer, the jet emission can outshine all the other radiation from the AGN.

In Fig. 5.12, the different green arrows indicate different lines-of-sight to observers, and they are labeled with the characteristic AGN class the corresponding observer will see. In the upper half of the figure, it is assumed that the AGN produces strong jets, whereas in the lower part, weaker jets (or none at all) are assumed. With this picture in mind, we shall now describe the various types of AGNs.

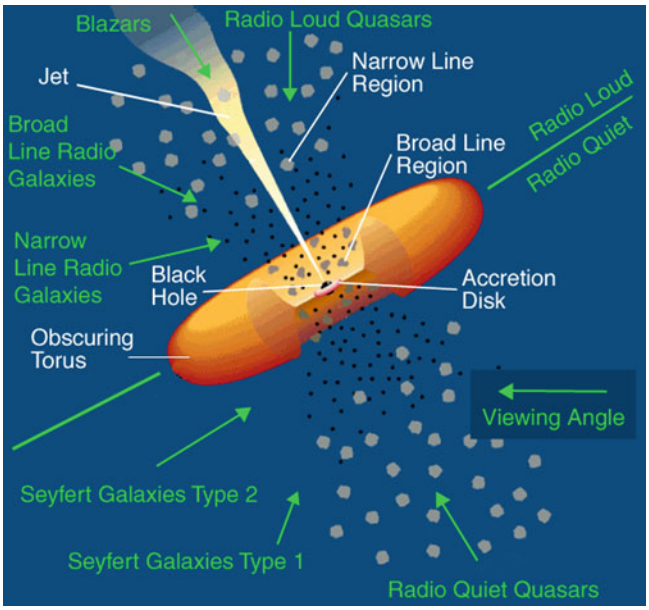

Fig. 5.12 Sketch of our current understanding of the unification of AGN types. The accretion disk is surrounded by a thick 'torus' containing dust which thus obscures the view to the center of the AGN. When looking from a direction near the plane of the disk, a direct view of the continuum source and the BLR is blocked, whereas it is directly visible from directions closer to the symmetry axis of the disk. The difference between Seyfert 1 (and BLRG) and Seyfert 2 (and NLRG) is therefore merely a matter of orientation relative to the line-of-sight. If an AGN is seen exactly along the jet axis, it appears as a blazar. Credit: NASA

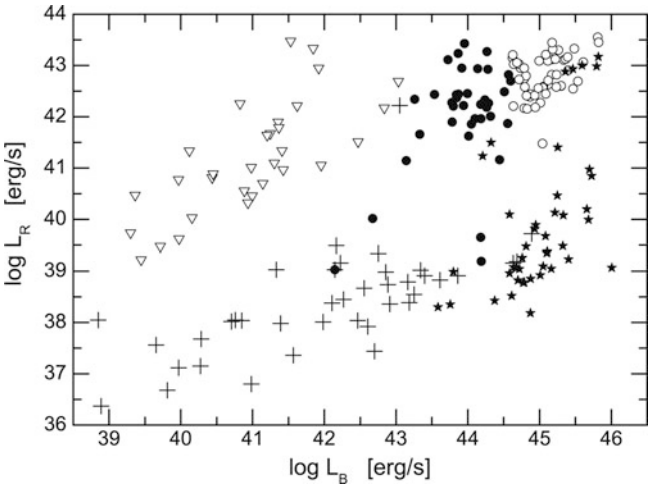

Fig. 5.13 Radio vs. optical luminosity of AGN, as measured at 5 GHz and in the B-band. Different types of AGNs are shown with *different symbols*: FR I radio galaxies (*open triangles*), Broad-Line Radio Galaxies (*filled circles*), radio-loud QSOs (*open circles*), Seyfert galaxies and LINERs (*crosses*), and a sample a $U - B$ color-selected bright QSOs, the Palomar-Green sample (*filled stars*). Apparently, the AGN population is divided into two populations, characterized by their radio-to-optical flux ratio R. Diagrams like this one suggest that there is a bimodal distribution in R, according to radio-loud and radio-quiet AGNs. Source: M. Sikora et al. 2007, *Radio Loudness of Active Galactic Nuclei: Observational Facts and Theoretical Implications*, ApJ 658, 815, p. 823, Fig. 1. ©AAS. Reproduced with permission

5.2.1 QSOs

The unusually blue color of quasars suggested the possibility of searching for them not only with radio observations but also at optical wavelengths, namely to look for point-like sources with a very blue $U - B$ color index. These photometric surveys were very successful. In fact, many more such sources were found than expected from radio counts. Most of these sources are (nearly) invisible in the radio domain of the spectrum; such sources are called radio-quiet. Their optical properties are virtually indistinguishable from those of quasars. In particular, they have a blue optical energy distribution (of course, since this was the search criterion!), strong and broad emission lines, and in general a high redshift.

Apart from their radio properties, these sources appear to be like quasars. Therefore they were called *radio-quiet quasars*, or quasi-stellar objects, QSOs. Today this terminology is no longer very common because the clear separation between sources with and without radio emission is not considered valid any more. Radio-quiet quasars also show radio emission if they are observed at sufficiently high sensitivity. In modern terminology, the expression QSO encompasses both the quasars and the radio-quiet QSOs. About 10 times more radio-quiet QSOs than quasars are thought to exist.

In fact, there is as yet not a clear consensus in whether QSOs show a bimodal distribution in their ratio of radio-to-optical luminosity. Figure 5.13 shows several different samples of AGN; in particular, optically-selected QSOs from the Palomar-Green survey (filled stars) and radio-loud QSOs (open circles). It seems that the ratio between radio and optical luminosity falls into two broad ranges, with a clear gap in between. Therefore, diagrams like that argue in favor of a bimodal distribution. However, this apparent division into two classes can at least partly be attributed to selection effects: the distribution of the radio-to-optical flux ratio depends on the selection of the QSO sample. Obviously, selecting them by their radio emission will favor those with a large $L_{\rm radio}/L_{\rm opt}$-ratio. Furthermore, the fraction of QSOs for which this ratio is large (i.e., which would be termed as radio-loud QSOs) depends on optical luminosity and on redshift: One finds a significantly higher radio-loud fraction amongst more luminous, and lower-redshift QSOs.

The QSOs are the most luminous AGNs. Their core luminosity can be as high as a thousand times that of an L^*-galaxy. Therefore they can outshine their host galaxy and appear point-like on optical images. For QSOs of lower L, their host galaxies were identified and spatially resolved with the HST (see Fig. 1.14). According to our current understanding, AGNs are the active cores of galaxies. These galaxies are supposed to be fairly normal galaxies, except for

their intense nuclear activity, and we will discuss possible reasons for the onset of this activity further below.

5.2.2 Seyfert galaxies

Seyfert galaxies are the AGNs that were detected first. Their nuclear luminosity is considerably lower than that of QSOs. On optical images they are identified as spiral galaxies which have an extraordinarily bright core (Fig. 5.5) whose spectrum shows strong emission lines which are broader than typical velocities in galaxies.

We distinguish between Seyfert galaxies of Type 1 and Type 2: Seyfert 1 galaxies have both very broad and also narrower emission lines, where 'narrow' still means several hundred km/s and thus a significantly larger width than characteristic velocities (like rotational velocities) found in normal galaxies. Seyfert 2 galaxies show only the narrower lines. Later, it was discovered that intermediate variants exist—one now speaks of Seyfert 1.5 and Seyfert 1.8 galaxies, for instance—in which very broad lines exist but with a smaller ratio of broad-to-narrow line flux than in Seyfert 1 galaxies. The classical Seyfert 1 galaxy is NGC 4151 (see Fig. 5.5), while NGC 1068 (Fig. 5.4) is a typical Seyfert 2 galaxy.

The optical spectrum of the nucleus of Seyfert 1 galaxies is very similar to that of QSOs. A smooth transition exists between (radio-quiet) QSOs and Seyfert 1 galaxies. Formally, these two classes of AGNs are separated at an absolute magnitude of $M_B = -21.5+5\log h$. The separation of Seyfert 1 galaxies and QSOs is historical since these two categories were introduced only because of the different methods of discovering them. However, except for the different core luminosity, no fundamental physical difference seems to exist. Often both classes are combined under the name Type 1 AGNs.

5.2.3 LINERs

The least luminous, and by far most common type of AGNs are the LINERs, low-ionization nuclear emission-line regions. In fact, at least one third of all nearby galaxies contain a LINER in their core, characterized by emission lines from neutral atoms or ions with rather low ionization energies. In contrast, emission from lines of strongly ionized ions is either weak or absent. Furthermore, the width of emission lines in LINERs is typically smaller than the narrow emission lines in Seyfert galaxies, and not much larger than the rotational velocity of the galaxy. However, in some LINERs one can find low-luminosity broad emission wings of the Balmer lines; these are sometimes called Type-1 LINERS.

Interestingly, spectra similar to LINERs are frequently found from low-density warm ionized gas in early-type galaxies, with the emission region being spatially extended. For these sources, it can be ruled out that the energy source of the line emission is due to a central AGN. Given the spectral similarity with the LINER emission from the center of spirals, it is sometimes questioned whether the latter phenomenon is indeed a signature of an AGN, or whether LINERs can be powered by star-formation activity, namely by post-AGB stars. In addition, if LINERs are AGNs, then there is no general consensus whether they form a distinct subclass, or whether they are the low-luminosity end of the distribution function of Seyfert galaxies.

5.2.4 Radio galaxies

Radio galaxies are elliptical galaxies with an active nucleus. They were the first sources that were identified with optical counterparts in the early radio surveys. Characteristic radio galaxies are Cygnus A (Fig. 5.6) and Centaurus A (see Fig. 5.48 below).

Similarly to Seyfert galaxies, for radio galaxies we also distinguish between those with and without broad emission lines: broad-line radio galaxies (BLRG) and narrow-line radio galaxies (NLRG), respectively. In principle, the two types of radio galaxies can be considered as radio-loud Seyfert 1 and Seyfert 2 galaxies. A smooth transition between BLRG and quasars also seems to exist, again separated by optical luminosity of the nucleus as for Seyfert galaxies.

Besides the classification of radio galaxies into BLRG and NLRG with respect to the optical spectrum, they are distinguished according to their radio morphology. As was discussed in Sect. 5.1.2, radio sources are divided into FR I and FR II sources.

5.2.5 OVVs

One subclass of QSOs is characterized by the very strong and rapid variability of its optical radiation. The flux of these sources, which are known as Optically Violently Variables (OVVs), can vary by a significant fraction on time-scales of days (see Fig. 5.14). Besides this strong variability, OVVs also stand out because of their relatively high polarization of optical light, typically a few percent, whereas the polarization of normal QSOs is below $\sim 1\,\%$. OVVs are usually strong radio emitters. Their radiation also varies in other wavelength regions besides the optical, with shorter time-scales and larger amplitudes at higher frequencies.

5.2.6 BL Lac objects

The class of AGNs called BL Lac objects (or short: BL Lacs) is named after its prototypical source BL Lacertae. They are

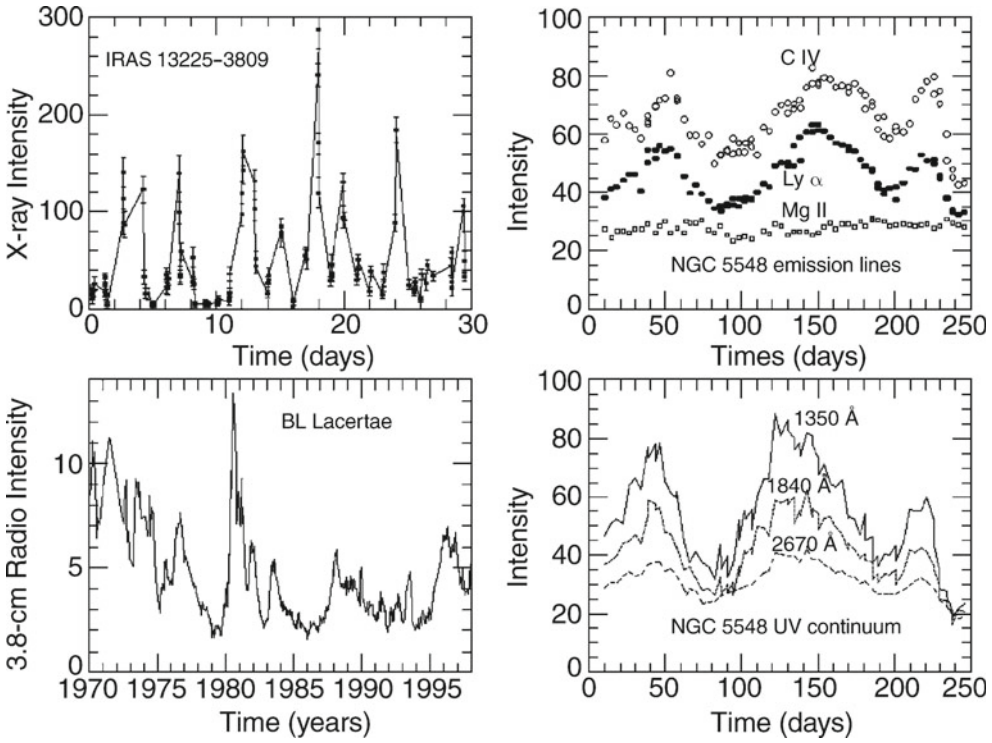

Fig. 5.14 Quasars, BL Lac objects, and Seyfert galaxies all show clear variability at many different wavelengths. In the *upper left panel*, the X-ray light curve of the Seyfert 1 galaxy IRAS 13225−3809 is plotted (observed by ROSAT); on time-scales of days, the source frequently varies by more than a factor of 20. The radio light curve of BL Lacertae at $\lambda = 3.8$ cm covering a period of 28 years is shown in the *lower left panel*. Short-term variations of such blazars are observed in a number of bursts, some overlapping (see, e.g., the burst in 1981). The UV variability of NGC 5548, a Seyfert 1 galaxy, observed by the IUE satellite is plotted for three wavelengths in the *lower right panel*. Variations at these frequencies appear to be in phase, but the amplitude becomes larger towards smaller wavelengths. Simultaneously, the line strengths of three broad emission lines of this Seyfert 1 galaxy have been measured and are plotted in the *upper right panel*. It is found that lines of high ionization potentials, like CIV, have higher variability amplitudes than those of low ionization potentials, like MgII. From the relative temporal shift in the line variability and the continuum flux, the size of the broad line region can be estimated—see Sect. 5.4.2. Credit: Webpage William C. Keel, University of Alabama

AGNs with very strongly varying radiation, like the OVVs, but without strong emission and absorption lines. As for OVVs, the optical radiation of BL Lacs is highly polarized. Since no emission lines are observed in the spectra of BL Lacs, the determination of their redshift is often difficult and sometimes impossible. In some cases, absorption lines are detected in the spectrum which are presumed to derive from the host galaxy of the AGN and are then identified with the redshift of the BL Lac.

The optical luminosity of some BL Lacs varies by several magnitudes if observed over a sufficiently long time period. Particularly remarkable is the fact that in epochs of low luminosity, emission lines are sometimes observed and then a BL Lac appears like an OVV. For this reason, OVVs and BL Lacs are collectively called *blazars*. All known blazars are radio sources. Besides the violent variability, blazars also show highly energetic and strongly variable γ-radiation (Fig. 5.15). Table 5.1 summarizes the fundamental properties of the different classes of AGNs.

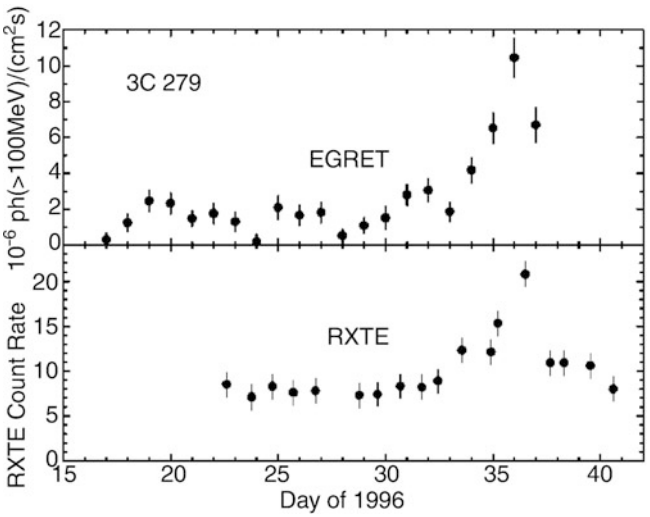

Fig. 5.15 Variability of the blazar 3C279 in X-ray (*bottom*) and in γ-radiation at photon energies above 100 MeV (*top*). On time-scales of a few days, the luminosity varies by a factor ~ 10. Credit: EGRET Team; NASA; research article: R.C. Hartman et al. 2001, ApJ 558, 583

Table 5.1 Overview of the classification of active galactic nuclei

	Normal galaxy	Radio galaxy	Seyfert galaxy	Quasar	Blazar
Example	Milky Way	M87, Cygnus A	NGC 4151	3C273	BL Lac, 3C279
Galaxy type	Spiral	Elliptical, Irregular	Spiral	Irregular	Elliptical?
$L_{\rm AGN}/L_\odot$	$< 10^4$	10^6–10^8	10^8–10^{11}	10^{11}–10^{14}	10^{11}–10^{14}
$M_{\rm BH}/M_\odot$	4×10^6	3×10^9	10^6–10^9	10^6–10^9	10^6–10^9
Radio emission	Weak	Core, jets, lobes	Only ≈5 % radio-loud	Only ≈5 % radio-loud	Strong, Short-time variable
X-ray emission	Weak	Strong	Strong	Strong	Strong
Gamma emission	Weak	Weak	Medium	Strong	Strong

5.3 The central engine: a black hole

We have previously mentioned that the energy production in AGNs must be related to a supermassive black hole (SMBH) in its center. We will present arguments for this conclusion in this section. To do this, we will first summarize some of the relevant observational facts for AGNs.

- The extent of some radio sources in AGNs may reach $\gtrsim$1 Mpc. From this length-scale a minimum lifetime for the activity in the nucleus of these objects can be derived, since even if the radio source expands outwards from the core with the speed of light, the age of such a source would be $\tau \gtrsim 10^7$ yr.
- Luminous QSOs have a luminosity of up to $L_{\rm bol} \sim 10^{47}$ erg/s. Assuming that the luminosity does not change substantially over the lifetime of the source, a total energy can be estimated from the luminosity and the minimum age,

$$E \gtrsim 10^{47}\,\mathrm{erg/s} \times 10^7\,\mathrm{yr} \sim 3 \times 10^{61}\,\mathrm{erg}\;; \qquad (5.8)$$

  however, the assumption of an essentially constant luminosity is not necessarily justified.
- The luminosity of some AGNs varies by more than 50 % on time-scales of a day. From this variability time-scale, an upper limit for the spatial extent of the source can be determined, because the source luminosity can change substantially only on such time-scales where the source as a whole, or at least a major part of the emitting region, is in causal contact. Otherwise 'one end' of the source does not know that the 'other end' is about to vary. This yields a characteristic extent of the central source of $R \lesssim$ 1 lightday $\sim 3 \times 10^{15}$ cm.

5.3.1 Why a black hole?

We will now combine the aforementioned observations and derive from them that the basic energy production in AGNs has to be of a gravitational nature. To do this, we note that the most efficient 'classical' method of energy production is nuclear fusion, as is taking place in stars. We will therefore make the provisional assumption (which will soon lead to a contradiction) that the energy production in AGNs is based on thermonuclear processes.

By burning hydrogen into iron—the nucleus with the highest binding energy per nucleon—8 MeV/nucleon are released, or $0.008\,m_{\rm p}c^2$ per nucleon. The maximumefficiency of nuclear fusion is therefore $\epsilon \lesssim 0.8\,\%$, where ϵ is defined as the mass fraction of 'fuel' that is converted into energy, according to

$$\boxed{E = \epsilon\, mc^2}\,. \qquad (5.9)$$

To generate the energy of $E = 3 \times 10^{61}$ erg by nuclear fusion, a total mass m of fuel would be needed, where m is given by

$$m = \frac{E}{\epsilon c^2} \sim 4 \times 10^{42}\,\mathrm{g} \sim 2 \times 10^9 M_\odot\,, \qquad (5.10)$$

where we used the energy estimate from (5.8). If the energy of an AGN was produced by nuclear fusion, burnt-out matter of mass m [more precisely, $(1-\epsilon)m$] must be present in the core of the AGN.

However, the Schwarzschild radius of this mass is (see Sect. 3.8.1)

$$r_{\rm S} = \frac{2Gm}{c^2} = \frac{2GM_\odot}{c^2}\frac{m}{M_\odot}$$
$$= 3 \times 10^5\,\mathrm{cm}\,\frac{m}{M_\odot} \sim 6 \times 10^{14}\,\mathrm{cm}\,,$$

i.e., the Schwarzschild radius of the 'nuclear cinder' is of the same order of magnitude as the above estimate of the extent of the central source. This argument demonstrates that the gravitational binding energy of the cinder is far higher than the energy released from nuclear burning. Hence, gravitational effects *must* play a crucial role—the assumption

of thermonuclear energy as prime energy source has been disproven because its efficiency ϵ is too low. The only known mechanism yielding larger ϵ is gravitational energy production.

Through the infall of matter onto a central black hole, potential energy is converted into kinetic energy. If it is possible to convert part of this inward-directed kinetic energy into internal energy (heat) and subsequently emit this in the form of radiation, ϵ can be larger than that of thermonuclear processes. From the theory of accretion onto black holes, a maximum efficiency of $\epsilon \sim 6\,\%$ for accretion onto a non-rotating black hole (also called a Schwarzschild hole) is derived. A black hole with the maximum allowed angular momentum can have an efficiency of $\epsilon \sim 29\,\%$.

5.3.2 Accretion

Due to its broad astrophysical relevance beyond the context of AGNs, we will consider the accretion process in somewhat more detail.

The principle of accretion. Gas falling onto a compact object loses its potential energy, which is first converted into kinetic energy. If the infall is not prevented, the gas will fall into the black hole without being able to radiate this energy. In general one can expect that the gas has finite angular momentum. Thus it cannot fall straight onto the compact object, since this is prevented by the angular momentum barrier. Through friction with other gas particles and by the resulting momentum transfer, the gas will assemble in a disk oriented perpendicular to the direction of the angular momentum vector. The frictional forces in the gas are expected to be much smaller than the gravitational force. Hence the disk will locally rotate with approximately the Kepler velocity. Since a Kepler disk rotates differentially, in the sense that the angular velocity depends on radius, the gas in the disk will be heated by internal friction. In addition, the same friction causes a slight deceleration of the rotational velocity, whereby the gas will slowly move inwards. The energy source for heating the gas in the disk is provided by this inward motion—namely the conversion of potential energy into kinetic energy, which is then converted into internal energy (heat) by friction.

According to the virial theorem, half of the potential energy released is converted into kinetic energy; in the situation considered here, this is the rotational energy of the disk. The other half of the potential energy can be converted into internal energy. We now present an approximately quantitative description of this process, specifically for accretion onto a black hole.

Temperature profile of a geometrically thin, optically thick accretion disk. When a mass m falls from radius $r + \Delta r$ to r, the energy

$$\Delta E = \frac{GM_\bullet m}{r} - \frac{GM_\bullet m}{r+\Delta r} \approx \frac{GM_\bullet m}{r}\,\frac{\Delta r}{r}$$

is released. Here $M_\bullet$ denotes the mass of the SMBH, assumed to dominate the gravitational potential, so that self-gravity of the disk can be neglected. Half of this energy is converted into heat, $E_{\rm heat} = \Delta E/2$. If we assume that this energy is emitted locally, the corresponding luminosity is

$$\Delta L = \frac{GM_\bullet \dot{m}}{2r^2}\Delta r\ , \tag{5.11}$$

where $\dot{m}$ denotes the accretion rate, which is the mass that falls into the black hole per unit time. In the stationary case, $\dot{m}$ is independent of radius since otherwise matter would accumulate at some radii. Hence the same amount of matter per unit time flows through any cylindrical radius.

If the disk is optically thick, the local emission corresponds to that of a black body. The ring between r and $r+\Delta r$ then emits a luminosity

$$\Delta L = 2 \times 2\pi r\,\Delta r\,\sigma_{\rm SB} T^4(r)\ , \tag{5.12}$$

where the factor 2 originates from the fact that the disk has two sides. Combining (5.11) and (5.12) yields the radial dependence of the disk temperature,

$$T(r) = \left(\frac{GM_\bullet \dot{m}}{8\pi\sigma_{\rm SB} r^3}\right)^{1/4}\ .$$

A more accurate derivation explicitly considers the dissipation by friction and accounts for the fact that part of the generated energy is used for heating the gas, where the corresponding thermal energy is also partially advected inwards. Except for a numerical correction factor, the same result is obtained,

$$\boxed{T(r) = \left(\frac{3GM_\bullet \dot{m}}{8\pi\sigma_{\rm SB} r^3}\right)^{1/4}}\ , \tag{5.13}$$

which is valid in the range $r \gg r_{\rm S}$. Scaling r with the Schwarzschild radius $r_{\rm S}$, we obtain

$$T(r) = \left(\frac{3GM_\bullet \dot{m}}{8\pi\sigma_{\rm SB} r_{\rm S}^3}\right)^{1/4} \left(\frac{r}{r_{\rm S}}\right)^{-3/4}\ .$$

By replacing $r_{\rm S}$ with (3.43) in the first factor, this can be written as

$$\boxed{T(r) = \left(\frac{3c^6}{64\pi\sigma_{\rm SB}G^2}\right)^{1/4} \dot{m}^{1/4}\, M_\bullet^{-1/2} \left(\frac{r}{r_{\rm S}}\right)^{-3/4}}\,. \tag{5.14}$$

Interpretation and conclusions. From this analysis, we can immediately draw a number of conclusions. The most surprising one may be the independence of the temperature profile of the disk from the detailed mechanism of the dissipation because the equations do not explicitly contain the viscosity. This fact allows us to obtain quantitative predictions based on the model of a *geometrically thin, optically thick accretion disk*.[4] The temperature in the disk increases inwards $\propto r^{-3/4}$, as expected. Therefore, the total emission of the disk is, to a first approximation, a superposition of black bodies consisting of rings with different radii at different temperatures. For this reason, the resulting spectrum does not have a Planck shape but instead shows a much broader energy distribution. Over a wide range of frequencies, the resulting spectrum from such an optically thick accretion disk is fairly flat, where the lower and upper bound of the frequency interval is determined by the lowest and highest temperature (at the outer and inner radius) of the disk.

Most of the luminosity from a disk comes from the inner parts, and thus depends critically on how far the disk extends inside. Around a black hole, there is a minimum radius $r_{\rm in}$ at which stable circular orbits can exist. For a black hole without rotation, this innermost stable orbit is at $r_{\rm in} = 3r_{\rm S}$, whereas it is smaller for a black hole with angular momentum. Accordingly, the efficiency

$$\epsilon = \frac{L}{\dot{m}c^2} \tag{5.15}$$

with which accreting mass is converted into luminosity depends on the black hole spin. It increases from $\sim 6\,\%$ for a non-rotating black hole to $\sim 29\,\%$ for one with maximum rotation.

For any fixed ratio $r/r_{\rm S}$, the temperature increases with the accretion rate. This again was expected: since the local emission is $\propto T^4$ and the locally dissipated energy is $\propto \dot{m}$, it must be $T \propto \dot{m}^{1/4}$. Furthermore, at fixed ratio $r/r_{\rm S}$, the temperature decreases with increasing mass $M_\bullet$ of the black hole. This implies that the maximum temperature attained in the disk is lower for more massive black holes. This may be unexpected, but it is explained by a decrease of the tidal forces, at fixed $r/r_{\rm S}$, with increasing $M_\bullet$. In particular, it implies that the maximum temperature of the disk in an AGN is much lower than in accretion disks around compact objects of stellar mass. Accretion disks around neutron stars and stellar-mass black holes emit in the hard X-ray part of the spectrum and are known as X-ray binaries. In contrast, the thermal radiation of the disk in an AGN extends to the UV or soft X-ray range only (see below).

Radiatively inefficient accretion. The disk accretion described above requires that the generated energy is emitted locally, which requires the disk to be optically thick. The optical depth of the disk depends on its surface density, which in turn depends on the accretion rate. In a system where the accretion rate $\dot{m}$ is low (in a sense quantified further below), the disk may be optically thin, and the emission process of the heated gas can become inefficient. In this case, the gas cannot efficiently cool, and the thermal energy generated by friction in the disk is advected inwards together with the gas. Such a disk (called 'advection-dominated accretion flow', or ADAF) is rather inefficient in converting rest mass into radiation, and so its corresponding ϵ can be quite small. However, one expects that such an accretion flow may be quite efficient in generating outflows, such that part of the accreted material is ejected in form of jets. Hence, this mode of accretion may play an important role for radio galaxies.

Bondi–Hoyle–Lyttleton accretion. In the absence of radiation pressure, the mean accretion rate is determined by the flux of matter that is added to the outer parts of the accretion disk. This quantity is difficult to estimate and depends on the rate with which gas in a galaxy can be transported inwards. Owing to the angular momentum of the gas, it presumably moves to the central region only through significant perturbations of the gravitational potential from axisymmetry. Details of these processes are not fully understood yet.

However, there is one simple situation where the accretion rate can be estimated analytically, namely the case of spherical accretion from a static medium. Assume a black hole being immersed into a spherically-symmetric gas distribution which for large radii is homogeneous with density ρ_∞ and sound speed $c_{\rm s}$. The gravitational pull by the black hole causes the gas to have an inward-directed velocity. Provided the gas is adiabatic, then the mass accretion rate can be calculated from the equations of fluid dynamics, yielding

[4] The physical mechanism that is responsible for the viscosity is unknown. The molecular viscosity is far too small to be considered as the primary process. Rather, the viscosity is probably produced by turbulent flows in the disk or by magnetic fields, which become spun up by differential rotation and thus amplified, so that these fields may act as an effective friction. In addition, hydrodynamic instabilities may act as a source of viscosity. Although the properties of the accretion disk presented here—luminosity and temperature profile—are independent of the specific mechanism of the viscosity, other disk properties definitely depend on it. For example, the temporal behavior of a disk in the presence of a perturbation, which is responsible for the variability in some binary systems, depends on the magnitude of the viscosity, which therefore can be estimated from observations of such systems.

$$\dot{m} = \frac{4\pi G^2 M_\bullet^2}{c_s^3}\rho_\infty \ . \tag{5.16}$$

This *Bondi–Hoyle–Lyttleton accretion rate* yields an indication of the mass influx onto the accretion disk, provided the angular momentum of the surrounding gas is sufficiently small. In this case, it can flow in at the rate given by (5.16), until it reaches a radius where the angular momentum becomes important and the gas is forced onto circular orbits, forming an accretion disk. Purely spherical accretion, i.e., where the gas has zero angular momentum and no disk is formed, is very inefficient; only a tiny fraction of the kinetic energy gets dissipated and radiated away.

5.3.3 Superluminal motion

Apparent velocities larger than c. Besides the generation of energy, another piece of evidence for the existence of SMBHs in the centers of AGNs results from observing relative motions of source components at *superluminal* velocities. These observations of central radio components in AGNs are mainly made using VLBI methods since they provide the highest available angular resolution. They measure a time dependence of the angular separation of source components, which often leads to values $> c$ if the angular velocity is translated into a transverse spatial velocity (Fig. 5.16). These superluminal motions caused some discomfort upon their discovery. In particular, they at first raised concerns that the redshift of QSOs may not originate from cosmic expansion. Only if the QSO redshifts are interpreted as being of cosmological origin can they be translated into a distance, which is needed to convert the observed angular velocity into a spatial velocity.

We consider two source components (e.g., the radio core and a component in the jet) which are observed to have a time-dependent angular separation $\theta(t)$. If D denotes the distance of the source, then the apparent relative transverse velocity of the two components is

$$v_{\rm app} = \frac{\mathrm{d}r}{\mathrm{d}t} = D\,\frac{\mathrm{d}\theta}{\mathrm{d}t} \ , \tag{5.17}$$

where $r = D\theta$ is the transverse separation of the two components. The final expression in (5.17) shows that $v_{\rm app}$ is directly observable if the distance D is assumed to be known.

Frequently, VLBI observations of compact radio sources yield values for $v_{\rm app}$ that are larger than c! Characteristic values for sources with a dominant core component are $v_{\rm app} \sim 5c$ (see Fig. 5.16). But according to the theory of Special Relativity, velocities $> c$ do not exist. Thus it is

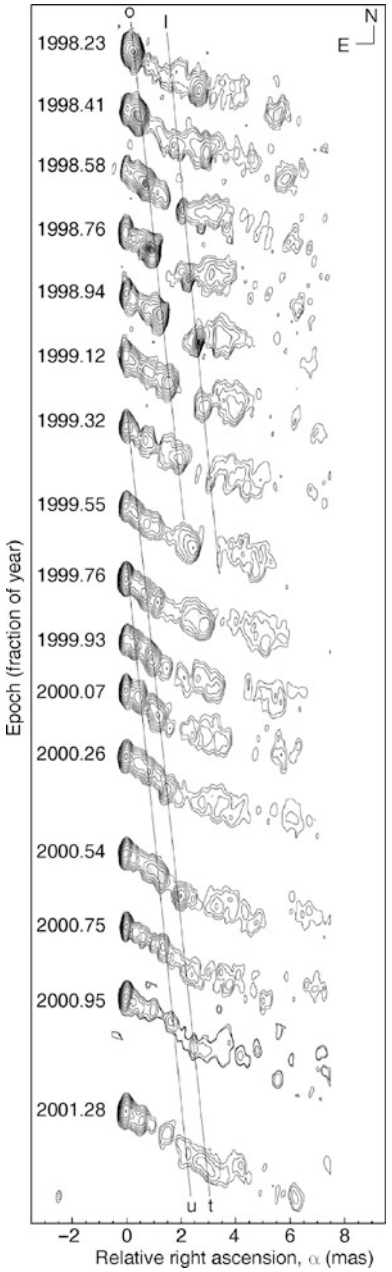

Fig. 5.16 Apparent superluminal velocities of source components in the radio jet of the source 3C120. VLBA observations of this source are presented for 16 different epochs (indicated by the *numbers* at the left of the corresponding radio map), observed at 7 mm wavelength. The ellipse at the *lower left* indicates the beam of the VLBA interferometer and thus the angular resolution of these observations. At the distance of 3C120 of 140 Mpc, a milliarcsecond corresponds to a linear scale of 0.70 pc. The four *straight lines*, denoted by l, o, t, and u, connect the same source components at different epochs. The linear motion of these components is clearly visible. The observed angular velocities of the components yield apparent transverse velocities in the range of $4.1c$ to $5c$. Source: A.P. Marscher et al. 2002, *Observational evidence for the accretion-disk origin for a radio jet in an active galaxy*, Nature 417, 625, Fig. 1

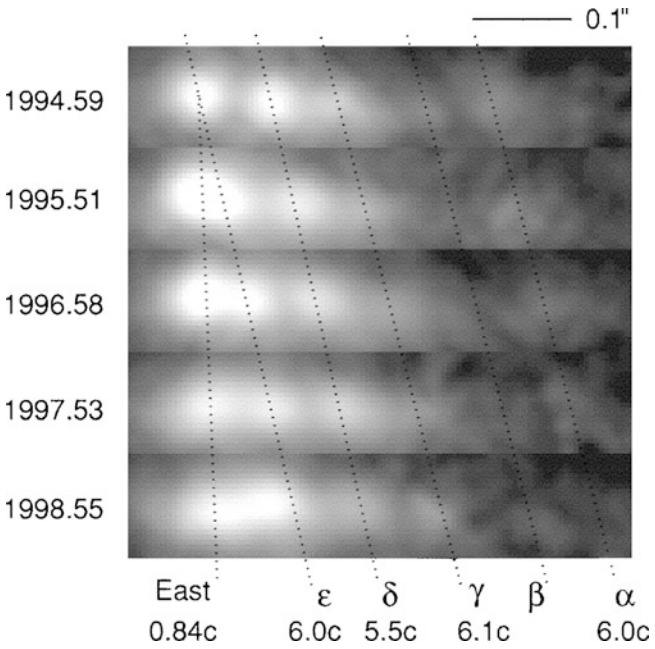

Fig. 5.17 Also at optical wavelengths, apparent superluminal motion was observed. The figure shows the optical jet in M87, based on HST images taken over a period of about 4 years. The angular velocity of the components is up to 23 mas/yr. Assuming a distance of M87 of $D = 16$ Mpc, velocities of up to $\sim 6c$ are obtained for the components. Source: J.A. Biretta et al. 1999, *Hubble Space Telescope Observations of Superluminal Motion in the M87 Jet*, ApJ 520, 621, p. 623, Fig. 2. ©AAS. Reproduced with permission

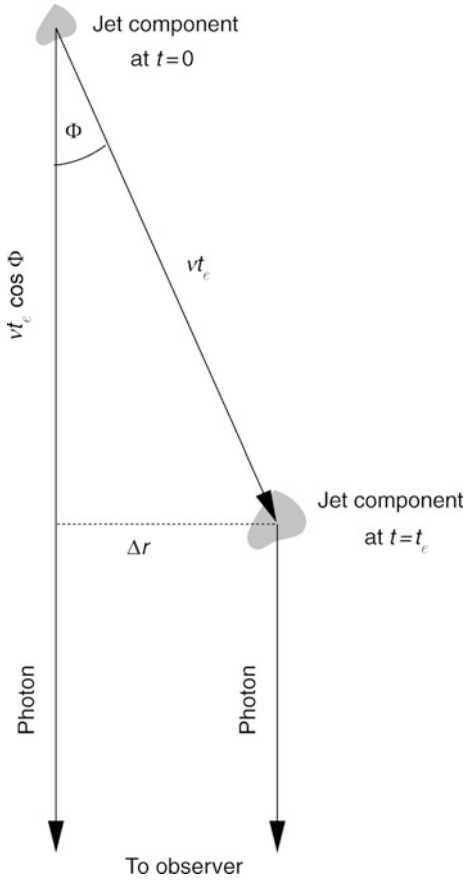

Fig. 5.18 Explanation of superluminal motion: a source component is moving at velocity v and at an angle ϕ relative to the line-of-sight. We consider the emission of photons at two different times $t = 0$ and $t = t_e$. Photons emitted at $t = t_e$ will reach us by $\Delta t = t_e(1-\beta\cos\phi)$ later than those emitted at $t = 0$. The apparent separation of the two source components then is $\Delta r = v t_e \sin\phi$, yielding an apparent velocity on the sky of $v_{app} = \Delta r/\Delta t = v\sin\phi/(1-\beta\cos\phi)$. Adapted from: B.W. Carroll & D.A. Ostlie 1996, *An introduction to Modern Astrophysics*, Reading

not surprising that the phenomenon of superluminal motion engendered various kinds of explanations upon its discovery. By now, superluminal motion has also been seen in optical observations of jets, as is displayed in Fig. 5.17.

For some time, one possibility that had been considered was that the cosmological interpretation of the redshifts may be wrong, because for a sufficiently small D velocities smaller than the speed of light would result from (5.17). However, no plausible alternative explanations for the observed redshifts of QSOs exist, and more than 40 years of QSO observations have consistently confirmed that redshift is an excellent measure for their distances—see Sect. 4.5.1.

However, Relativity only demands that no *signal* may propagate with velocities $> c$. It is easy to construct a thought-experiment in which superluminal velocities occur. For instance, consider a laser beam or a flashlight that is rotating perpendicular to its axis of symmetry. The corresponding light point on a screen changes its position with a speed proportional to the angular velocity and to the distance of the screen from the light source. If we make the latter sufficiently large, it is 'easy' to obtain a superluminal light point on the screen. But this light point does not carry a signal along its track. Therefore, the superluminal motions in compact radio sources may be explained by such a screen effect, but what is the screen and what is the laser beam? We point out that apparent superluminal velocities are seen in the center of the Milky Way in form of X-ray echos—see Sect. 2.6.5—a phenomenon where a screen (of scattering material) explains the effect.

Explanation of superluminal motion. The generally accepted explanation of apparent superluminal motion combines very fast motions of source components with the finite speed of light. For this, we consider a source component moving at speed v at an angle ϕ with respect to the line-of-sight (see Fig. 5.18). We arbitrarily choose the origin of time $t = 0$ to be the time at which the moving component is close to the core component. At time $t = t_e$, the source has a distance $v\,t_e$ from the original position. The observed separation is the transverse component of this distance,

$$\Delta r = v\,t_e \sin\phi \ .$$

Since at time t_e the source has a smaller distance from Earth than at $t = 0$, the light will accordingly take slightly less time to reach us. Photons emitted at times $t = 0$ and $t = t_e$ will reach us with a time difference of

$$\Delta t = t_e - \frac{v\, t_e \cos\phi}{c} = t_e\,(1 - \beta\cos\phi) \ ,$$

where we define

$$\beta := \frac{v}{c} \tag{5.18}$$

as the velocity in units of the speed of light. Equation (5.17) then yields the apparent velocity,

$$\boxed{v_{app} = \frac{\Delta r}{\Delta t} = \frac{v\sin\phi}{1 - \beta\cos\phi}} \ . \tag{5.19}$$

We can directly draw some conclusions from this equation. The apparent velocity v_{app} is a function of the direction of motion relative to the line-of-sight and of the true velocity of the component. For a given value of v, the maximum velocity v_{app} is obtained if

$$(\sin\phi)_{max} = \frac{1}{\gamma} \ , \tag{5.20}$$

where the *Lorentz factor* $\gamma = (1 - \beta^2)^{-1/2}$ was already defined in (5.4). The corresponding value for the maximum apparent velocity is then

$$\left(v_{app}\right)_{max} = \gamma\, v \ . \tag{5.21}$$

Since γ may become arbitrarily large for values of $v \to c$, the apparent velocity can be much larger than c, even if the true velocity v is—as required by Special Relativity—smaller than c. In Fig. 5.19, v_{app} is plotted as a function of ϕ for different values of the Lorentz factor γ. To get $v_{app} > c$ for an angle ϕ, we need

$$\beta > \frac{1}{\sin\phi + \cos\phi} \geq \frac{1}{\sqrt{2}} \approx 0.707 \ .$$

Hence, superluminal motion is a consequence of the finiteness of the speed of light. Its occurrence implies that source components in the radio jets of AGNs are accelerated to velocities close to the speed of light.

In various astrophysical situations we find that the outflow speeds are of the same order as the escape velocities from the corresponding sources. Examples are the Solar wind, stellar winds in general, or the jets of neutron stars, such as in the famous example of SS433 (in which the jet velocity is $0.26\,c$). Therefore, if the outflow velocity of the jets in AGNs is close c, the jets should originate in a region

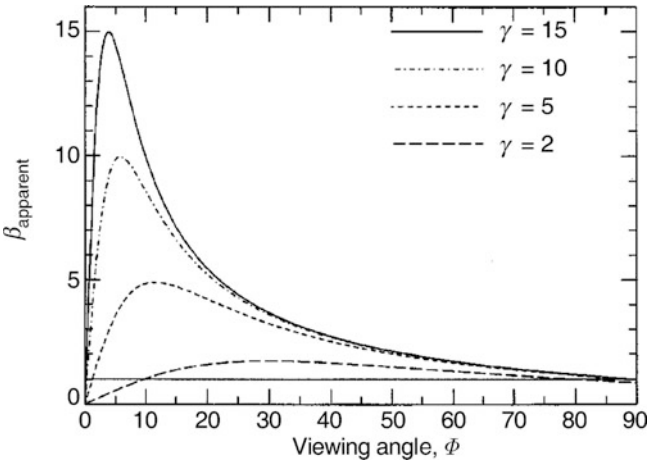

Fig. 5.19 Apparent velocity $\beta_{app} = v_{app}/c$ of a source component moving with Lorentz factor γ at an angle ϕ with respect to the line-of-sight, for four different values of γ. Over a wide range in θ, $\beta_{app} > 1$, thus apparent superluminal motion occurs. The maximum values for β_{app} are obtained if $\sin\theta = 1/\gamma$. Source: C.M. Urry & P. Padovani 1995, *Unified Schemes for Radio-Loud Active Galactic Nuclei*, PASP 107, 803, p. 839, Fig. 21. ©ASP. Reproduced with permission

where the escape velocity has a comparable value. The only objects compact enough to be plausible candidates for this are neutron stars and black holes. And since the central mass in AGNs is considerably larger than the maximum mass of a neutron star, a SMBH is the only option left for the central object. This argument, in addition, yields the conclusion that jets in AGNs must be formed and accelerated very close to the Schwarzschild radius of the SMBH.

The processes that lead to the formation of jets are still subject to intensive research. Most likely magnetic fields play a central role. Such fields may be anchored in the accretion disk, and then spun up and thereby amplified. The wound-up field lines may then act as a kind of spring, accelerating plasma outwards along the rotation axis of the disk (see Sect. 5.5.2 below). In addition, it is possible that rotational energy is extracted from a rotating black hole, a process in which magnetic fields again play a key role. As is always the case in astrophysics, detailed predictions in situations where magnetic fields dominate the dynamics of a system (like, e.g., in star formation) are extremely difficult to obtain because the corresponding coupled equations for the plasma and the magnetic field are very hard to solve.

5.3.4 Further arguments for SMBHs

A black hole is not only the simplest solution of the equations of Einstein's General Relativity, it is also the natural final state of a very compact mass distribution. The occurrence of SMBHs is thus highly plausible from a theoretical point of view. The evidence for the existence of SMBHs in the

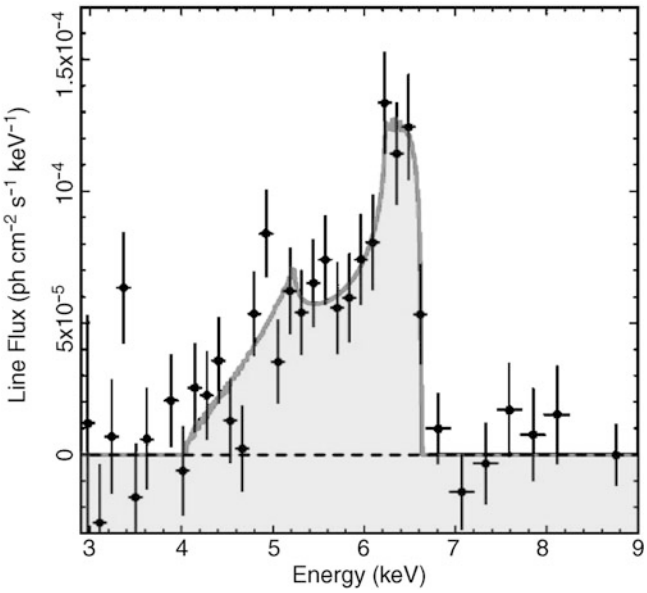

Fig. 5.20 The spectral form of the broad iron line in the Seyfert 1 galaxy MCG-6-30-15 as observed with the ASCA satellite. If the material emitting the line were at rest we would observe a narrow line at $h_P\nu = 6.35\,\text{keV}$. We see that the line is (a) broad, (b) strongly asymmetric, and (c) shifted to smaller energies. A model for the shape of the line, based on a disk around a black hole that is emitting in the radius range $r_S \leq r \leq 20r_S$, is sketched in Fig. 5.21. Source: A.C. Fabian et al. 2000, *Broad Iron Lines in Active Galactic Nuclei*, PASP 112, 1145, Fig. 6. ©ASP. Reproduced with permission

center of galaxies that has been detected in recent years (see Sect. 3.8) provides an additional argument for the presence of SMBHs in AGNs.

Furthermore, we find that the direction of the jets on a milliarcsecond scale, as observed by VLBI, is essentially identical to the direction of jets on much larger scales and to the direction of the corresponding radio lobes. These lobes often have a huge distance from the core, indicating a long lifetime of the source. Hence, the central engine must have some long-term memory because the outflow direction is stable over $\sim 10^7$ yr. A rotating SMBH is an ideal gyroscope, with a direction being defined by its angular momentum vector.

X-ray observations of an iron line of rest energy $h_P\nu = 6.35\,\text{keV}$ in Seyfert galaxies indicates that the emission must be produced in the inner region of an accretion disk, within only a few Schwarzschild radii of a SMBH. An example for this is given in Fig. 5.20. The shape of the line can be explained by a combination of a strong Doppler effect due to high rotation velocities in the disk and by the strong gravitational field of the black hole, as is illustrated in Fig. 5.21.

This iron line is not only detected in individual AGNs, but also in the average spectrum of an ensemble of AGNs. In a deep ($\sim 7.7 \times 10^5$ s) XMM-Newton exposure of the Lockman hole, a region of very low column density of Galactic hydrogen, a large number of AGNs were identified and spectroscopically verified. The X-ray spectrum of these AGNs in the energy ranges of 0.2–3 keV and of 8–20 keV (each in the AGN rest-frame) was modeled by a power law plus intrinsic absorption. The ratio of the measured spectrum of each individual AGN and the fitted model spectrum was then averaged over the AGN population, after transforming the spectra into the rest-frame of the individual sources. As shown in Fig. 5.22, this ratio clearly shows the presence of a strong and broad emission line. The shape of this average emission line can be very well modeled by emission from an accretion disk around a black hole where the radiation originates from a region lying between $\sim$ 3 and $\sim$ 400 Schwarzschild radii. The strength of the iron line indicates a high metallicity of the gas in these AGNs.

The spin of black holes. The spectral shape of the line is affected by the spin of the SMBH. General Relativity predicts that the geometrical properties of space-time around a black hole are determined by its mass and its spin, which affects the properties of the accretion disk in its innermost part as well as the propagation of light rays around the black hole. Furthermore, according to General Relativity, there is a maximum spin a black hole can have. The ratio of the black hole spin to its maximally possible value is called the spin parameter a_{spin}. With sufficiently well-observed spectra, the spin parameter can in fact be estimated, using the model indicated in Fig. 5.21. It is found that a large fraction of SMBHs have a spin parameter $a_{\text{spin}} \gtrsim 0.9$. One would expect this result if the SMBH attained most of its mass through accretion events with almost constant orientation, since accreting matter transfers, beside mass, also angular momentum to the black hole, thus spinning it up. On the other hand, if the mass growth occurred predominantly through merger processes of black holes during the merging of galaxies (see Chap. 10), then smaller values of a_{spin} are expected to result.

5.3.5 A first mass estimate for the SMBH: the Eddington luminosity

Radiation force. As we have seen, the primary energy production in AGNs occurs through accretion of matter onto a SMBH, where the largest part of the energy is produced in the innermost region, close to the Schwarzschild radius. The energy produced in the central region then propagates outwards and can interact with infalling matter by absorption or scattering. Through this interaction of outward-directed radiation with matter, the momentum of the radiation is transferred to the matter, i.e., the infalling matter experiences an outwards-directed radiation force. In order for matter to fall onto the SMBH at all, this radiation force needs to be smaller than the gravitational force. This condition can be

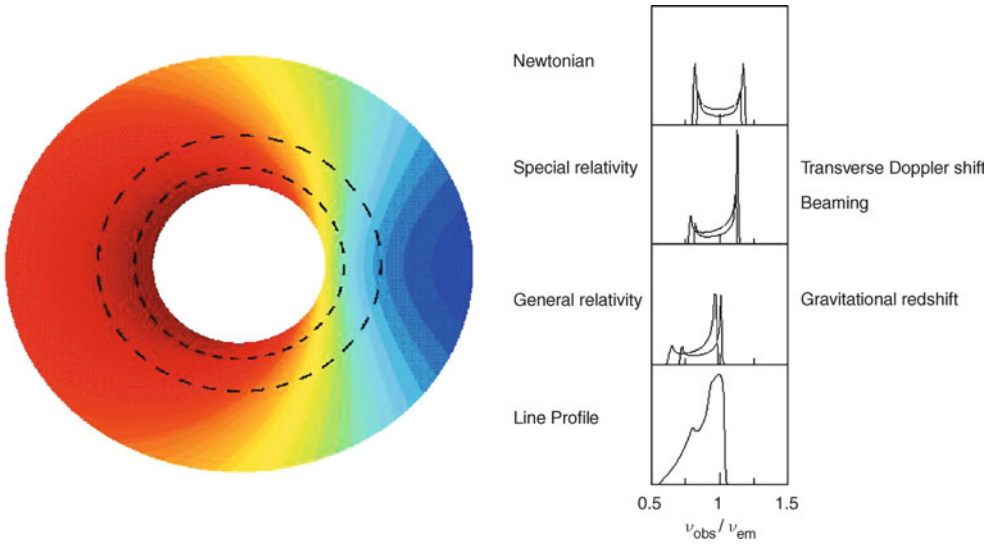

Fig. 5.21 The profile of the broad iron line is caused by a combination of Doppler shift, relativistic beaming, and gravitational redshift. *On the left*, the observed energy of the line as a function of position on a rotating disk is indicated by *colors*. Here, the energy in the right part of the disk which is moving towards us is blueshifted, whereas the left part of the disk emits redshifted radiation. Besides this Doppler effect, all radiation is redshifted because the photons must escape from the deep potential well. The smaller the radius of the emitting region, the larger this gravitational redshift. The line profile we would obtain from a ring-shaped section of the disk (*dashed ellipses*) is plotted in the panels *on the right*. The uppermost panel shows the shape of the line we would obtain if no relativistic effects occurred besides the non-relativistic Doppler effect. Below, the line profile is plotted taking the relativistic Doppler effect and beaming [see (5.37)] into account. This line profile is shifted towards smaller energies by gravitational redshift so that, in combination, the line profile shown at the bottom results. Source: A.C. Fabian et al. 2000, *Broad Iron Lines in Active Galactic Nuclei*, PASP 112, 1145, Fig. 3. ©ASP. Reproduced with permission

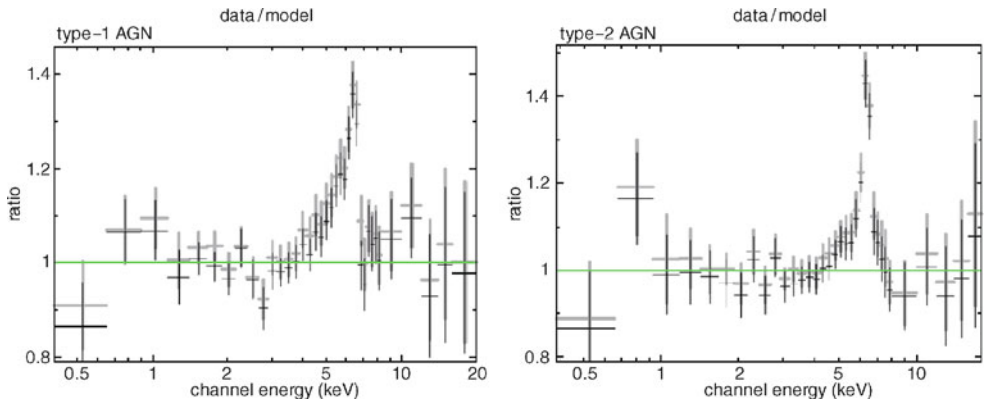

Fig. 5.22 The ratio of the X-ray spectrum of AGNs and a fitted power law averaged over 53 Type 1 AGNs (*left panel*) and 41 Type 2 AGNs (*right panel*). The *gray* and *black* data points are from two different detectors on-board the XMM-Newton observatory. In both AGN samples, a broad relativistic iron line is visible; in the Type 2 AGNs, an additional narrow line component at 6.4 keV can be identified. The line strength indicates that the average iron abundance in these sources is about three times the Solar value. Source: A. Streblyanska et al. 2005, *XMM-Newton observations of the Lockman Hole. III. A relativistic Fe line in the mean X-ray spectra of type-1 and type-2 AGN*, A&A 432, 395, p. 397, Figs. 2, 3. ©ESO. Reproduced with permission

translated into a minimum mass of the SMBH, required for its gravity to dominate the total force at a given luminosity.

We consider a fully ionized gas, so that the interaction of radiation with this infalling plasma is basically due to scattering of photons by free electrons. This is called *Thomson scattering*. The mean radiation force on an electron at radius r is then

$$F_{\rm rad} = \sigma_{\rm T} \frac{L}{4\pi\, r^2\, c}\,, \tag{5.22}$$

where

$$\sigma_{\rm T} = \frac{8\pi}{3}\left(\frac{e^2}{m_{\rm e}c^2}\right)^2 = 6.65\times 10^{-25}\,{\rm cm}^2 \tag{5.23}$$

denotes the *Thomson cross section* (in cgs units). This cross section is independent of photon frequency.[5] To derive (5.22), we note that the flux $S = L/(4\pi r^2)$ is the radiation energy which flows through a unit area at distance r from the central source per unit time. Then S/c is the momentum of photons flowing through this unit area per time, or the radiation pressure, because the momentum of a photon is given by its energy divided by the speed of light. Thus the momentum transfer to an electron per unit time, or the radiation force, is given by $\sigma_T S/c$. From (5.22), we can see that the radiation force has the same dependence on radius as the gravitational force, $\propto r^{-2}$, so that the ratio of the two forces is independent of radius.

Eddington luminosity. For matter to be able to fall in—the condition for energy production—the radiation force must be smaller than the gravitational force. For each electron there is a proton, and these two kinds of particles are electromagnetically coupled. The gravitational force per electron-proton pair is given by

$$F_{\rm grav} = \frac{GM_\bullet m_{\rm p}}{r^2} \,.$$

where we have neglected the mass of the electron since it is nearly a factor of 2000 smaller than the proton mass $m_{\rm p}$. Hence, the condition

$$F_{\rm rad} < F_{\rm grav} \tag{5.24}$$

for the dominance of gravity can be written as

$$\frac{\sigma_T L}{4\pi\, r^2\, c} < \frac{GM_\bullet m_{\rm p}}{r^2} \,,$$

or

$$\boxed{L < \frac{4\pi G\, c\, m_{\rm p}}{\sigma_T} M_\bullet =: L_{\rm edd} \approx 1.26\times 10^{38} \left(\frac{M_\bullet}{M_\odot}\right) {\rm erg/s}} \,, \tag{5.25}$$

where we have defined the *Eddington luminosity* $L_{\rm edd}$ *of a black hole of mass* $M_\bullet$. Since σ_T is independent of photon frequency, the luminosity referred to above is the bolometric luminosity.

A lower limit on $M_\bullet$. For accretion to occur at all, we need $L < L_{\rm edd}$. Remembering that the Eddington luminosity is proportional to $M_\bullet$, we can turn the above argument around: if a luminosity L is observed, we conclude $L_{\rm edd} > L$, or

$$\boxed{M_\bullet > M_{\rm edd} := \frac{\sigma_T}{4\pi G c m_{\rm p}} L \approx 8\times 10^7 \left(\frac{L}{10^{46}\,{\rm erg/s}}\right) M_\odot} \,. \tag{5.26}$$

Therefore, a lower limit for the mass of the SMBH can be derived from the luminosity. For luminous AGNs, like QSOs, typical masses are $M_\bullet \gtrsim 10^8 M_\odot$, while Seyfert galaxies have lower limits of $M_\bullet \gtrsim 10^6 M_\odot$. Hence, the SMBH in our Galaxy could in principle provide a Seyfert galaxy with the necessary energy.

In the above definition of the Eddington luminosity we have implicitly assumed that the emission of radiation is isotropic. In principle, the above argument of a maximum luminosity can be avoided, and thus luminosities exceeding the Eddington luminosity can be obtained, if the emission is highly anisotropic. A geometrical concept for this would be, for example, accretion through a disk in the equatorial plane and the emission of a major part of the radiation along the polar axes. Models of this kind have indeed been constructed. It was shown that the Eddington limit may be exceeded by this, but not by a large factor. However, the possibility of anisotropic emission has another very important consequence. To derive a value for the luminosity from the observed flux of a source, the relation $L = 4\pi D_{\rm L}^2\, S$ is applied, which is explicitly based on the assumption of isotropic emission. But if this emission is anisotropic and thus depends on the direction to the observer, the true luminosity may differ considerably from that which is derived under the assumption of isotropic emission. Later we will discuss the evidence for anisotropic emission in more detail.

Eddington accretion rate. If the conversion of infalling mass into energy takes place with an efficiency ϵ [see (5.15)], the accretion rate can be determined,

$$\dot m = \frac{L}{\epsilon\, c^2} \approx 0.18\, \frac{1}{\epsilon} \left(\frac{L}{10^{46}\,{\rm erg/s}}\right) M_\odot/{\rm yr} \,. \tag{5.27}$$

[5] When a photon scatters off an electron at rest, this process is called Thomson scattering. To a first approximation, the energy of the photon is unchanged in this process, only its direction is different after scattering. This is not really true, though. Due to the fact that a photon with energy E_γ carries a momentum E_γ/c, scattering will impose a recoil on the electron. After the scattering event the electron will thus have a non-zero velocity and a corresponding kinetic energy. Owing to energy conservation the photon energy after scattering is therefore slightly smaller than before. This energy loss of the photon is very small as long as $E_\gamma \ll m_{\rm e}c^2$. When this energy loss becomes appreciable, this scattering process is then called Compton scattering. If the electron is not at rest, the scattering can also lead to net energy transfer to the photon, such as it happens when low-frequency photons propagate through a hot gas (as we will discuss in Sect. 5.4.4 for the case of AGNs, and in Sect. 6.4.4 for galaxy clusters) or through a distribution of relativistic electrons. In this case one calls it the inverse Compton effect. The physics of all these effects is the same, only their kinematics are different.

Since the maximum efficiency is of order $\epsilon \sim 0.1$, this implies accretion rates of typically several Solar masses per year for very luminous QSOs. If L is measured in units of the Eddington luminosity, we obtain with (5.25)

$$\dot{m} = \frac{L}{L_{\rm edd}} \left(\frac{1.26 \times 10^{38}\,{\rm erg/s}}{\epsilon c^2} \right) \left(\frac{M_\bullet}{M_\odot} \right) \equiv \frac{L}{L_{\rm edd}}\, \dot{m}_{\rm edd}\,, \tag{5.28}$$

where in the last step the Eddington accretion rate has been defined,

$$\dot{m}_{\rm edd} = \frac{L_{\rm edd}}{\epsilon c^2} \approx \frac{1}{\epsilon}\, 2 \times 10^{-9}\, M_\bullet\, {\rm yr}^{-1}\,. \tag{5.29}$$

Growth rate of the SMBH mass. The Eddington accretion rate is the maximum accretion rate if isotropic emission is assumed, and it depends on the assumed efficiency ϵ. We can now estimate a characteristic time in which the mass of the SMBH will significantly increase,

$$t_{\rm evo} := \frac{M_\bullet}{\dot{m}} \approx \epsilon \left(\frac{L}{L_{\rm edd}} \right)^{-1} 5 \times 10^8\,{\rm yr}\,, \tag{5.30}$$

i.e., even with efficient energy production ($\epsilon \sim 0.1$), the mass of a SMBH can increase greatly on cosmologically short time-scales by accretion (see problem 5.4). However, this is not the only mechanism that can produce SMBHs of large mass. They can also be formed through the merger of two black holes, each of smaller mass, as would be expected after the merger of two galaxies if both partners hosted a SMBH in its center. This aspect will be discussed more extensively later.

5.4 Components of an AGN

In contrast to stars, which have a simple geometry, we expect several source components in AGNs with different, sometimes very complex geometric configurations to produce the various components of the spectrum; this is sketched in Fig. 5.23. Accretion disks and jets in AGNs are clear indicators for a significant deviation from spherical symmetry in these sources. The relation between source components and the corresponding spectral components is not always obvious. However, combining theoretical arguments with detailed observations has led to quite satisfactory models.

5.4.1 The IR, optical, and UV-continuum

In Sect. 5.3.2 we considered an accretion disk with a characteristic temperature, following from (5.14), of

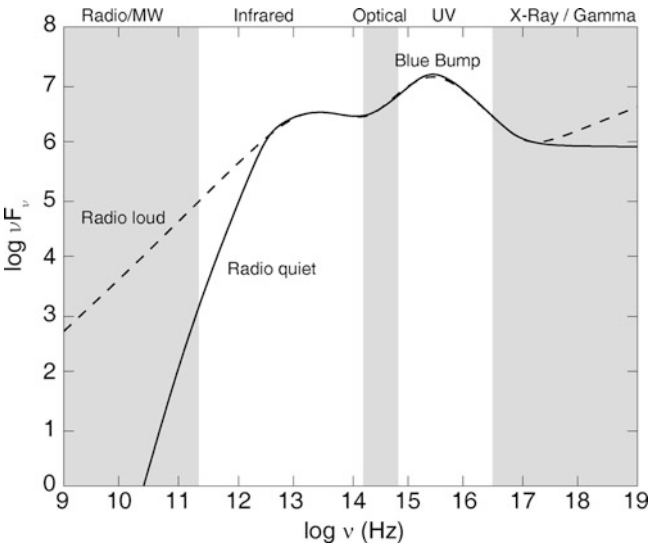

Fig. 5.23 Sketch of the characteristic spectral emission of a QSO. We distinguish between radio-loud (*dashed curve*) and radio-quiet (*solid curve*) QSOs. Plotted is νS_ν (in arbitrary units), so that flat sections in the spectrum correspond to equal energy per logarithmic frequency interval. The most prominent feature is the big blue bump, a broad maximum in the UV up to the soft X-ray domain of the spectrum. Besides this maximum, a less prominent secondary maximum is found in the IR. The spectrum increases towards higher energies in the X-ray domain of the spectrum—typically $\sim 10\,\%$ of the total energy is emitted as X-rays. For blazars, the spectrum can rise at even higher energies, yielding a large fraction of the total flux being radiated at gamma-rays

$$T(r) \approx 6.3 \times 10^5\,{\rm K} \left(\frac{\dot{m}}{\dot{m}_{\rm edd}} \right)^{1/4} \left(\frac{M_\bullet}{10^8 M_\odot} \right)^{-1/4} \left(\frac{r}{r_{\rm S}} \right)^{-3/4}. \tag{5.31}$$

The thermal emission of an accretion disk with this radial temperature profile produces a broad spectrum with its maximum in the UV. The continuum spectrum of QSOs indeed shows an obvious increase towards UV wavelengths, up to the limit of observable wavelengths, $\lambda \gtrsim 1000\,$Å. (This is the observed wavelength; QSOs at high redshifts can be observed at significantly shorter wavelengths in the QSO rest-frame.) At wavelengths $\lambda \leq 912\,$Å, photoelectric absorption by neutral hydrogen in the ISM of the Galaxy sets in, so that the Milky Way is opaque for this radiation. Only at considerably higher frequencies, namely in the soft X-ray band ($h_{\rm P}\nu \gtrsim 0.2\,$keV), does the extragalactic sky become observable again.

If the UV radiation of a QSO originates mainly from an accretion disk, which can be assumed because of the observed increase of the spectrum towards the UV, the question arises whether the thermal emission of the disk is also visible in the soft X-ray regime. In this case, the spectrum in the range hidden from observation, at $13\,{\rm eV} \lesssim h_{\rm P}\nu \lesssim 0.2\,$keV, could be interpolated by such an accretion

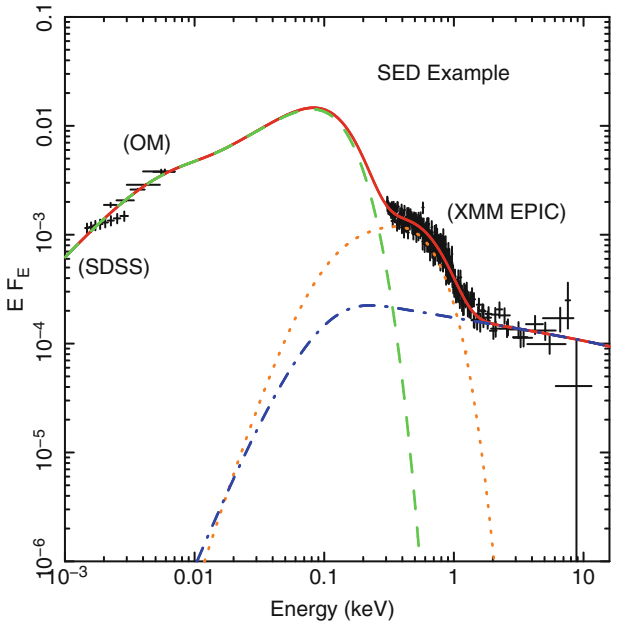

Fig. 5.24 A possible interpretation for the connection between the optical-UV spectrum and the X-ray spectrum of an AGN, interpolating through the unobservable spectral region between 13.6 eV and $\sim$ 0.2 keV. The total spectral energy distribution (*red curve*) in this model is composed of the accretion disk emission (*green dashed curve*) and the X-ray emission through low-temperature Comptonization with high optical depth (*orange dotted curve*) and high-temperature Comptonization at low optical depth (*blue dot-dashed curve*). Source: C. Jin et al. 2012, *A combined Optical and X-ray Spectra Study for Type 1 AGN. III. Broadband SED Properties*, MNRAS 425, 907, Fig. 1. Reproduced by permission of Oxford University Press on behalf of the Royal Astronomical Society

disk spectrum. This seems indeed to be the case. The X-ray spectrum of QSOs often shows a very simple spectral shape in the form of a power law, $S_\nu \propto \nu^{-\alpha}$, where $\alpha \sim 0.7$ is a characteristic value (see Sect. 5.4.4 below). However, the spectrum follows this power law only at energies down to $\sim$ 0.5 keV. At lower energies, the spectral flux can be higher than predicted by the extrapolation of the power-law spectrum observed at higher energies. One interpretation of this finding is that the (non-thermal) source of the X-ray emission produces a simple power law, and the additional flux at lower X-ray energies is emission from the innermost part of the accretion disk (see Fig. 5.23).

Perhaps these two spectral properties—the increase of the spectrum towards the UV and the radiation excess in the soft X-ray—have the same origin, being two wings of a broad maximum in the energy distribution, which itself is located in the spectral range unobservable for us. This maximum is called the *big blue bump* (BBB). A description of the BBB is possible using detailed models of accretion disks (see Fig. 5.24 for an example). For this modeling, however, the assumption of a local Planck spectrum at all radii of the disk is too simple because the structure of the accretion disk is more complicated. The spectral properties of an accretion disk have to be modeled by an 'atmosphere' for each radius, similar to that in stars. Indeed, since the radial temperature distribution (5.31) extends to soft X-ray energies only for very low $M_\bullet$, the X-rays (and the BBB) are probably not due to the thermal disk emission, but originates from a hot atmosphere (corona) of the disk, as will be discussed in more detail in Sect. 5.4.4.

Besides the BBB, an additional maximum exists in the MIR (IR-bump). This can be ascribed to thermal emission of warm dust ($T \lesssim 1000$ K). As we will discuss below, other observations provide additional evidence for this dust component, which may be associated with the absorbing torus (see Fig. 5.12).

The optical continuum of blazars is different from that of Seyfert galaxies and QSOs. It often features a spectral pattern that follows, to very good approximation, a power law and is strongly variable and polarized. This indicates that the radiation is predominantly non-thermal. The origin of this radiation thus probably does not lie in an accretion disk. Rather, the radiation presumably has its origin in the relativistic jets which we already discussed for the radio domain, with their synchrotron radiation extending up to optical wavelengths. This assumption was strongly supported by many sources where observations discovered optical and X-ray emission from jets (see Fig. 5.17 and Sect. 5.5.4).

Gravitational microlensing: Microscopy of the accretion disk. In Sect. 2.5 we discussed the Galactic microlensing effect, where a star is lensed by a compact object in our Milky Way. The observational signature of this effect is the flux variation of the source, which occurs due to a time-varying magnification of the background star caused by the lens.

In a strong lensing event where a QSO is mapped into several images by a foreground galaxy (see Sect. 3.11), another kind of microlensing can occur (and in fact had been investigated before Galactic microlensing was discussed). One needs to realize that the mass distribution in a galaxy is composed of a baryonic and a dark matter component, the former being dominated by stars. Thus, the mass distribution responsible for the gravitational light deflection is 'grainy', possessing small-scale structure. Depending on the size of the source being lensed, this graininess can be relevant.

To see why, we first consider the typical length scales of the situation. Consider first a single star of mass M in the lens; then the Einstein angle θ_E of this star is given by (2.82). The corresponding length-scale in the source plane is obtained by projecting this angle onto the source plane,

$$R_\mathrm{E} = D_\mathrm{s}\theta_\mathrm{E} = \sqrt{\frac{4GM}{c^2}\,\frac{D_\mathrm{s}D_\mathrm{ds}}{D_\mathrm{d}}} \sim 9\times10^{16}\,\mathrm{cm}\,\sqrt{\frac{M}{M_\odot}}\sqrt{\frac{D_\mathrm{s}}{c/H_0}}\sqrt{\frac{D_\mathrm{ds}}{D_\mathrm{d}}}, \tag{5.32}$$

where we assumed a Hubble constant of $h = 0.7$. Since the angular-diameter distance of QSOs is of order the Hubble radius, and the lens is somewhere in the middle between us and the source, the last two factors in (5.32) are of order unity. The typical stellar mass in an early-type galaxy is of order $M_\odot/2$; thus we conclude that the Einstein radius in the source plane is of the order of a few times 10^{16} cm, or about 10 light-days. Sources of that size or smaller can be significantly magnified by the lensing effect of the star.

This length-scale can be compared to typical sizes of QSOs. The rapid variability of QSOs in the X-rays implies that the X-ray emitting source is typically smaller than R_E; the same applies to the source component emitting UV-radiation. As we shall see below, the size of the region where the broad emission lines originate is typically larger than R_E. From these size considerations, we expect that the UV- and X-ray emission of QSOs can be magnified by stars in the lens galaxy.

There is a major difference between Galactic microlensing and QSO microlensing. We saw in Sect. 3.11 that the probability that a lens in our Galaxy is located close to the line-of-sight to a distant star is tiny—that is one of the reasons why Galactic microlensing surveys are so difficult. This is no longer true in QSO microlensing; the density of stars in the lens galaxy at the location where the multiple images occur is rather large. In fact, the mean separation between stars is not much larger than the Einstein radius of each star, somewhat depending on the local fraction of the surface mass density contained in stars compared to that in the form of dark matter. The consequence of this high microlens density is that the individual stars can no longer be considered as isolated point-mass lenses. Instead, an ensemble of microlenses needs to be considered whose joint lensing action causes the microlensing effect. This phenomenon can be studied with numerical simulations, by 'shooting' light rays through an ensemble of point-mass lenses.

In the upper part of Fig. 5.25, typical magnification patterns are shown, for two different values of the stellar density in the line-of-sight. The magnification pattern is a plot of the magnification $\mu(\boldsymbol{\beta})$ as a function of source position $\boldsymbol{\beta}$. The magnification of a source thus depends on its location relative to the stellar field of the lens (the stars were assumed to be randomly placed in the lens), as well as on the size of the source: The magnification of an extended source is given as the average of the point-source magnification across the brightness profile of the source. We see that characteristic lines of high magnification occur, the so-called caustics, that we already saw for the case of binary lenses (see Fig. 2.40 for an example).

Since the system of source, lens and observer has a relative motion transverse to the line-of-sight, the relative location of the source in the magnification pattern will change in time. Hence, the magnification of each image of a source varies due to microlensing. Synthetic light curves for two different source sizes are shown in the lower part of Fig. 5.25, for the two stellar densities. We see that in the low-density case, strong variations are rare and occur only when the source crosses a caustic. For the high-density case, the frequency of caustic crossings is much higher, and the periods of small flux variations become shorter.

Furthermore, we see from the light curves that there is another relevant scale in the problem: the fastest magnification variations are defined by the time it takes the source to cross a caustic. This time-scale is reflected in the sharp rise (or fall) of the magnification during caustic crossing. Thus, the rise/fall-time is given by the size of the source divided by the transverse velocity. The time-variable magnification of course corresponds to a change of the observed flux of the images.

Indeed, flux monitoring of multiply-imaged QSOs are carried out, predominantly to determine the time delays in lens systems (see Sect. 3.11.4). After accounting for the relative time delay, the light curves of multiply-imaged QSOs are not identical, but they vary independently of each other. This cannot be due to a variation of the source flux—as that would show up in all images with the respective time delay—but must be due to microlensing. Since we have good estimates for typical peculiar cosmic velocities, we can relate time-scales of variations to length-scales in the source plane, and thus estimate the size of the emission region (at a given spectral band).

Results. For any single lens system, the quantitative analysis of microlensing light curves is hampered by uncertainties about the transverse velocity needed to relate observed time-scales to intrinsic length scales. Thus, for any parameter to be estimated from the light curves, one obtains a probability distribution—also because there is stochasticity in the spatial location of the stars. For the lens system $2237 + 0305$, where the lens is at low redshift ($z_\mathrm{d} \approx 0.04$), the effective transverse velocity is probably dominated by that of the Solar System, well known through the CMB dipole. The optical light curves of this system (Fig. 5.26) show pronounced microlensing in all four images. From matching the observed light curves with synthetic ones from ray-shooting simulations, the size of the emitting region can then be estimated. In the left panel of Fig. 5.27, the solid red curve shows the probability distribution of the half-light radius of the optical emission region, here corresponding to a rest wavelength of $\lambda \approx 2000$ Å. The characteristic scale thus obtained is $R \approx 3 \times 10^{16}$ cm. From the available X-ray light curves of this lens system, the size of the X-ray emitting region can be estimated as well, as shown by the other two curves in this figure. The right-hand panel of Fig. 5.27 shows the estimated optical (filled black squares) and (for some systems with

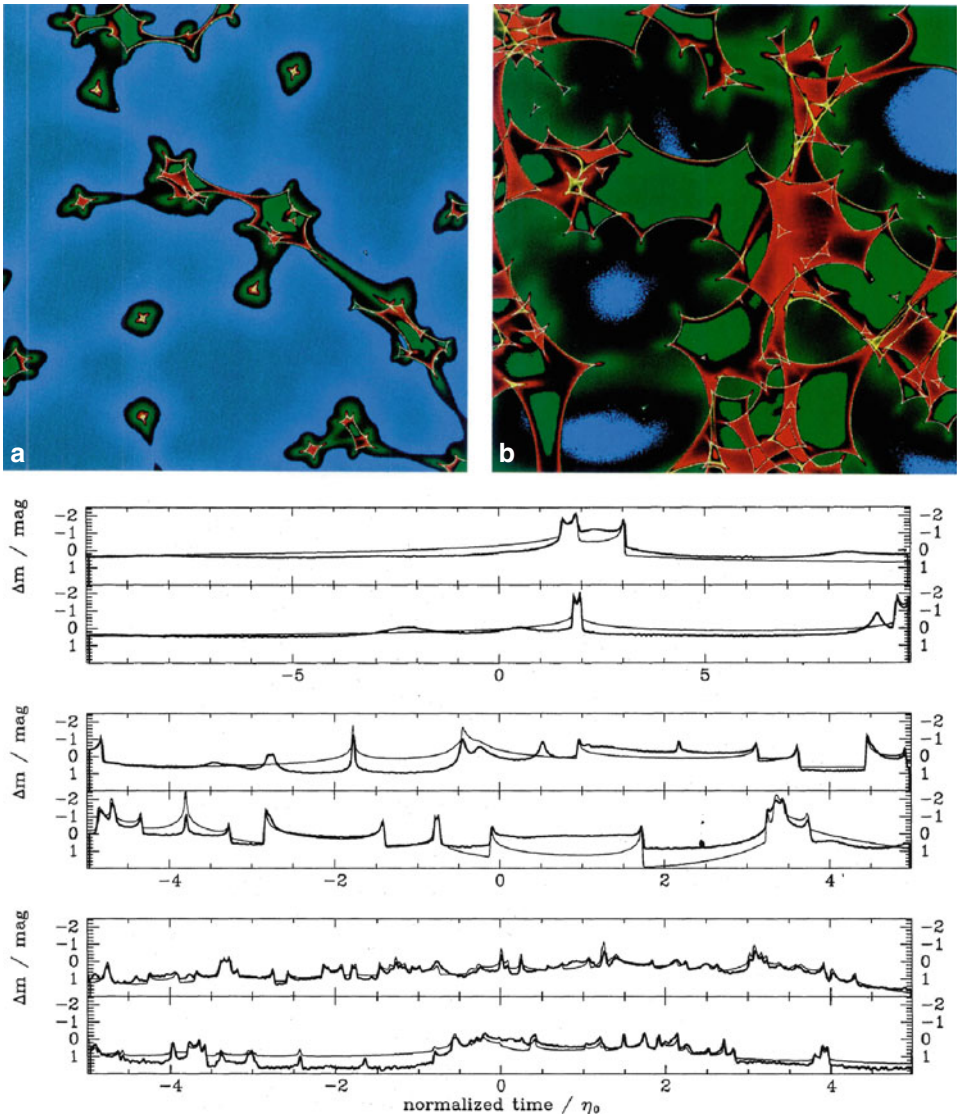

Fig. 5.25 The *upper panels* show magnification patterns from microlensing by stars in a lens galaxy, assuming a relatively low (*left*) or high (*right*) density of microlenses in the line-of-sight. Shown is the magnification as a function of position of a (small) source in the source plane, with *blue* indicating a low magnification, and increasingly higher ones are indicated in *green*, *red* and *yellow*. In the low-density case, one can identify several rather isolated point masses with their characteristic, almost axi-symmetric magnification pattern. However, many regions of larger magnification are connected. In the high-density case, there are no longer isolated point masses, and a complex pattern of magnification occurs. The highest magnifications occur along lines, the so-called caustics (see also Fig. 2.40). In the *lower two panels*, synthetic light curves, Δm as a function of time, normalized by the time is takes the source to move by one Einstein radius of the microlenses, are plotted as they occur if a source moves through the magnification patterns shown in the *upper panels*; the upper (lower) set of light curves correspond to the low (high) density microlensing field. In each case, light curves for two different source sizes are plotted, with the smoother curve corresponding to the large source. Source: J. Wambsganss et al. 1992, *Gravitational microlensing - Powerful combination of ray-shooting and parametric representation of caustics*, A&A 258, 591, Figs. 2,3. ©ESO. Reproduced with permission

available X-ray data) X-ray half-light radius as a function of estimated black hole mass, as obtained from microlensing studies.

With a characteristic size of the emitting region of $\sim 10^{16}$ cm, corresponding to about micro-arcsecond angular resolution (thus the name 'microlensing'), these microlensing studies yield the highest resolution observations available, almost three orders-of-magnitude higher than VLBI. The size of the optical source scales with black hole mass roughly as

$$R_{1/2} \approx 6 \times 10^{15} \left(\frac{M_\bullet}{10^9 M_\odot} \right)^{0.8} , \qquad (5.33)$$

which is shown as dashed curve in the right panel of Fig. 5.27. Given the uncertainty in the parameters of this

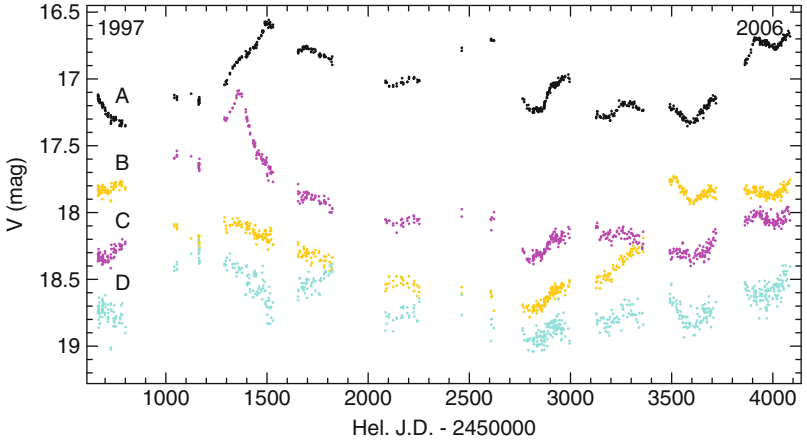

Fig. 5.26 Light curves of the four images of the lens system 2237 + 0305 (Fig. 3.59), taken over ten seasons 1997–2006. The time delay between the images is less than a day, thus all uncorrelated variability is due to microlensing. Source: A. Udalski et al. 2006, *The Optical Gravitational Lensing Experiment. OGLE-III Long Term Monitoring of the Gravitational Lens QSO 2237 + 0305*, Acta Astronomica 56, 293, p. 303. Reproduced by permission of the Copernicus Foundation for Polish Astronomy

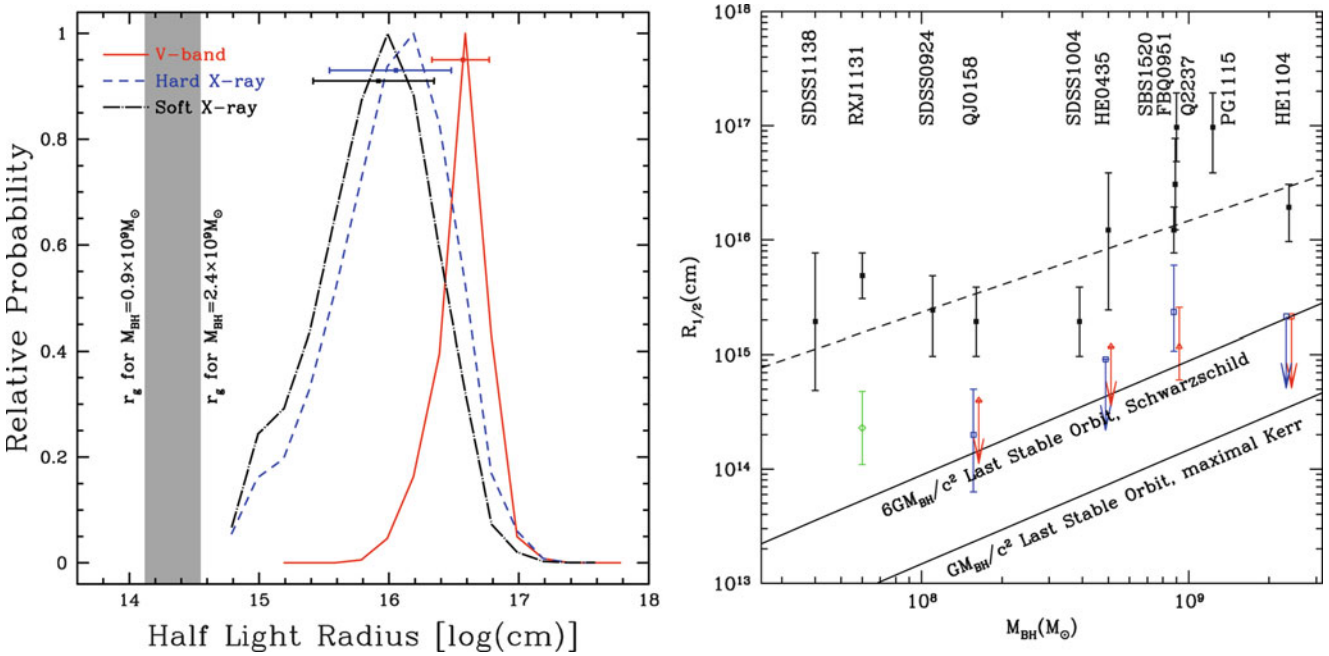

Fig. 5.27 *Left panel:* From an analysis of the microlensing light curves of the lens system 2237 + 0305, the probability distribution for the half-light radius is shown, for the optical emission region by the *solid red curve*, and the soft and hard X-ray emission by the *dash-dotted black* and *dashed blue curves*, respectively. The *grey band* indicates the 'gravitational radius' $r_g = GM_\bullet/c^2$, i.e., half the Schwarzschild radius, for a range of black hole masses between 0.9 and 2.4 × $10^9 M_\odot$, as estimated from the broad emission lines (see Sect. 5.4.2). *Right panel:* Estimates of the half-light radius for 11 multiply-imaged QSOs, plotted against the estimated black hole mass. *Filled black squares* correspond to the optical emitting region, the *other symbols* to X-ray emission, where *arrows* indicate upper limits. The *two solid lines* show the size of the last stable circular orbit for a black hole without spin (Schwarzschild black hole) and one with maximum spin parameter ('maximal Kerr'). Source: A. Mosquera et al. 2013, *The Structure of the X-Ray and Optical Emitting Regions of the Lensed Quasar Q 2237 + 0305*, ApJ 769, 53, p. 5, 7, Figs. 4, 8. ©AAS. Reproduced with permission

power-law fit, this scaling of emission size with $M_\bullet$ is compatible with our model of accretion disks: for a fixed temperature, (5.13) predicts $r^3 \propto \dot{m} M_\bullet \propto (\dot{m}/\dot{m}_{\rm edd})\epsilon^{-1} M_\bullet^2$, where we scaled the accretion rate by the Eddington rate (5.29). Thus, for a given wavelength, standard accretion disk theory predicts $r \propto M_\bullet^{2/3}$, not very different from the scaling in (5.33). However, taken at face value, the prefactor in (5.33) implies that the efficiency ϵ is surprisingly

small, $\epsilon \sim 2\,\%$, compared to the usually expected $\sim 10\,\%$. Whereas the estimate is affected by the aforementioned stochasticity of the effect, as well as factors like inclination angle distribution of the accretion disks, this difference may indicate a deviation from the simple picture of the optical emission coming solely from an accretion disk – for example, some fraction of the optical light may be scattered by gas at larger radius, thus effectively increasing the source size.

Furthermore, we see from Fig. 5.27 that the emission region of X-rays is significantly smaller than the optical emitting region. In fact, the estimated size of the X-ray region is comparable to the radius of the last stable circular orbit in a non-rotating black hole, $R = 3r_S = 6GM_\bullet/c^2$. Hence, in agreement with the size estimates from the broad iron line (Sect. 5.3.4), the X-rays originate from the very central region around the SMBH. Indeed, the estimated X-ray size is compatible with a scaling $R_{1/2} \propto M_\bullet$.

To conclude, microlensing allows us to study the smallest scales in AGNs, and the estimated sizes will become more accurate with increased length of the observed light curves and increasing the sample size.

5.4.2 The broad emission lines

Characteristics of the broad line region. One of the most surprising characteristics of AGNs is the presence of very strong emission lines, except for BL Lac objects which show an almost featureless spectrum. Furthermore, in Type 1 AGN, such as QSOs and Seyfert 1s, the emission lines are very broad. Interpreted as Doppler velocities, the corresponding width of the velocity distribution of the components in the emitting region is of order $\Delta v \lesssim 10\,000\,\mathrm{km/s}$ (or $\Delta\lambda/\lambda \lesssim 0.03$). These lines cannot be due to thermal line broadening because that would imply $k_B T \sim m_p(\Delta v)^2/2 \sim 1\,\mathrm{MeV}$, or $T \sim 10^{10}\,\mathrm{K}$—no emission lines would be produced at such high temperatures because all atoms would be fully ionized (plus the fact that at such temperatures a plasma would efficiently produce e^+e^--pairs, and the corresponding annihilation line at 511 keV should be observable in Gamma radiation). Therefore, the observed line width is interpreted as Doppler broadening. The gas emitting these lines then has large-scale velocities of order $\sim 10\,000\,\mathrm{km/s}$. Velocities this high are indicators of the presence of a strong gravitational field, as would occur in the vicinity of a SMBH. If the emission of the lines occurs in gas at a distance r from a SMBH, we expect characteristic velocities of

$$v_{\mathrm{rot}} \sim \sqrt{\frac{GM_\bullet}{r}} = \frac{c}{\sqrt{2}} \left(\frac{r}{r_S}\right)^{-1/2} ,$$

so for velocities of $v \sim c/30$, we obtain a radial distance of $r \sim 500\, r_S$.

Hence, the Doppler broadening of the broad emission lines can be produced by Kepler rotation at radii of about $500\, r_S$. Although this estimate is based on the assumption of a rotational motion, the infall velocity for free fall does not differ by more than a factor $\sqrt{2}$ from this rotational velocity. Thus the kinematic state of the emitting gas is of no major relevance for this rough estimate if only gravity is responsible for the occurrence of high velocities.

The region in which the broad emission lines are produced is called the *broad line region* (BLR). The density of the gas in the BLR can be estimated from the lines that are observed. To see this, it must be pointed out that allowed and semi-forbidden transitions are found among the broad lines. Examples of the former are Lyα, MgII, and CIV, whereas CIII] and NIV] are semi-forbidden transitions. However, forbidden transitions are essentially absent among the broad lines.[6]

An excited atom can transit into its ground state (or another lower-lying state) either by spontaneous emission of a photon or by losing energy through collisions with other atoms. The probability for a radiative transition is defined by the atomic parameters, whereas the collisional de-excitation depends on the gas density. If the density of the gas is high, the mean time between two collisions is much shorter than the average lifetime of forbidden or semi-forbidden radiative transitions. Therefore the corresponding line photons are not observed.[7] The absence of forbidden lines is then used to derive a lower limit for the gas density, and the occurrence of semi-forbidden lines yields an upper bound for the density. To minimize the dependence of this argument on the chemical composition of the gas, transitions of the same element are preferentially used for these estimates. However, this is not always possible. From the presence of the CIII] line and the non-existence of the [OIII] line in the BLR, combined with model calculations, a density estimate of $n_e \sim 3 \times 10^9\,\mathrm{cm}^{-3}$ is obtained. However, as we shall see shortly, the conditions in the BLR are not uniform, but the BLR extends over a range of scales. The CIII] line originates

[6] The classification into allowed, semi-forbidden, and forbidden transitions is done by means of quantum-mechanical transition probabilities, or the resulting mean time for a spontaneous radiational transition. Allowed transitions correspond to electric dipole radiation, which has a large transition probability, and the lifetime of the excited state is then typically only 10^{-8} s. For forbidden transitions, the time-scales are considerably larger, typically 1 s, because their quantum-mechanical transition probability is substantially lower. Semi-forbidden transitions have a lifetime between these two values. To mark the different kinds of transitions, a double square bracket is used for forbidden transitions, like in [OIII], while semi-forbidden lines are marked by a single square bracket, like in CIII].

[7] To make forbidden transitions visible, the gas density needs to be very low. Such low densities cannot be produced in the laboratory. Forbidden lines are in fact not observed in laboratory spectra; they are 'forbidden'.

from rather large radii. In the inner-most part of the BLR, the electron density is higher, $n_e \sim 10^{11}$ to $10^{12}\,\mathrm{cm}^{-3}$

Furthermore, from the ionization stages of the line-emitting elements, a temperature can be estimated, typically yielding $T \sim 20\,000$ K. Detailed photoionization models for the BLR are very successful and are able to reproduce details of line ratios very well.

From the density of the gas and its temperature, the emission measure can then be calculated (i.e., the number of emitted line photons per unit time and per unit volume element). From the observed line strength and the distance to the AGN, the total number of emitted line photons can be calculated, and by dividing through the emission measure, the volume of the line-emitting gas can be determined. This estimated volume of the gas is much smaller than the total volume ($\sim r^3$) of the BLR. We therefore conclude that the BLR is not homogeneously filled with gas; rather, the gas has a very small filling factor. The gas in which the broad lines originate fills only a small fraction (estimates range from $\sim 10^{-7}$ to ~ 0.1) of the total volume of the BLR; hence, it must be concentrated in clouds.

Geometrical picture of the BLR. From the previous considerations, a picture of the BLR emerges in which it contains gas clouds with a characteristic particle density of $n_e \sim 10^{10}\,\mathrm{cm}^{-3}$. In these clouds, heating and cooling processes take place. Probably the most important cooling process is the observed emission of broad emission lines, with $\sim 25\,\%$ of the cooling due to the iron lines (see Fig. 5.3). Heating of the gas is provided by energetic continuum radiation from the AGN which photoionizes the gas, similar to processes in Galactic gas clouds. The difference between the energy of a photon and the ionization energy yields the energy of the released electron, which is then thermalized by collisions and leads to gas heating. In a stationary state, the heating rate equals the cooling rate, and this equilibrium condition defines the temperature which the clouds attain.

The comparison of continuum radiation and line emission yields the fraction of ionizing continuum photons which are absorbed by the BLR clouds; a value of about $10\,\%$ is obtained. Since the clouds are optically thick to ionizing radiation, the fraction of absorbed continuum photons is also the fraction of the solid angle subtended by the clouds, as seen from the central continuum source. From the filling factor and this solid angle, the characteristic size of the clouds can be estimated, from which we obtain typical values of $\sim 10^{11}$ to $\sim 10^{14}$ cm, depending on the filling factor. In addition, based on these arguments, the number of clouds in the BLR can be estimated. This yields a typical value of $\sim 10^{10}$. An independent argument for a very large number of clouds comes from the fact that the observed line profiles are very smooth. Since the width of the emission line from an individual cloud is very much smaller (of order $\sim 20\,\mathrm{km/s}$ as thermal Doppler width for temperatures of $T \sim 2 \times 10^4$ K) than the observed line width, the broad lines must be produced by the superposition of the line emission from many clouds. If the number of clouds is not much larger than the ratio of observed line width to individual cloud line width, then the observed emission line will contain a kind of 'Poisson noise'; the smoothness of the lines clearly shows that this noise must be small, yielding a lower limit of $\sim 10^7$ clouds in the BLR.[8]

The characteristic velocity of the clouds corresponds to the line width, hence several thousand km/s. However, the kinematics of the clouds is unknown. We do not know whether they are rotating around the SMBH, whether they are infalling or streaming outwards, or whether their motion is rather chaotic. Reasonably good arguments are found for several of these possibilities. The fact that higher ionization lines exhibit a blueshift relative to the narrow emission lines may indicate an outflow of the BLR gas, with the part of the gas streaming away from us being covered by the accretion disk—so that we can see predominantly that part of the flow moving towards us. However, the systematic blueshift could also be produced by extinction of the red wing of the emission line, which would argue for an inflow of the absorbing material. There are some Type 1 AGNs which show broad Balmer emission lines with a double peak, such as would be produced if the emitting region was in a disk-like flow. It is also possible that different regions within the BLR exist with different kinematic properties.

Reverberation mapping: the principle. A direct method to examine the extent of the BLR is provided by *reverberation mapping*. This observational technique utilizes the fact that heating and ionization of the gas in the BLR are both caused by the central continuum source of the AGN. Since the UV radiation of AGNs varies, we expect corresponding variations of the physical conditions in the BLR. In this picture, a decreasing continuum flux should then lead to a lower line flux, as is demonstrated in Fig. 5.28. Due to the finite extent of the BLR, the observed variability in the lines will be delayed in time compared to the ionizing continuum. This delay Δt can be identified with the light travel time across the BLR, $\Delta t \sim r/c$. In other words, the BLR feels the variation in the continuum source only after a delay of Δt. From the observed correlated variabilities of continuum and line emission, Δt can be determined for

[8] Note, however, that this argument essentially pictures the clouds as having some random velocities. It is not unlikely that the picture of 'clouds' is somewhat misleading; instead, the BLR could consist of a turbulent gas, with a large-scale velocity field, in which condensations are present. These condensations then take the roles of the 'clouds' in the simple picture.

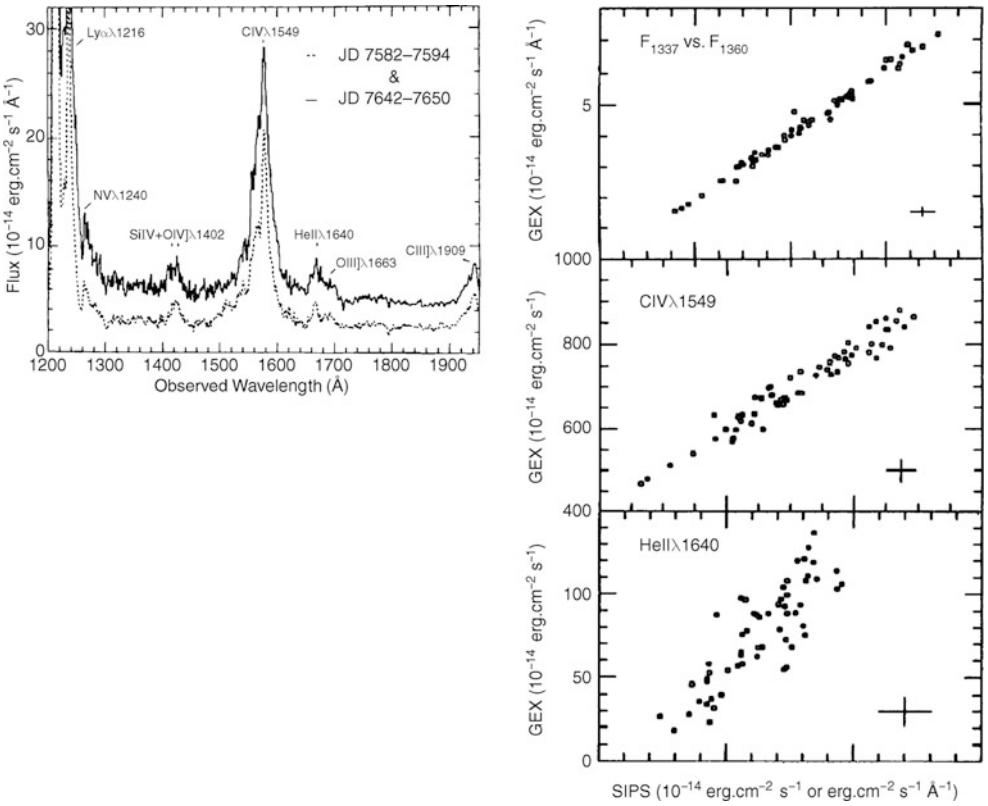

Fig. 5.28 In the *left-hand panel*, the UV spectrum of the Seyfert 1 galaxy NGC 5548 is plotted for two different epochs in which the source radiated strongly and weakly, respectively. It can clearly be seen that not only does the continuum radiation of the source vary but also the strength of the emission lines. The *right-hand panels* show the flux of the continuum at ~ 1300 Å, the CIV line at $\lambda = 1549$ Å, and the HeII line at $\lambda = 1640$ Å, as a function of the near-UV flux at different epochs during an 8-month observational campaign with the IUE. Source: J. Clavel et al. 1991, *Steps toward determination of the size and structure of the broad-line region in active galactic nuclei. I - an 8 month campaign of monitoring NGC 5548 with IUE*, ApJ 366, 64, p. 69, 76, Figs. 1, 2. ©AAS. Reproduced with permission

different line transitions, and so the corresponding values of r can be estimated.[9]

Such analyses of reverberation mapping are extremely time-consuming and complex because one needs to continuously monitor the continuum light and, simultaneously, the line fluxes of an AGN over a long period. The relevant time-scales are typically months for Seyfert 1 galaxies (see Fig. 5.29). To perform such measurements, coordinated campaigns involving many observatories are necessary, because the light curves have to be observed without any gaps, and one should not depend on the local weather conditions at any single observatory.

Reverberation mapping: results. From the results of such campaigns and the correlation of the light curves in the UV continuum and the different line fluxes (Fig. 5.30), the picture of an inhomogeneous BLR is obtained which extends over a large range in r and which consists of different 'layers'. The various emission lines are emitted at different radii, because the ionization structure of the BLR varies with r; the higher the ionization energy of a transition, the smaller the corresponding radius r. For the Seyfert 1 galaxy NGC 5548, one obtains $\Delta t \sim 12$ d for Lyα, about $\Delta t \sim 26$ d for CIII], and about 50 d for MgII. This may not come as a surprise because the ionizing flux increases for smaller r. The fact that lines of higher ionization energy are located closer to the central continuum source implies that they are also broader than low-ionization lines, according to the scaling $v \sim \sqrt{GM_\bullet / r}$.

We can understand this behavior by considering the two processes which are responsible for the ionization state of the gas. On the one hand, ionization is produced by the ionizing flux of the continuum source; thus, the ionization rate is proportional to L/r^2, where L is the luminosity (strictly speaking, L should be taken as the luminosity in ionizing photons, but if we assume that the shape of the continuum does not vary strongly, one can also use the luminosity in some optical or UV band). On the other hand, ions undergo recombination, with the recombination rate being proportional to the number density of electrons. Thus, we expect that the ionization state depends on the ratio of these two rates. One therefore defines the ionization parameter

$$\Xi = \frac{L}{r^2 n_\mathrm{e}} , \tag{5.34}$$

which describes the relative efficiency of ionization and recombination. The larger Ξ, the more abundant are ions

[9] The emissivity of the gas in the BLR reacts very quickly to a change of the ionizing radiation: if the ionizing flux onto a cloud in the BLR decreases, the corresponding line emission from the cloud decreases on the recombination time scale. For a gas density of $n \sim 10^{11}$ cm^{-3}, this time scale is about a minute—that is, almost instantaneously. Thus, the line emission from a cloud depends on the instantaneous ionizing flux at the cloud.

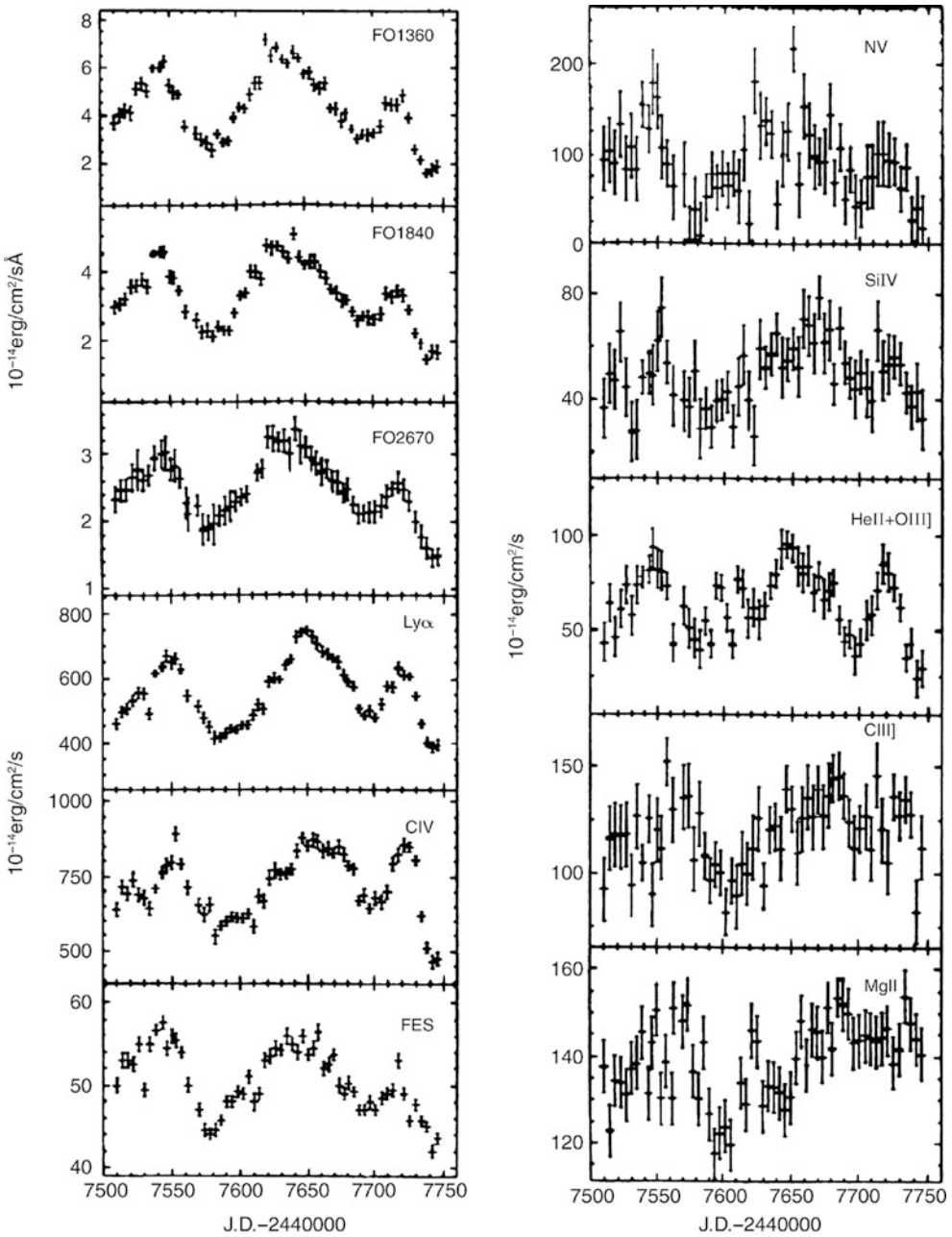

Fig. 5.29 Light curve of NGC 5548 over a period of 8 months at different wavelengths. In the *left-hand panels*, from top to bottom, the continuum at $\lambda = 1350$ Å, $\lambda = 1840$ Å, and $\lambda = 2670$ Å, the broad and strong emission lines Lyα and CIV, as well as the optical light curve are plotted. The *right-hand panels* show the weaker lines NV at $\lambda = 1240$ Å, SiIV at $\lambda = 1402$ Å, HeII+OIII] at $\lambda = 1640$ Å, CIII] at $\lambda = 1909$ Å, and MgII at $\lambda = 2798$ Å. Source: J. Clavel et al. 1991, *Steps toward determination of the size and structure of the broad-line region in active galactic nuclei. I - an 8 month campaign of monitoring NGC 5548 with IUE*, ApJ 366, 64, p. 78, Figs. 3, 4. ©AAS. Reproduced with permission

with high ionization energies, in agreement with the finding from reverberation studies.

If the ionization parameter Ξ is indeed the relevant quantity for the ionization structure of the BLR, then one expects that the size of the region from which a particular line is emitted increases with increasing continuum luminosity. In fact, if one assumes that the density n_e is a constant, one would infer that $r \propto L^{1/2}$ to keep Ξ constant. The Seyfert 1 galaxy NGC 5548 has been monitored for many years, and in addition to short-term, low-amplitude variations of its flux, which is used for the reverberation mapping, its flux varies by larger factors ($\sim$ 2 mag) on longer time scales. Hence, the source has been observed at different levels of activity, i.e., different L. In the left panel of Fig. 5.31, the delay time Δt as measured for the Hβ line is plotted against the continuum luminosity of this source. Clearly, the time lag is correlated with the optical luminosity of this source. The best power-law fit of the relation between time lag ($\propto r$) and luminosity yields a slope of $\sim 0.66 \pm 0.13$, slightly steeper that inferred from the foregoing argument of constant Ξ. However, the uncertainty in the slope is appreciable. In addition, we had to assume that n_e is constant. If n_e decreases with r, a larger slope would be expected. As the bottom line, we see that the extent of the region from which a specific

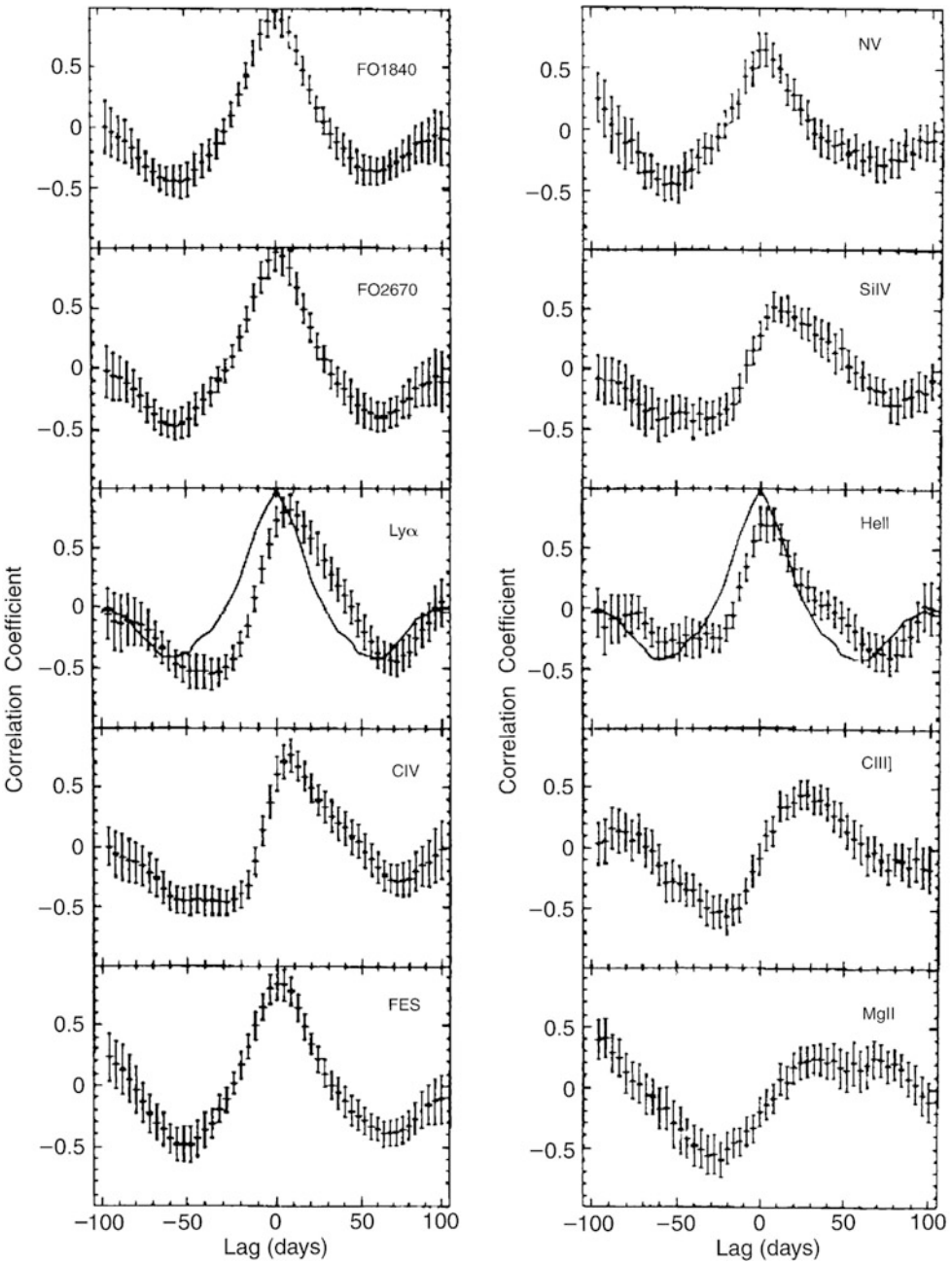

Fig. 5.30 The different light curves from Fig. 5.29 are correlated with the continuum flux at $\lambda = 1350$ Å. The autocorrelation function is shown by the *solid line* in the central panels, the others are cross-correlation functions. We can see that the maximum of the correlation is shifted towards positive times—variations in the continuum flux are not simultaneously followed by the emission lines but appear only after a delay. This delay corresponds to the light travel time from the center of the AGN to the clouds of the BLR where the lines are emitted. The smaller the ionization level of the respective ion, the longer the delay. For example, we obtain a delay of 12 days for Lyα, 26 days for CIII], and about 50 days for MgII, where the latter value could not be measured exactly because the relative flux variations of this line are small and thus the correlation function does not show a very prominent maximum. Source: J. Clavel et al. 1991, *Steps toward determination of the size and structure of the broad-line region in active galactic nuclei. I - an 8 month campaign of monitoring NGC 5548 with IUE*, ApJ 366, 64, p. 79, Figs. 5, 6. ©AAS. Reproduced with permission

line is emitted increases with the source luminosity. This fact has also been demonstrated by comparing the time lags from reverberation mapping of AGNs with different luminosity—the larger L, the larger is the extent of the BLR, scaling roughly as expected from the constancy of Ξ, as seen in the right-hand panel of Fig. 5.31.

The relative flux variations in lines of higher ionization energy are larger, as can also be seen in Fig. 5.14. In addition, lines of higher ionization energy have a mean wavelength systematically shifted bluewards compared to narrower emission lines. As mentioned before, this fact could be interpreted as an indication for outflowing motion of the BLR; however, velocity-resolved reverberation mapping in some Seyfert 1 galaxies hints towards an inflow motion.

Thus, the nature of the clouds in the BLR is to some degree still unknown, as are their geometrical distribution and their kinematic behavior. Their small extent and high temperature imply that they should vaporize on very small time-scales unless they are somehow stabilized. Therefore these clouds need to be either permanently replenished or they have to be stabilized, either by external pressure, e.g., from a very hot but thin medium in the BLR in between the clouds, by magnetic fields, or even gravitationally. One possibility is that the clouds are the extended atmospheres of

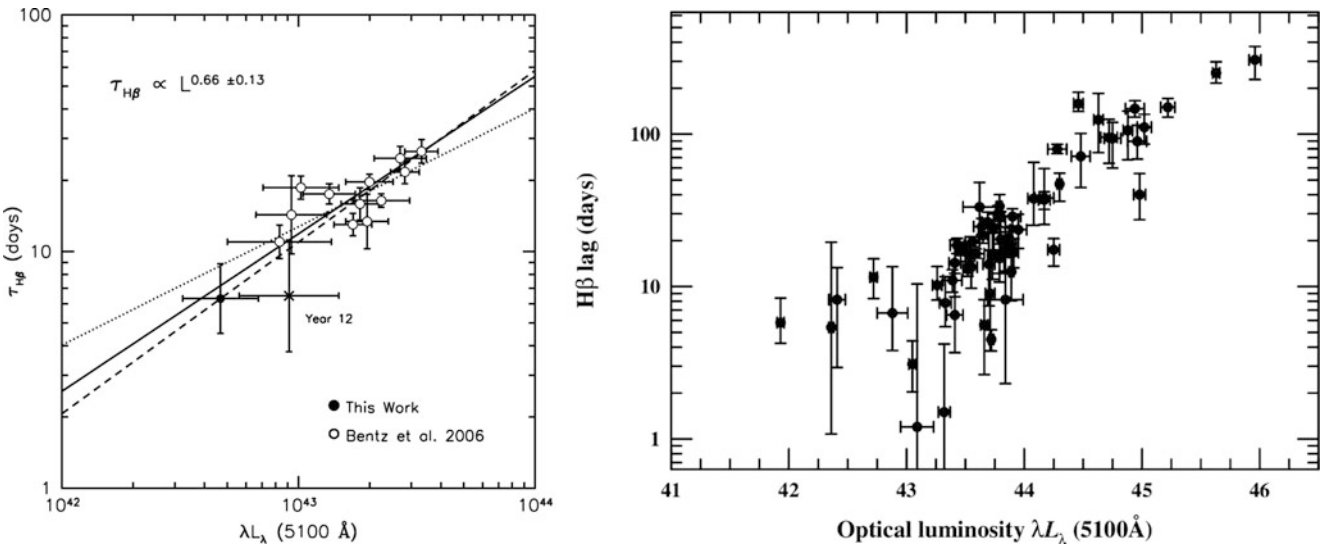

Fig. 5.31 *Left panel*: For the source NGC 5548, the time lag between line- and continuum variations are shown for the Hβ line, as a function of the optical continuum luminosity of the source. The latter was obtained after correcting the optical flux for the contribution by stellar light of the host galaxy. The data were obtained by monitoring this Seyfert 1 galaxy for almost two decades. The *dotted line* indicates the scaling $\delta t \propto L^{1/2}$, whereas the *solid line* is the best fitting power law, with the data point 'Year 12' (for which the sampling during the reverberation mapping campaign has been worse than in other years) excluded. *Right panel*: Measured time lags of the Hβ line as a function of optical luminosity, for a sample of 35 AGNs with reverberation mapping. Note the large range of luminosity. The slope of the best-fitting power law is again 0.6 ± 0.1 as for the case of NGC 5548. Source: *Left*: M.C. Bentz et al. 2007, *NGC 5548 in a Low-Luminosity State: Implications for the Broad-Line Region*, ApJ 662, 205, p. 210, Fig. 5. ©AAS. Reproduced with permission. *Right*: B.M. Peterson 2006, *Emission-Line Variability in Active Galactic Nuclei*, ASPC 360, 191, p. 196. Reproduced by permission of the Astronomical Society of the Pacific

stars; this would, however, imply a very high (probably too high) total mass of the BLR.

5.4.3 Narrow emission lines

Besides the broad emission lines that occur in QSOs, Seyfert 1 galaxies, and broad-line radio galaxies, most AGNs (with exception of the BL Lacs) show narrow emission lines. Their typical width is $\sim 400\,\mathrm{km/s}$. This is considerably narrower than lines of the BLR, but still significantly broader than characteristic velocities in normal galaxies. In analogy to the BLR, the region in which these lines are produced is known as the *narrow line region* (NLR). The strongest line from the NLR is, besides Lyα and CIV, the forbidden [OIII] line at $\lambda = 5007\,\text{Å}$. The existence of forbidden lines implies that the gas density in the NLR is significantly lower than in the BLR.

The gas in the NLR is also assumed to be ionized by UV-radiation from the central continuum source. From estimates analogous to those for the BLR, the characteristic properties of the NLR can then be determined. It should be noted that no reverberation mapping can be applied, since the NLR extends over a region of $\sim 100\,\mathrm{pc}$ for Seyfert 1 galaxies. Because of this large size, no variability of the narrow line intensities is expected on time-scales accessible to observation, and none has been found. The line ratios of allowed and forbidden lines yield $n_\mathrm{e} \sim 10^4\,\mathrm{cm}^{-3}$ for the typical density of the gas in which the lines originate. The characteristic temperature of the gas is likewise obtained from line ratios, $T \sim 15\,000\,\mathrm{K}$, probably slightly lower than in the BLR. The filling factor here is also significantly smaller than unity, about 10^{-2}. Hence, the geometrical picture of clouds in the NLR also emerges. Like in the BLR, the properties of the NLR are not homogeneous but vary with r.

With a size of $r \sim 100\,\mathrm{pc}$, the NLR can be spatially resolved for nearby Seyfert galaxies. The morphology of the NLR is very interesting: it is not spherical, but appears as two cone-shaped regions (Fig. 5.32). It seems as if the ionization of the NLR by the continuum radiation of the AGN is not isotropic, but instead depends strongly on the direction, and is confined largely to a cone-shaped region, called 'ionization cone'.

The BPT diagram. Active galaxies are not the only galaxies which show emission lines; whenever a galaxy undergoes active star formation, its spectrum will contain emission lines from the ionized regions (HII-regions) around newly-born, hot stars. If a galaxy shows broad emission lines, or other clear signs of nuclear activity (such as strong non-thermal radio emission), its identification as an AGN is straightforward. However, for many emission-line galaxies,

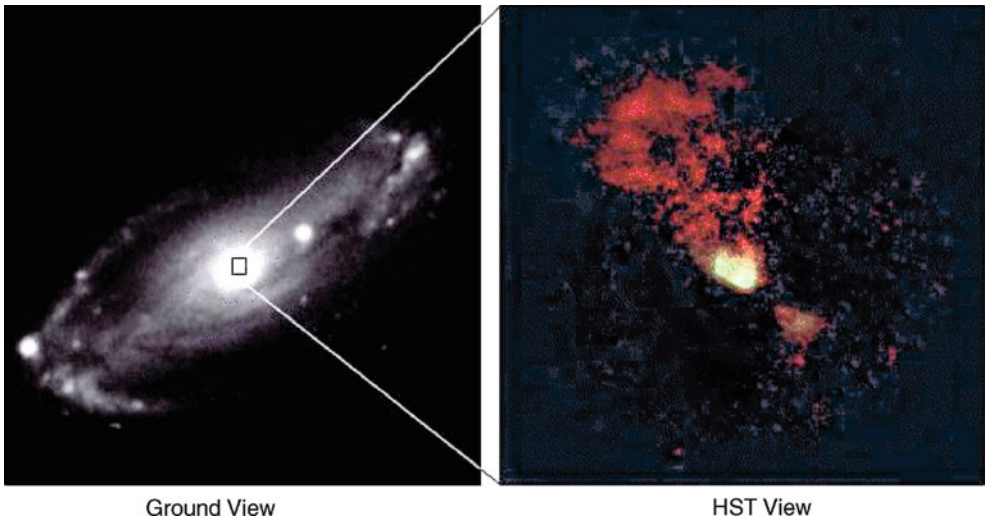

Fig. 5.32 Image of the Seyfert galaxy NGC 5728. On the *left*, a large-scale image showing the disk galaxy; on the *right*, an HST image of its central region taken through a filter with a small bandwidth (narrow-band filter) centered on a narrow emission line. This image shows the spatially resolved NLR. We can see that it is not spherical but consists of two cones ('ionization cones'). From this, it is concluded that the ionizing radiation of the AGN is not isotropic, but is emitted in two preferred directions which appear to be perpendicular, at least in this case, to the disk of the Galaxy. Credit: Allan Sandage, Observatories of the Carnegie Institution of Washington & Andrew S. Wilson, Department of Astronomy, University of Maryland; STScI/NASA

distinguishing between the possible sources of emission lines is non-trivial. For the physical interpretation of a galaxy, this distinction is essential: if a galaxy shows bright emission lines from its central region, it can be due to a burst of star formation, or due to a central AGN.

As pointed out by Baldwin, Phillips, and Terlevich in 1981, the ratios of line strengths can be employed for identifying the different origins of emission lines. The basic idea behind these BPT diagrams is that the source of photons, which ionizes the gas producing the emission lines, is different in these two cases: Massive stars have a clear cut-off in their ionizing spectrum, at the Lyman-limit of helium (corresponding to $\lambda = 228\,\text{Å}$), whereas the non-thermal radiation from AGNs extends to much higher photon energies. One can show that, as a consequence, the ratio of collisionally excited lines to that of lines which are produced in the course of recombination is larger in the case of an AGN-like ionizing radiation field.

Figure 5.33 shows an example of such a BPT diagram, based on SDSS galaxy spectra. As diagnostics, the line ratios [NII]/Hα and [OIII]/Hβ are chosen here; in both cases, the wavelengths of the two lines are quite similar. Thus, these line ratios should only weakly be affected by extinction (of course, ratios of fluxes are independent of the distance to the sources). One sees that the distribution of galaxies in that diagram shows a distinctive pattern: On the one hand, galaxies are distributed along an 'arc' (shown in blue), with an upper envelope indicated by the dashed curve; on the other hand, a second major concentration extends from the lower-right part of the arc towards larger line ratios, shown in grey and red. From numerical modeling, coupled with stellar population synthesis models, it was found that star-forming regions cannot produce line ratios which are above the dotted line. Hence, objects located above the dotted line are essentially powered by AGN radiation. The morphology of the galaxy distribution in the BPT diagram suggests that the galaxies below the dashed curve form a class of its own; therefore, galaxies in the region indicated in blue are considered to be powered solely by star formation. Objects in the region between the two curves, shown in grey, can originate from both, star formation and AGN activity, and they are frequently considered as 'composite' objects.

Furthermore, the distribution of AGN-powered objects in Fig. 5.33 seems to display a further substructure; there seems to be a bimodal distribution in the line ratio of [OIII]/Hβ. Objects with a larger value of this line ratio are found to be Seyfert 2 galaxies (and Seyfert 1 galaxies, if only their narrow line ratios are considered), whereas those with a lower line ratio correspond to LINERs.

The BPT diagram, and variants thereof (in which different line ratios are considered) are a useful tool for classifying emission-line galaxies and are thus extensively used.

5.4.4 X-ray emission

The most energetic radiation of an AGN is expected to be produced in the immediate vicinity of the SMBH. Therefore, the X-ray emission of AGNs is of special interest for probing the innermost regions of these objects, as we have already

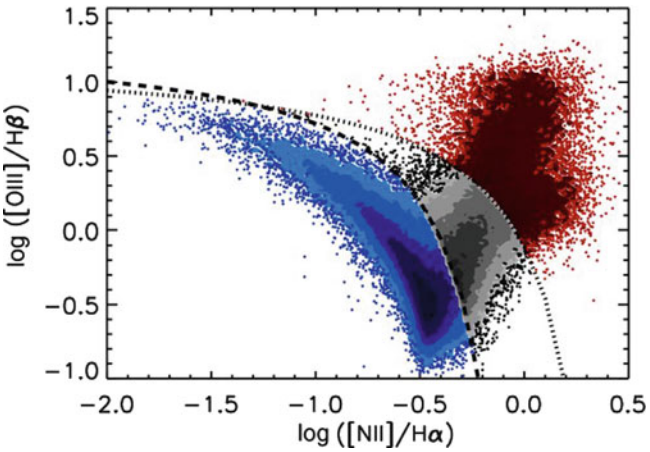

Fig. 5.33 The BPT diagram, a diagnostic for the primary source of emission lines in galaxies. Plotted is the distribution of $\sim$ 240 000 galaxies in the SDSS with $0.02 \leq z \leq 0.35$ for which the indicated emission lines were detected with S/N $>$ 5, with the line ratio [NII]/Hα against [OIII]/Hβ. The *shading* indicates the number density of galaxies, and individual galaxies are plotted as point in regions of low density. The *dashed curve* displays an empirically found division between star-forming galaxies and AGNs, whereas the *dotted curve* shows the division based on theoretical considerations. One frequently considers galaxies below the *dashed curve* (shown in *blue*) as star-formation galaxies, those above the *dotted curve* (*red*) as AGNs, and those between the *curves* (*grey*) as composite objects. Source: L. Trouille, A.J. Barger & C. Tremonti 2011, *The OPTX Project. V. Identifying Distant Active Galactic Nuclei*, ApJ 742, 46, p. 4, Fig. 4. ©AAS. Reproduced with permission

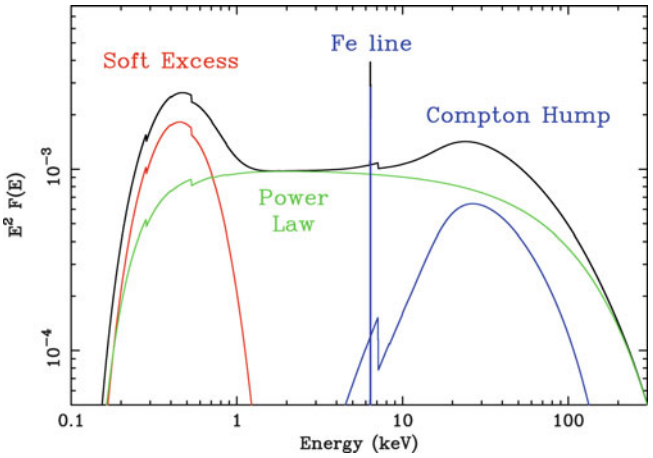

Fig. 5.34 Sketch of a typical X-ray spectrum of an AGN, together with the four components of a standard model, as discussed in the text. Source: A.C. Fabian 2006, ESASP 604, 463, p. 463, Fig. 1. Reproduced by permission of the author

seen from the relativistic iron line shown in Fig. 5.20. In fact, the variability on very short time-scales (see Fig. 5.14) as well as the microlensing results (Fig. 5.27) are a clear indicator of a small extent of the X-ray source.

To a first approximation, the X-ray spectrum in the few keV range is characterized by a power law, $S_\nu \propto \nu^{-\alpha}$, with mean slope $\alpha \sim 0.7$ (see Fig. 5.34). There is a trend that the slope is somewhat steeper for radio-quiet AGNs, and flatter for radio-loud ones. At energies $h_\mathrm{P}\nu \gtrsim 10\,\mathrm{keV}$, the spectrum exceeds the extrapolation of this power law, i.e., it becomes flatter. Towards lower X-ray energies, the spectrum seems to be steeper than the power law; this feature is called the 'soft excess'. The X-ray emission extends up to energies of $\sim$ 100 keV, beyond which there is a spectral cut-off; however, for blazars the spectrum can extend to much higher energies (see below). At low X-ray energies, the observed spectrum is heavily cut off due to photoelectric absorption in the ISM of the Milky Way.

Origin of the X-rays. The decomposition of the continuum spectrum in Fig. 5.34 into a soft excess, a power law, and a Compton hump is based on ideas of how the X-rays in AGNs are generated. Details of the model are yet uncertain; nevertheless, the basic ingredients are probably well established. The origin of the soft excess may be related to the Big Blue Bump, as discussed before; here we will discuss the other components indicated in Fig. 5.34.

Accretion disk corona. Our description of an optically thick, geometrically thin accretion disk in Sect. 5.3.2 led to the picture that at every radius, the disk is characterized by a temperature, and radiates locally almost as a blackbody. It thus can locally be compared to the surface of a star, and stellar atmosphere models have been used to calculate a more accurate spectrum from such disks. In the outer layers of such disks, the gas density is small. Furthermore, we argued that magnetic fields in the disk are probably responsible for the friction which is needed to transport angular momentum through the disk and allow the accretion of material. One may thus expect, in analogy to stars (like in our Sun), that a hot layer of gas forms above the optically thick part of the disk, a corona. The gas in the corona is so thin that it cannot cool efficiently, and hence its temperature can be much higher than that of the disk; perhaps it may even approach the virial temperature, such that $k_\mathrm{B}T \sim GM_\bullet m/r = (2r/r_\mathrm{S})^{-1}\,mc^2$, where m is the mass of a particle. If $m = m_\mathrm{e}$, the temperature of the corona can reach tens or hundreds of keV in the innermost regions of the disk.[10] A sketch of a possible geometry of the corona above the disk is shown in Fig. 5.35.

Inverse Compton effect. As mentioned before, if a photon scatters off an electron at rest, then due to recoil, a (small) fraction of the photon energy is transferred to the electron; this is the Compton effect. However, if the electron has a much larger energy than the photon, part of the electron energy can be transferred to the photon, which is called the

[10] The virial temperature of the protons is much higher, but it is not clear how well electrons and protons are coupled in this hot, thin plasma.

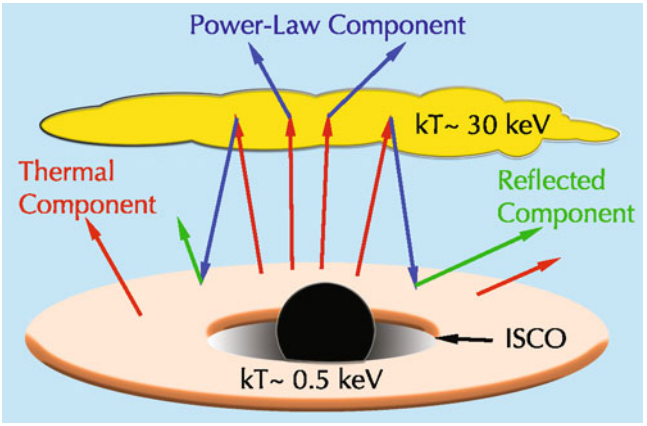

Fig. 5.35 Illustration of an accretion disk and a hot corona, where the possible origin of the various X-ray components of an AGN are indicated; see text for details. Source: L. Gou et al. 2011, *The Extreme Spin of the Black Hole in Cygnus X-1*, ApJ 742, 85, p. 5, Fig. 2. ©AAS. Reproduced with permission

inverse Compton effect. Thus, if photons propagate through a distribution of energetic electrons, they will on average gain energy. For a thermal plasma of temperature T, the mean energy gain of a photon with incoming energy E_γ can be shown to be

$$\Delta E_\gamma = \left(4k_\mathrm{B}T - E_\gamma\right) \frac{E_\gamma}{m_\mathrm{e}c^2} \,, \tag{5.35}$$

so that a photon can either gain or lose energy on average, depending on its energy relative to the electron temperature. If a low-energy photon is scattered multiple times, it can increase its energy is every scattering, until its energy becomes of the same order as the temperature of the gas.

Blackbody photons from the accretion disk propagate through the corona and may scatter off the hot electron distribution. Since the original photon energy is much smaller than $k_\mathrm{B}T$ in the corona, after one scatting is will have an energy

$$E_1 = E_\gamma + \Delta E_\gamma \approx \left(1 + \frac{4k_\mathrm{B}T}{m_\mathrm{e}c^2}\right) E_\gamma \approx E_\gamma \exp\left(\frac{4k_\mathrm{B}T}{m_\mathrm{e}c^2}\right) .$$

A photon may actually scatter more than once, and each time its energy is increased; after N scatterings, the mean photon energy is

$$E_N \approx E_\gamma \exp\left(\frac{4Nk_\mathrm{B}T}{m_\mathrm{e}c^2}\right) .$$

The mean number of scatterings $\langle N \rangle$ a photon undergoes before it leaves the corona depends on the optical depth $\tau_\mathrm{T} = \sigma_\mathrm{T} N_\mathrm{e}$ with respect to electron scattering, where N_e is the electron column density of the corona. If $\tau_\mathrm{T} \lesssim 1$, then $\langle N \rangle \approx \tau_\mathrm{T}$; in the opposite case of $\tau_\mathrm{T} \gg 1$, the mean number of scatterings is $\langle N \rangle \sim \tau_\mathrm{T}^2$. But of course, for a given $\langle N \rangle$, there will be a probability distribution $p(N)$ of the number of scatterings which photons will experience before they leave the corona.

The power-law component. Hence, the emission from the corona will be the sum of photons which scattered N times and which have energy E_N. One can show that the resulting spectrum is a power law in energy whose slope depends on the temperature in the corona and the scattering optical depth τ_T. There is an upper cut-off of this power law, given by $E_\gamma \sim k_\mathrm{B}T$. The power-law component of the X-ray spectrum is interpreted in terms of this thermal Comptonization. The fact that the observed cut-off in the spectrum occurs at $E_\gamma \sim 100\,\mathrm{keV}$ implies that this radiation must originate from the very inner part of the accretion disk, where the virial temperature of the electrons is that high.

The reflection component. Photons scattering in the corona may escape the disk and then form the observable power-law component. However, since the electron scattering is basically isotropic, these photons may be scattered back to the disk. If they have a low energy, they will simply be absorbed. However, if their energy is high, they will scatter inside the disk. Since in this case, the situation $E_\gamma \gg k_\mathrm{B}T$ applies, these photons will lose energy due to scattering. In analogy to what was said above, there is a probability distribution such that photons which hit the disk are scattered N times before they can leave the disk again, and the larger N, the more energy they lose. Furthermore, as they are scattered down in energy, their probability of being absorbed is increased. Thus, this *reflection component* is appreciable only at rather high photon energies. The primary photon energy must be larger than that of the reflection component, indicating that this component must originate from the innermost region of the accretion flow. In addition, line radiation from photoionized gas in the disk, together with the relativistic effects due to relativistic orbital velocities in the inner disk region, may give rise to the soft X-ray excess.

Line emission. Besides this continuum radiation, emission and absorption lines are also found in the X-ray domain, the strongest line being the 6.4 keV iron line (see Sect. 5.3.4). Together with other emission lines at lower energy, it is probably produced in the accretion disk through reflection (note that the energy of this line is much larger than the maximum temperature of a standard accretion disk around SMBHs): A high-energy photon from the corona hits an iron atom in the disk and removes one of its inner (K-shell) electrons. After this K-shell ionization, the electronic K-shell is refilled by a transition of the ion, emitting the 6.4 keV Kα line. The improved sensitivity and spectral resolution of the X-ray telescopes Chandra and XMM-Newton compared to earlier X-ray observatories have greatly advanced the X-ray

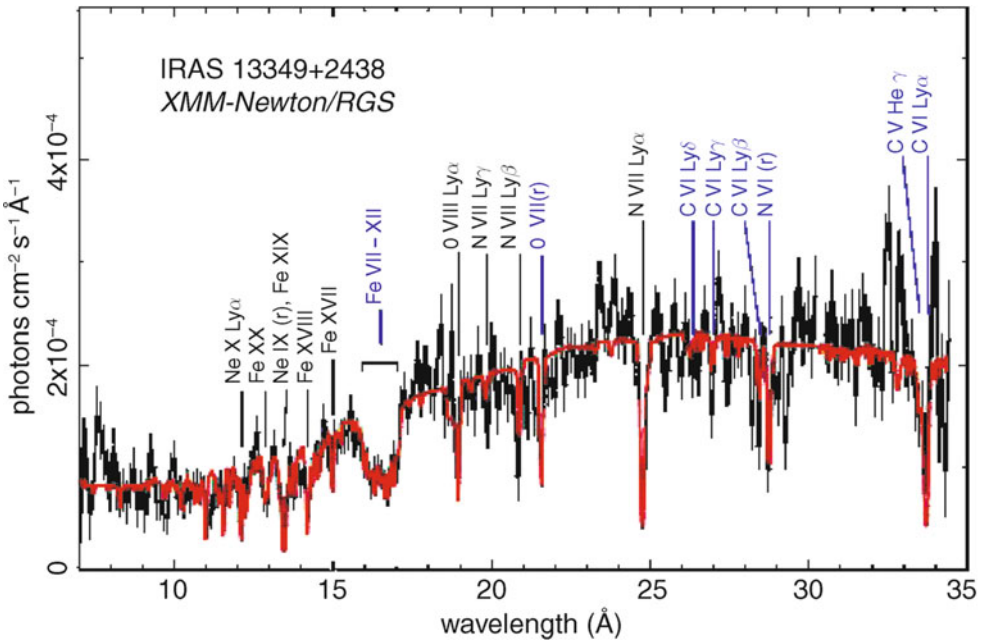

Fig. 5.36 X-ray spectrum of the quasar IRAS 13349 + 2438 ($z = 0.108$), observed by the XMM satellite. Various absorption lines are marked. Source: M. Sako et al. 2001, *Complex resonance absorption structure in the X-ray spectrum of IRAS 13349 + 2438*, A&A 365, L168, p. L170, Fig. 1. ©ESO. Reproduced with permission

spectroscopy of AGNs. Figure 5.36 shows an example of the quality of these spectra.

The X-ray emission of Seyfert 1 and Seyfert 2 galaxies is very different. In the energy range of the ROSAT X-ray satellite ($0.1\,\text{keV} \leq h_P\nu \leq 2.4\,\text{keV}$), significantly more Seyfert 1 galaxies were discovered than Seyfert 2 galaxies. The origin of this was later uncovered by Chandra and XMM-Newton. In contrast to ROSAT, these two satellites are sensitive up to energies of $h_P\nu \sim 10\,\text{keV}$ and they have found large numbers of Seyfert 2 galaxies. However, their spectrum differs from that of Seyfert 1 galaxies because it is cut off towards lower X-ray energies. The spectrum indicates the presence of absorbing material with a hydrogen column density of $\gtrsim 10^{22}\,\text{cm}^{-2}$ and in some cases even orders of magnitude higher.[11] This fact will be used in the context of unified models (Sect. 5.5) of AGNs.

5.4.5 The host galaxy

As the term 'active galactic nuclei' already implies, AGNs are considered the central engine of otherwise quite normal galaxies. This nuclear activity is nourished by accretion of matter onto a SMBH. Since it seems that all galaxies (at least those with a spheroidal component) harbor a SMBH, the question of activity is rather one of accretion rate. What does it take to turn on a Seyfert galaxy, and why are most SMBHs virtually inactive? And by what mechanism is matter brought into the vicinity of the SMBH to serve as fuel?

For a long time it was not clear as to whether QSOs are also hosted in a galaxy. Their high luminosity renders it difficult to identify the surrounding galaxy on images taken from the ground, with their resolution being limited by seeing to $\sim 1''$. In the 1980s, the surrounding galaxies of some QSOs were imaged for the first time, but only with the HST, it became possible to obtain detailed images of QSO host galaxies (see Fig. 5.37) and thus to include them in the class of galactic nuclei. In these investigations, it was also found that the host galaxies of QSOs are often heavily disturbed, e.g., by tidal interaction with other galaxies or even by merging processes. These disturbances of the gravitational potential are considered essential for the gas to overcome the angular momentum barrier and to flow towards the center of the galaxy. At the same time, such disturbances seem to increase the star-formation rate enormously, because starburst galaxies are also often characterized by disturbances and interactions. A close connection seems to exist between AGN activity and nuclear starbursts—in fact, they both have in common that they require the presence of gas in the central region of a galaxy. Optical and NIR images of QSOs (see Fig. 5.37) cannot unambiguously answer the question of whether QSO hosts are spirals or ellipticals.

Today it seems established that the hosts of low-redshift luminous QSOs are predominantly massive and bulge-dominated galaxies. This finding is in good agreement with the fact that the black hole mass in 'normal' galaxies scales with the mass of the spheroidal component of the galaxies,

[11] The absorption of X-rays is due to ionization of metals. Whereas the photoelectric effect is also present for hydrogen, the corresponding cross section for X-rays is small, due to their high energy and the strong frequency dependence of the cross section, $\propto \nu^{-3}$ above the energy threshold. Despite the fact that metals have a much smaller abundance than hydrogen and helium, they dominate the optical depth for X-ray absorption. Nevertheless, the absorber is characterized in terms of a hydrogen column density, implicitly assuming that the gas has Solar metallicity.

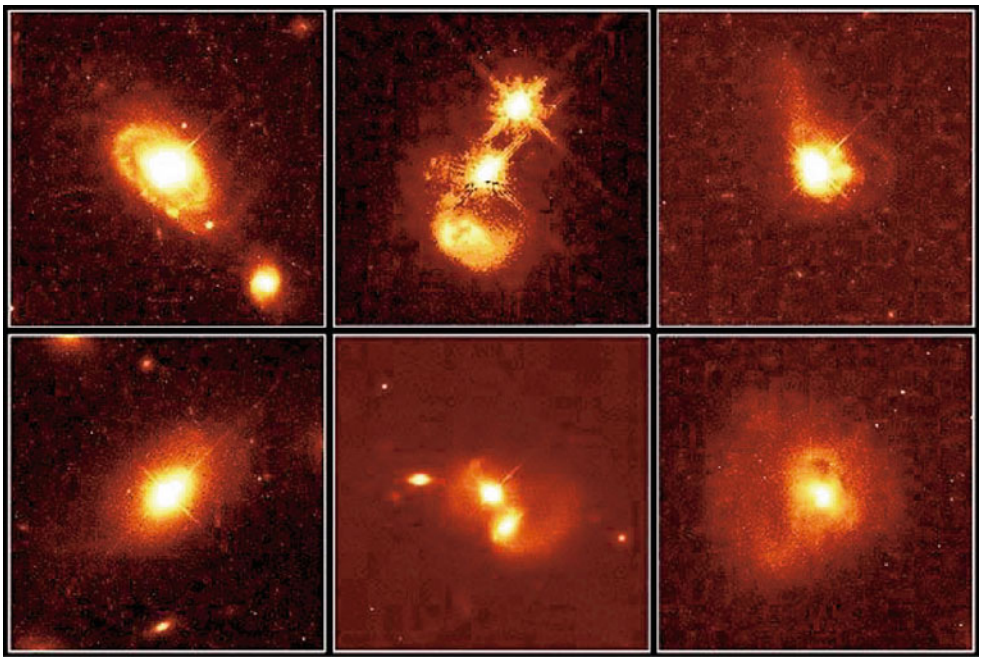

Fig. 5.37 HST images of QSOs. In all cases the host galaxy can clearly be identified, with the QSO itself being visible as a (central) point source in these images. *Top left*: PG 0052 + 251 is located in the center of an apparently normal spiral galaxy. *Bottom left*: PHL 909 seems to be located in the center of a normal elliptical galaxy. *Top center*: the QSO IRAS 04505−2958 is obviously part of a collision of two galaxies and may be provided with 'fuel' by material ripped from the galaxies by tidal forces. Surrounding the QSO core, a region of active star formation is visible. PG 1012 + 008 (*bottom center*) is also part of a pair of merging galaxies. *Top right*: the host galaxy of QSO 0316−346 seems to be about to capture a tidal tail. *Bottom right*: the QSO IRAS 13218 + 0552 seems to be located in a galaxy which just went through a merger process. Credit: J. Bahcall (Institute for Advanced Study, Princeton), M. Disney (University of Wales), and NASA

as the Eddington limit sets a lower bound on $M_\bullet$ for a given luminosity. On the other hand, evolved early-type galaxies are gas-poor, and one might thus not expect that a quiescent elliptical can host a luminous QSO; perhaps the central SMBH is 'switched on' in these galaxies only after some interaction with other galaxies. There are some indications that higher-redshift QSOs are also hosted by massive elliptical galaxies. Figure 5.38 shows three gravitational lens systems where the lens is a (low-luminosity) QSO; from analyzing the lensing geometry, one finds that the overall mass properties (stars plus dark matter) of the three host galaxies are very similar to normal early-type galaxies.

For somewhat less luminous, and thus more abundant AGNs, the situation may be different. Those AGNs seem to have hosts whose color places them mostly in the blue cloud (see Sect. 3.6) or the green valley, but they seem to largely avoid the red sequence. This may indicate a connection between AGNs and star-forming galaxies, both relying on the supply of gas. In contrast, weak AGNs are found preferentially in massive red galaxies, which also seem to be the preferred host for radio-loud AGNs.

Binary QSOs. The connection between the activity of galaxies and the presence of close neighbors is also seen from the clustering properties of QSOs. In surveys for gravitational lens systems, pairs of QSO images have been detected which have angular separations of a few arcseconds and very similar redshifts, but sufficiently different spectra to exclude them being gravitationally lensed images of the same source. The number of binary QSOs thus found is considerably larger than the expectation from the large-scale correlation function of QSOs. This conclusion was further strengthened by an extensive analysis of QSOs in the Sloan Digital Sky Survey. The correlation function of QSOs at separations below $\sim 30h^{-1}$ kpc exceeds that of the extrapolation of the correlation function from larger scales by a factor of 10 or more. Hence it seems that the small-scale clustering of QSOs is very much enhanced, say compared to normal galaxies, which could be due to the triggering of activity by the proximity of the neighbor: in this case, both galaxies attain a perturbed gravitational potential and start to become active.

5.4.6 The black hole mass in AGNs

We now return to the determination of the mass of the central black hole in AGNs. In Sect. 5.3.5, a lower limit on the mass was derived, based on the fact that the luminosity of an

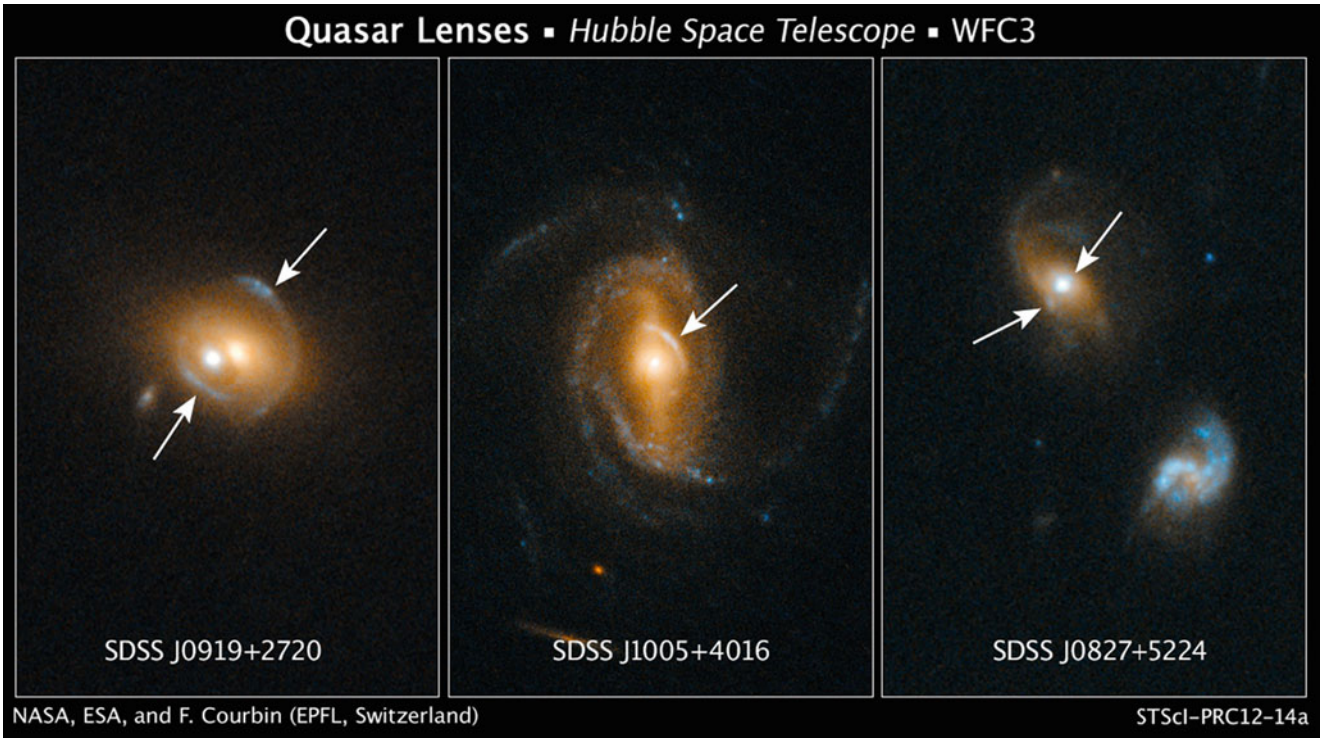

Fig. 5.38 Three QSOs from the SDSS which act as gravitational lenses. Indicated by the *arrows* are the images of star-forming background galaxies ($z_s \sim 0.5$) lensed by the host galaxies of nearby ($z_d \sim 0.2$) QSOs, as obtained from spectroscopy. Analysis of the lensing geometry yields the characteristic mass scale of the lenses, which is very similar to that of early-type gravitational lens galaxies. Credit: NASA, ESA, and F. Courbin (EPFL, Switzerland)

AGN cannot exceed the Eddington luminosity. However, this estimate cannot be very precise, for at least two reasons. The first is related to the anisotropic appearance of an AGN. The observed flux can be translated into a luminosity only on the assumption that the emission from the AGN is isotropic, and we have discussed several reasons why this assumption may not be justified in many cases. Second, we do not have a clear idea what the ratio of AGN luminosity to its Eddington luminosity is. It is clear that this ratio can vary a lot from one system to another. For example, the black hole at the center of our Galaxy could power a luminosity of several 10^{44} erg/s if radiating with the Eddington luminosity—and we know that the true luminosity is many orders of magnitude below this value.

$M_\bullet$ from reverberation mapping. A far more accurate method for estimating the black hole mass in AGNs comes from reverberation mapping which we described in Sect. 5.4.2. The principal quantity that is derived from this technique is the size r of the BLR for a given atomic line or for a given ionization state of a chemical element. Furthermore, the relative line width $\Delta\lambda/\lambda$ can be measured, and can be related to the characteristic velocity dispersion σ in the BLR, $\sigma = c\,\Delta\lambda/\lambda$. Assuming that the gas is virialized, or moving approximately on Keplerian orbits around the black hole, the mass of the latter can be estimated to be

$$M_\bullet = f\, r\, \sigma^2/G\ , \tag{5.36}$$

where $f \sim 1$ for circular orbits and for the observer being located in the plane of the orbit. However, the geometry and kinematics of the BLR may be much more complex than that, and there are orientation effects; the quantity f accounts for this added complexity. If f could be determined, then the black hole mass can be estimated with very reasonable accuracy from reverberation mapping.

One can check whether the functional form of (5.36) is valid, irrespective of the value of f, by studying the relation between time lag Δt (or radius r) and the observed line width *in the same source*, i.e., for a constant value of $M_\bullet$. This was done for the Seyfert 1 galaxy NGC 5548 for which long-term monitoring was carried out (see Fig. 5.31). As shown in Fig. 5.39, one finds indeed a strong correlation between line width and radius, very close to the expected form $\sigma \propto r^{-1/2}$.

The value of f can in principle be estimated from models of the BLR, but these carry considerable uncertainties. A more reliable method is the determination of f empirically. Whereas f cannot be determined for any individual source—and is expected to vary from source to source, e.g.,

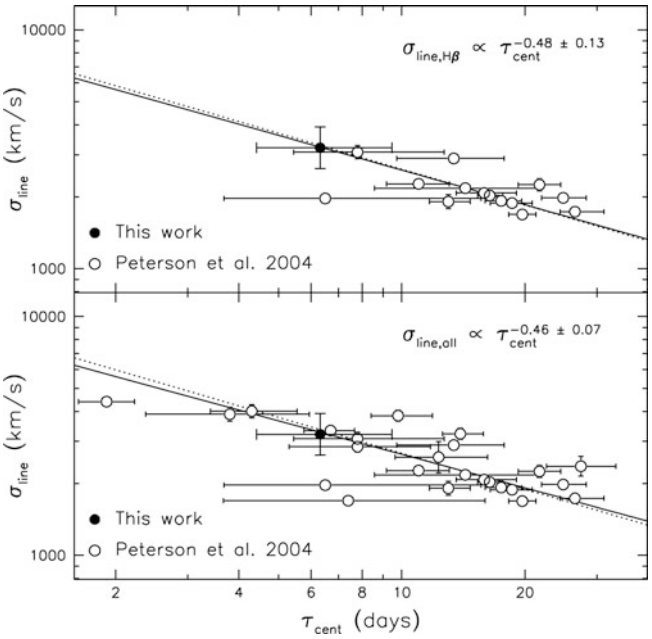

Fig. 5.39 The emission line width as a function of time lags for the Seyfert 1 galaxy NGC 5548. In the *upper panel*, the measurements of the Hβ line are shown, based on the same data used for Fig. 5.31, whereas in the *lower panel* all emission lines are plotted. In both cases, the line width scales with time lag approximately as $\sigma \propto (\Delta t)^{-1/2}$, i.e., obeying a virial relation (5.36). Source: M.C. Bentz et al. 2007, *NGC 5548 in a Low-Luminosity State: Implications for the Broad-Line Region*, ApJ 662, 205, p. 210, Fig. 4. ©AAS. Reproduced with permission

due to different inclination angles relative to the line-of-sight—its mean value can be estimated from a set of AGNs for which the stellar velocity dispersion of the AGN host can be measured. As we have seen in Sect. 3.8.3, there is a well-defined relation between SMBH mass and stellar velocity dispersion. Assuming that the same relation holds for AGNs, then the SMBH mass can be determined from stellar velocity dispersion measurements. This yields a first estimate of $M_\bullet$. From reverberation measurements, (5.36) yields a second, independent mass estimate, which, however, depends on f. Minimizing the difference between these two independent estimates of $M_\bullet$ for a sample of AGNs then yields an estimate for the mean of $f = 5.9 \pm 2.0$; see the left panel of Fig. 5.40.

With the value of f determined, the SMBH mass can now be estimated from (5.36) also for those AGNs for which no measurements of the stellar velocity dispersion σ_* of the host are available. Corresponding results are shown in the right-hand panel of Fig. 5.40, where $M_\bullet$ is plotted as a function of optical continuum luminosity, from which the stellar contribution of the host galaxy was subtracted. Assuming a constant ratio between the optical and bolometric luminosity of 1/9, lines of constant Eddington ratio $\lambda_{\rm Edd} = L_{\rm bol}/L_{\rm edd}$ can be drawn in this figure. We see that the measurements cover a very broad range of luminosities and estimated black hole masses, extending over more than three orders of magnitude. None of the sources has an estimated Eddington ratio larger than unity, but they are concentrated around $\lambda_{\rm Edd} \sim 0.1$. Also, only a single source has an estimated $\lambda_{\rm Edd}$ smaller than 0.01. It thus seems that AGNs with broad emission lines are accreting at a fairly high rate.

However, reverberation mapping is a fairly expensive observing technique. Furthermore, the effort required for this technique increases with AGN luminosity, since the size of the BLR, and thus the time delay and the necessary length of the monitoring campaign, increases with the black hole mass. We might therefore want to look at alternative methods for estimating $M_\bullet$.

$M_\bullet$ from scaling relations. We saw from Fig. 5.40 that nearby AGNs, for which reverberation data are available, obey the same relation (3.49) between $M_\bullet$ and the stellar velocity dispersion σ_* of the host as is obtained for inactive galaxies. This scaling relation then yields a useful estimate of the black hole mass from the stellar velocity dispersion. Unfortunately, even this method cannot be applied to a broad range of AGNs, since the velocity dispersion of stars cannot be measured in AGNs which are either too luminous—since then the nuclear emission outshines the stellar light, rendering spectroscopy of the latter impossible—or too distant, so that a spatial separation of nuclear light from stellar light is no longer possible.

However, the scaling relation between the size of the BLR for a given transition and the optical continuum luminosity, $r \propto L^\beta$, with $\beta \approx 0.6$ (see Fig. 5.31), is very useful for estimating black hole masses. From the continuum luminosity, the size r is estimated using this scaling relation; combined with the measured emission line width σ, the relation (5.36) can be applied. In particular, this method can be extended to luminous and high-redshift sources.

The Eddington ratio. Once an estimate for $M_\bullet$ is obtained, the Eddington luminosity can be calculated and compared with the observed luminosity. The ratio of these two is the Eddington ratio, $\lambda_{\rm Edd} \equiv L_{\rm bol}/L_{\rm Edd}$. For the estimate of $\lambda_{\rm Edd}$, the observed luminosity in the optical band needs to be translated into a bolometric luminosity, which can be done with the help of the average spectral energy distribution of AGNs of a given class. If one can ignore strongly beamed emission, $\lambda_{\rm Edd}$ should be smaller than unity, which seems to be indeed the case as seen from Fig. 5.40. Whereas the various steps in deriving $\lambda_{\rm Edd}$ involve statistical and systematic uncertainties by factors $\gtrsim 2$, the results indicate that most broad-line AGNs have an Eddington ratio of order $\lambda_{\rm Edd} \sim 0.1$.[12]

[12]There might be a trend that radio-loud QSOs have a somewhat larger $\lambda_{\rm Edd}$, but these correlations are controversial and might be based on

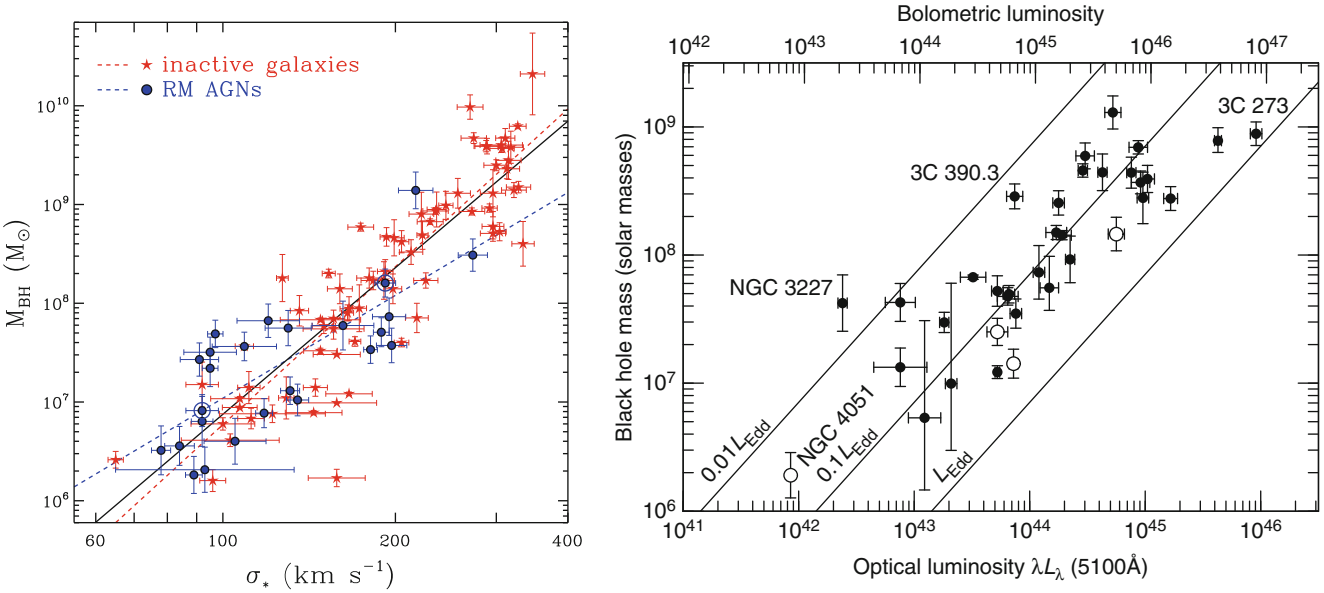

Fig. 5.40 *Left panel*: The SMBH mass $M_\bullet$ as a function of stellar velocity dispersion, for a sample of nearby normal, quiescent galaxies (shown in *red*) and for AGNs for which reverberation mapping data are available (*blue symbols*). The two *dashed lines* correspond to the best power-law fits for these two samples, where the virial factor $f \approx 5.9 \pm 2$ yields the best agreement between these two samples. The *black solid curve* is the best power-law fit to both data sets. *Right panel*: The relation between black hole mass $M_\bullet$ and continuum luminosity, for a sample of 35 AGNs with reverberation mapping, using a value of $f = 5.5$. The upper axis is an estimate of the bolometric luminosity under the assumption that $L_{bol} = 9\lambda L_\lambda(5100\,\text{Å})$. *Diagonal lines* indicate constant Eddington ratio $\lambda_{Edd} = L_{bol}/L_{Edd}$ of 0.01, 0.1 and 1. Source: Left: J.-H. Woo et al. 2013, *Do quiescent and active galaxies have different M_{BH}-σ relations?*, arXiv:1305.2946, Fig. 4. Reproduced by permission of the author. Right: B.M. Peterson et al. 2004, *Central Masses and Broad-Line Region Sizes of Active Galactic Nuclei. II. A Homogeneous Analysis of a Large Reverberation-Mapping Database*, ApJ 613, 682, p. 696, Fig. 16. ©AAS. Reproduced with permission

The fact that λ_{Edd} is confined to a fairly narrow interval implies that the luminosity of a QSO can be used to estimate $M_\bullet$, just by setting $M_\bullet = \lambda_{Edd} M_{Edd}(L)$. This mass estimate has a statistical uncertainty of at least a factor of three in individual sources, but requires only the measurement of the continuum luminosity. However, it is not clear whether the mean Eddington ratio is approximately constant in redshift; it may well be that there is a cosmological evolution of the properties of AGNs. We will come back to this point in Sect. 10.6.2.

The Galactic center black hole. The Eddington ratio of the SMBH in the Galactic center is many orders of magnitude smaller than unity, at least at the current epoch; in fact, with its total luminosity of $5 \times 10^{36}\,\text{erg/s}$, $\lambda_{Edd} \sim 10^{-8}$. Such a small value indicates that the SMBH in our Galaxy is starved; the accretion rate must be very small. However, one can estimate a minimum mass rate with which the SMBH in the Galactic center is fed, by considering the mass-loss rate of the stars in its vicinity. This amounts to $\sim 10^{-4} M_\odot/\text{yr}$, enough material to power an accretion flow with $L \sim 10^{-2} L_{Edd}$. The fact that the observed luminosity is so much smaller than this value leads to two implications. The first of these is that there must be other modes of accretion which are far less efficient than that of the geometrically thin, optically thick accretion disk. Such models for accretion flows were indeed developed, such as the ADAF briefly described in Sect. 5.3.2. The second conclusion is that the central mass concentration must indeed be a black hole—a black hole is the only object which does not have a surface. If, for example, one would postulate a hypothetical object with $M \sim 4 \times 10^6 M_\odot$ which has a hard surface (like a scaled-up version of a neutron star), the accreted material would fall onto the surface, and its kinetic and internal energy would be deposited there. Hence, this surface would heat up and radiate thermally. Since we have strict upper limits on the radius of the object, coming from mm-VLBI observations, we can estimate the minimum luminosity such a source would have. This estimate is again several orders of magnitude larger than the observed luminosity from Sgr A*, firmly ruling out the existence of such a solid surface.

Evolution of the $M_\bullet$ scaling relations. As we have seen in Sect. 3.8.3, the black hole mass in normal, nearby galaxies is correlated with the bulge (or spheroidal) luminosity. As this component of galaxies consists of an old stellar population, its luminosity is very closely related to its stellar mass. Esti-

selection effects. On the other hand, radio galaxies have a lower value of λ_{Edd} than QSOs.

mating the black hole mass from the continuum luminosity of the QSOs, and observing the spheroidal luminosity of their host galaxies (which requires the high angular resolution of HST), one can investigate whether such a scaling relation already existed at earlier epochs, i.e., at high redshifts.

The results from such studies indicate that the ratio of black hole mass and stellar mass of the spheroidal component of the host galaxy evolves with redshift, in the sense that $M_\bullet/M_*$ was larger in the past. Furthermore, the scatter in the scaling relation increases with redshift. There is some debate on how strong these evolutionary effects are: Whereas evolution as strong as $M_\bullet/M_* \propto (1+z)^2$ is claimed, the scatter in the relation impacts on the sample of QSOs with well measured stellar and black hole masses by selection effects which bias the apparent redshift evolution. A milder evolution of the form $M_\bullet/M_* \propto (1+z)^{0.7}$ was deduced from QSOs in the COSMOS field, when these selection effects are accounted for.

5.5 Family relations of AGNs

5.5.1 Unified models

In Sect. 5.2, several different types of AGNs were mentioned. We saw that many of their properties are common to all types, but also that there are considerable differences. Why are some AGNs seen as broad line radio galaxies, others as BL Lac objects? The obvious question arises as to whether the different classes of AGNs consist of rather similar objects which differ in their appearance due to geometric or light propagation effects, or whether more fundamental differences exist. At the beginning of Sect. 5.2, we summarized a classification scheme for AGNs, called unified model. In this section, we will collect the various differences and similarities of the various classes of AGNs, and provide evidence for the unified scheme, which will be explained in more detail below.

Common properties. Common to all AGNs is a SMBH in the center of the host galaxy, the supposed central engine, and also an accretion disk that is feeding the black hole. This suggests that a classification can be based on $M_\bullet$ and the accretion rate $\dot{m}$, or perhaps more relevantly the ratio $\dot{m}/\dot{m}_{\rm edd}$. $M_\bullet$ defines the maximum (isotropic) luminosity of the SMBH in terms of the Eddington luminosity, and the ratio $\dot{m}/\dot{m}_{\rm edd}$ describes the accretion rate relative to its maximum value. Furthermore, the observed properties, in particular the seemingly smooth transition between the different classes, suggest that radio-quiet QSOs and Seyfert 1 galaxies basically differ only in their central luminosity. From this, we would then deduce that they have a similar value of $\dot{m}/\dot{m}_{\rm edd}$ but differ in $M_\bullet$. An analogous argument may be valid for the transition from BLRGs to radio-loud quasars.

The difference between these two classes may be due to the nature of the host galaxy. Radio galaxies (and maybe radio-loud quasars?) are situated in elliptical galaxies, Seyfert nuclei (and maybe radio-quiet quasars?) preferentially in spirals. A correlation between the luminosity of the AGN and that of the host galaxy also seems to exist. This is to be expected if the luminosity of the AGN is strongly correlated with the respective Eddington luminosity, because of the correlation between the SMBH mass in normal galaxies and the properties of the galaxy (Sect. 3.8.3). Another question is how to fit blazars and Seyfert 2 galaxies into this scheme.

Anisotropic emission. In the context of the SMBH plus accretion disk model, another parameter exists that affects the observed characteristics of an AGN, namely the inclination, i.e., the angle between the rotation axis of the disk and the direction from which we observe the AGN. We should mention that in fact there are many indications that the radiation of an AGN is not isotropic and thus its appearance depends on the viewing angle. Among these are the observed ionization cones in the NLR (see Fig. 5.32) and the morphology of the radio emission, as the radio lobes define a preferred direction. Furthermore, our discussion of superluminal motion has shown that the observed superluminal velocities are possible only if the direction of motion of the source component is close to the direction of the line-of-sight. The X-ray spectrum of many AGNs shows intrinsic (photoelectric) absorption caused by high column density gas, where this effect is mainly observed in Seyfert 2 galaxies. Because of these clear indications it seems obvious to examine the dependence of the appearance of an AGN on the viewing direction. For example, the observed difference between Seyfert 1 and Seyfert 2 galaxies may simply be due to a different orientation of the AGN relative to the line-of-sight.

Broad emission lines in polarized light. In fact, another observation of anisotropic emission provides a key to understanding the relation between AGN types, which supports the above idea. The galaxy NGC 1068 has no visible broad emission lines and is therefore classified as a Seyfert 2 galaxy. Indeed, it is considered an archetype of this kind of AGN. However, the optical spectrum of NGC 1068 in polarized light shows broad emission lines (Fig. 5.41) such as one would find in a Seyfert 1 galaxy. Obviously the galaxy must have a BLR, but it is only visible in polarized light. The photons that are emitted by the BLR are initially unpolarized. But polarization may be induced through scattering of the

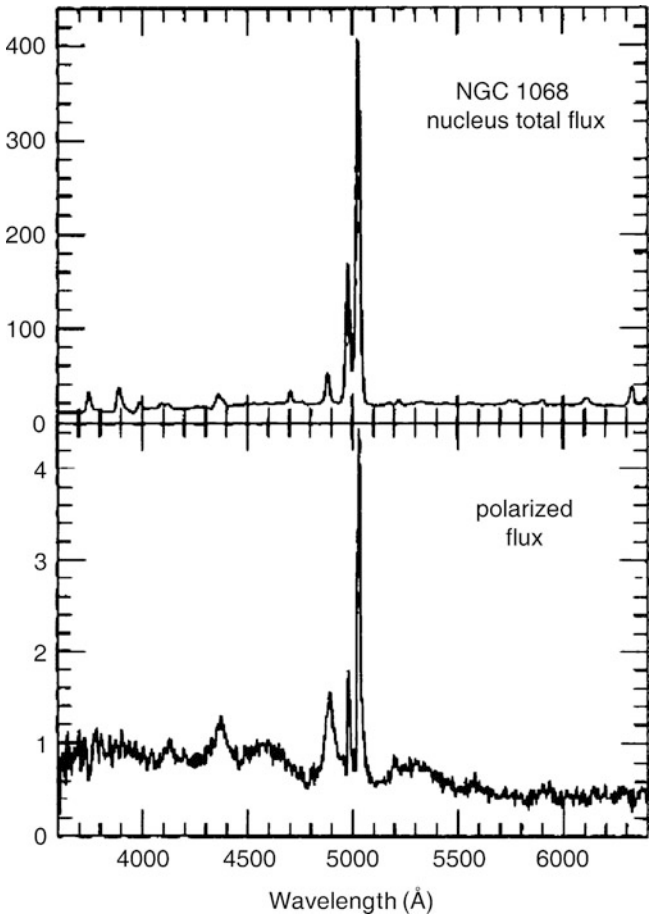

Fig. 5.41 Spectrum of the Seyfert 2 galaxy NGC 1068. The *top panel* displays the total flux which, besides the continuum, also shows narrow emission lines, in particular [OIII] at $\lambda = 5007$ Å and $\lambda = 4959$ Å. However, in polarized light (*bottom panel*), broad emission lines (like Hβ λ4861 Å and Hγ λ4340 Å) typical of a Seyfert 1 galaxy are also visible. Therefore, it is concluded that the BLR becomes visible in light polarized via scattering; the BLR is thus visible only indirectly. Source: J.S. Miller et al. 1991, *Multidirectional views of the active nucleus of NGC 1068*, ApJ 378, 47, p. 50, Fig. 6. ©AAS. Reproduced with permission

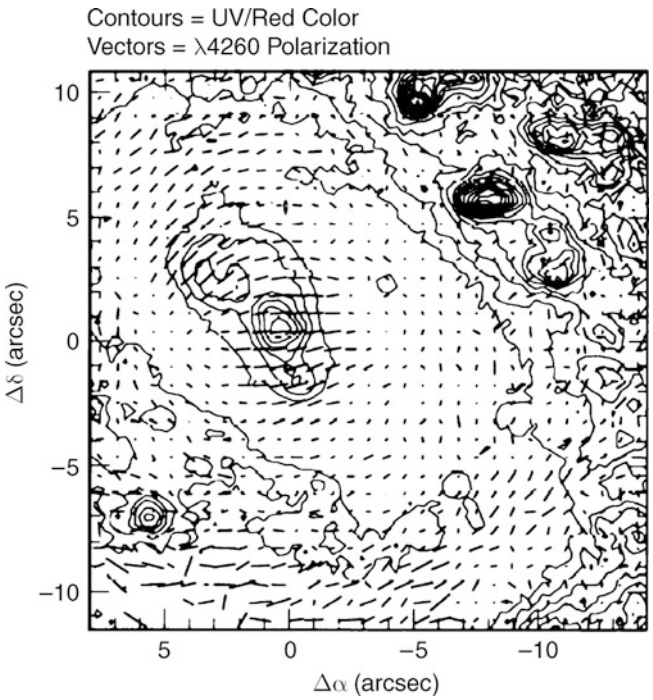

Fig. 5.42 The contours show the color of the optical emission in the Seyfert 2 galaxy NGC 1068, namely the flux ratio in the U- and R-bands. The sticks indicate the strength and orientation of the polarization in B-band light. The center of the galaxy is located at $\Delta\alpha = 0 = \Delta\delta$. At its bluest region (*center left*), the polarization of the optical emission is strongest and is perpendicular to the direction to the center of the galaxy; this is the direction of polarization expected for local scattering by electrons. Hence, where the scattering is strongest, the largest fraction of direct light from the AGN is also observed, and the optical spectrum of AGNs is considerably bluer than the stellar light from galaxies. Source: R.W. Pogge & M.M. De Robertis 1993, *Extended near-ultraviolet continuum emission and the nature of the polarized broad-line Seyfert 2 galaxies*, ApJ 404, 563, p. 568, Fig. 4. ©AAS. Reproduced with permission

light, where the direction perpendicular to the directions of incoming and scattered photons defines a preferred direction, which then yields the polarization direction.

The interpretation of this observation (see Fig. 5.12) now is that NGC 1068 has a BLR but our direct view of it is obscured by absorbing material. However, this absorber does not fully engulf the BLR in all directions but only within a solid angle of $<4\pi$ as seen from the central core. If photons from the BLR are scattered by dust or electrons in a way that we are able to observe the scattered radiation, then the BLR would be visible in this scattered light. Direct light from the AGN completely outshines the scattered light, which is the reason why we cannot identify the latter in the total flux. By scattering, however, this radiation is also polarized. Thus in observations made in polarized light, the (unpolarized) direct radiation is suppressed and the BLR becomes visible in the scattered light.[13]

This interpretation is additionally supported by a strong correlation of the spatial distribution of the polarization and the color of the radiation in NGC 1068 (see Fig. 5.42). We can conclude from this that the differences between Seyfert 1 and Seyfert 2 galaxies originate in the orientation of the accretion disk and thus of the absorbing material relative to the line-of-sight.

From the abundance ratio of Seyfert 1 to Seyfert 2 galaxies (which is about 1:2), the fraction of solid angle in which the view to the BLR is obscured, as seen from the AGN, can be estimated. This ratio then tells us that about 2/3 of the

[13]Not all Seyfert 2 galaxies show broad emission lines in polarized flux, which may be either due to the fact that there is no appropriate scattering medium which makes the BLR visible for us, or that some of the sources intrinsically lack a BLR. The discussion about the possible existence of such 'true Seyfert 2' objects has not yet come to a clear conclusion.

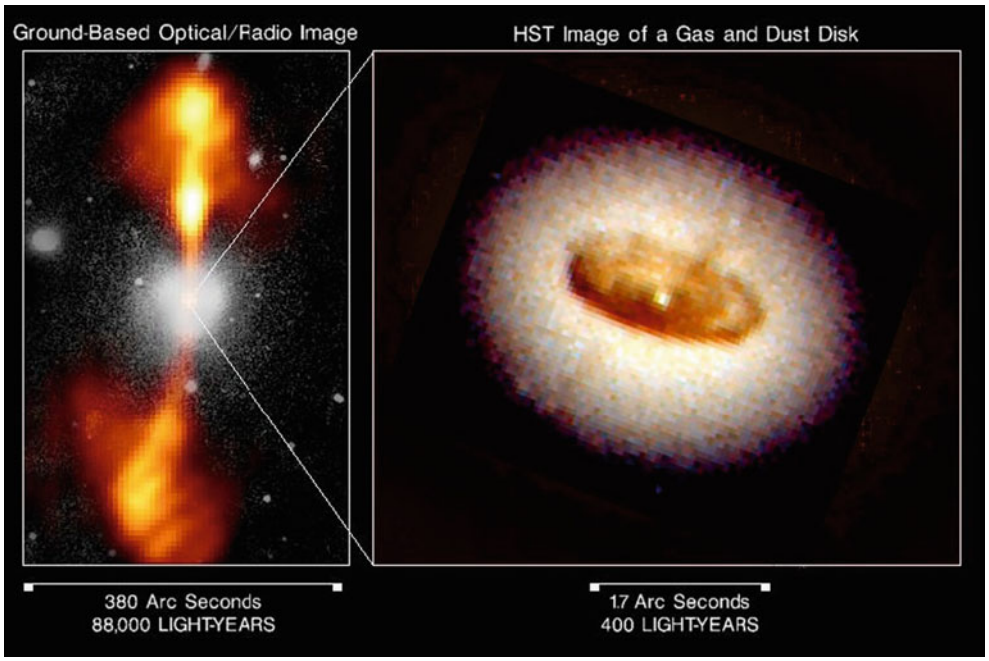

Fig. 5.43 The elliptical galaxy NGC 4261. The *left-hand panel* shows an optical image of this galaxy together with the radio emission (shown in *orange*). An HST image showing the innermost region of the galaxy is shown on the *right*. The jet is virtually perpendicular to the central disk of gas and dust, which is in agreement with the theoretical picture in the context of a unification model. Credit: National Radio Astronomy Observatory, California Institute of Technology, Walter Jaffe/Leiden Observatory, Holland Ford/JHU/STScI, and NASA

solid angle is covered by an absorber. Such a blocking of light may be caused by dust imbedded in a gas distribution. It is assumed that the dust is located in the plane of the accretion disk in the form of a thick 'torus' (see Figs. 5.12 and 5.43 for a view of this geometry).

Originally, the 'torus' was imagined as a rather big and smooth distribution of gas and dust. However, from the properties of the infrared emission from QSOs, it is now believed to be considerably smaller and instead clumpy. The 'torus' is now visioned as a region around the central AGN, perhaps an order-of-magnitude larger than the BLR, which is filled with rather dense clouds. The dust in the clouds absorbs radiation from the AGN, gets heated, and reradiates the energy in form of infrared radiation. Models assuming a size distribution of the clouds can reproduce the observed spectral shapes in the infrared spectral regime. Whereas there is considerable variation between individual objects, one finds that some 25 % of the total power emitted by an AGN emerges in the infrared regime. Taken at face value, this would imply that the material in the torus blocks the light in a quarter of all directions; however, this estimate is probably too simple. Instead, one expects a broad probability distribution of absorbing optical depths for light rays traversing the torus region.

Search for Type 2 QSOs. If the difference between Seyfert galaxies of Type 1 and Type 2 is caused merely by their orientation, and if likewise the difference between Seyfert 1 galaxies and QSOs is basically one of absolute luminosity, then the question arises as to whether a luminous analog for Seyfert 2 galaxies exists, a kind of Type 2 QSO. Until around 2000, such Type 2 QSOs had not been observed, from which it was concluded that either no dust torus is present in QSOs due to the high luminosity (and therefore no Type 2 QSOs exist) or that Type 2 QSOs are not easy to identify.

This question is now settled: the current X-ray satellites Chandra and XMM-Newton have identified the population of Type 2 QSOs. Due to the high column density of hydrogen which is distributed in the torus together with the dust, low-energy X-ray radiation is almost completely absorbed by the photoelectric effect if the line-of-sight to the center of these sources passes through the obscuring torus. These sources were therefore not visible for ROSAT ($E \leq 2.4\,\mathrm{keV}$), but the energy ranges of Chandra and XMM-Newton finally allowed the X-ray detection and identification of these Type 2 QSOs.

Other candidates for Type 2 QSOs are the ultra-luminous infrared galaxies (ULIRGs), in which extreme IR-luminosity is emitted by large amounts of warm dust which is heated either by very strong star formation or by an AGN. Since ULIRGs have total luminosities comparable to QSOs, the latter interpretation is possible. In fact, distinguishing between the two possibilities is not easy for individual ULIRGs, and in many sources indicators of both strong star formation and non-thermal emission (e.g., in the form of X-ray emission) are found. This discovery indicates that in many objects, the processes of strong star formation and accretion onto a SMBH are linked. For both processes, large amounts of gas are necessary, and the fact that both starburst galaxies and AGNs are often found in interacting galaxies, where the disturbance in the gravitational field provides the conditions for a gas flow into the center of the galaxy, suggests a link between the two phenomena.

Next we will examine how blazars fit into this unified scheme. A first clue comes from the fact that all blazars are radio sources. Furthermore, in our interpretation of superluminal motion (Sect. 5.3.3) we saw that the appearance and apparent velocity of the central source components depend on the orientation of the source with respect to us, and that it requires relativistic velocities of the source components. To obtain an interpretation of the blazar phenomenon that fits into the above scheme, we first need to discuss an effect that results from Special Relativity.

5.5.2 Beaming

Due to relativistic motion of the source components relative to us, another effect occurs, known as *beaming*. Because of beaming, the relation between source luminosity and observed flux from a moving source depends on its velocity with respect to the observer. One aspect of this phenomenon is the Doppler shift in frequency space: the measured flux at a given frequency is different from that of a non-moving source because the measured frequency corresponds to a Doppler-shifted frequency in the rest-frame of the source. Another effect described by Special Relativity is that a moving source which emits isotropically in its rest-frame has an anisotropic emission pattern, with the angular distribution depending on its velocity. The radiation is emitted preferentially in the direction of the velocity vector of the source (thus, in the forward direction), so that a source will appear brighter if it is moving towards the observer. In Sect. 4.3.2, we already mentioned the relation (4.46) between the radiation intensity in the rest-frame of a source and in the system of the observer. Due to the strong Doppler shift, this implies that a source moving towards us appears brighter by a factor (*Doppler factor*)

$$\mathcal{D}_+ = \left(\frac{1}{\gamma(1 - \beta \cos\phi)} \right)^{2+\alpha} \qquad (5.37)$$

than the source at rest, where α is the spectral index. Here, $\beta = v/c$, ϕ is the angle between the velocity vector of the source component and the line-of-sight to the source, and the Lorentz factor $\gamma = (1 - \beta^2)^{-1/2}$ was already defined in Sect. 5.3.3. Even at modest relativistic velocities ($\beta \sim 0.9$) this can already be a considerable factor, i.e., the radiation from the relativistic jet may appear highly amplified. Another consequence of beaming is that if a second jet exists which is moving away from us (the so-called counter-jet), its radiation will be weakened by a factor

$$\mathcal{D}_- = \left(\frac{1}{\gamma(1 + \beta \cos\phi)} \right)^{2+\alpha} \qquad (5.38)$$

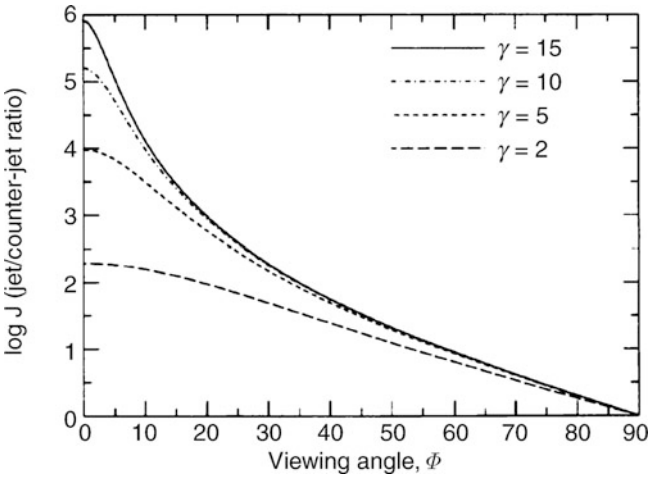

Fig. 5.44 The logarithm of the flux ratio of jet and counter-jet (5.39) is plotted as a function of the angle ϕ for different values of the Lorentz factor γ. Even at relatively small values of γ, this ratio is large if ϕ is close to 0, but even at $\phi \sim 30°$ the ratio is still appreciable. Hence the plot shows the Doppler favoritism and explains why, in most compact radio AGNs, one jet is visible but the counter-jet is not. Source: C.M. Urry & P. Padovani 1995, *Unified Schemes for Radio-Loud Active Galactic Nuclei*, PASP 107, 803, p. 839, Fig. 22. ©ASP. Reproduced with permission

relative to the stationary source. Obviously, $\mathcal{D}_-$ can be obtained from $\mathcal{D}_+$ by replacing ϕ by $\phi + \pi$, since the counter-jet is moving in the opposite direction. In particular, the flux ratio of jet and counter-jet is

$$\frac{\mathcal{D}_+}{\mathcal{D}_-} = \left(\frac{1 + \beta \cos\phi}{1 - \beta \cos\phi} \right)^{2+\alpha} , \qquad (5.39)$$

and this factor may easily be a hundred or more (Fig. 5.44). The large flux ratio (5.39) for relativistic jets is the canonical explanation for VLBI jets being virtually always only one-sided. This effect is also denoted as 'Doppler favoritism'—the jet pointing towards us is observed preferentially because of the beaming effect and the resulting amplification of its flux.

Beaming and the blazar phenomenon. If we observe a source from a direction very close to the jet axis and if the jet is relativistic, its radiation can outshine all other radiation from the AGN because $\mathcal{D}_+$ can become very large in this case. Especially if the beamed radiation extends into the optical/UV part of the spectrum, the line emission may also become invisible relative to the jet emission, and the source will appear to us as a BL Lac object. If the line radiation is not outshined completely, the source may appear as an OVV. The synchrotron nature of the optical light is also the explanation for the optical polarization of blazars since synchrotron emission can be polarized, in contrast to thermal emission.

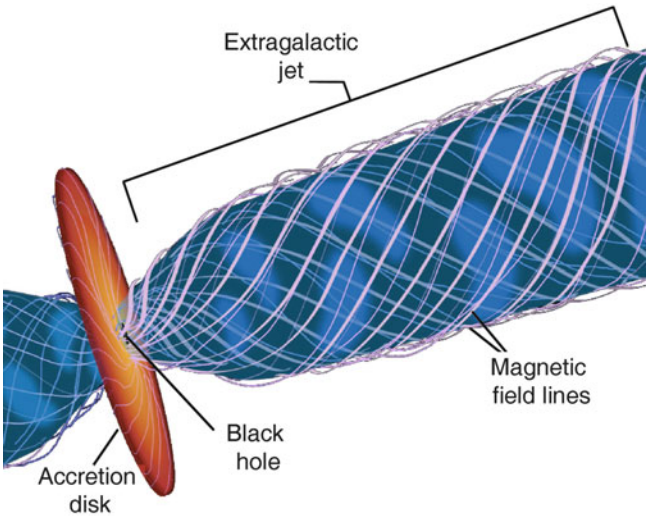

Fig. 5.45 Illustration of the relativistic jet model. The acceleration of the jet to velocities close to the speed of light is probably caused by a combination of very strong gravitational fields in the vicinity of the SMBH and strong magnetic fields which are rotating rapidly because they are anchored in the accretion disk. Shock fronts within the jet lead to acceleration processes of relativistic electrons, which then strongly radiate and become visible as 'blobs' in the jets. By rotation of the accretion disk in which the magnetic field lines are frozen in, the field lines obtain a characteristic helical shape. It is supposed that this process is responsible for the focusing (collimation) of the jet. Credit: NASA/ESA and Ann Feild, Space Telescope Science Institute

The strong beaming factor also provides an explanation for the rapid variability of blazars. If the velocity of the emitting component is close to the speed of light, $\beta \lesssim 1$, even small changes in the jet velocity or its direction may noticeably change the Doppler factor $\mathcal{D}_+$. Such small changes in the direction are expected because there is no reason to expect a smooth outflow of material along the jet at constant velocity. In addition, we argued that, very probably, magnetic fields play an important role in the generation and collimation of jets. These magnetic fields are toroidally spun-up, and emitting plasma can, at least partially, follow the field lines along helical orbits (see Fig. 5.45).

Hence beaming can explain the dominance of radiation from the jet components if the gas is relativistic, and also the absence or relative weakness of emission lines. At the same time, it provides a plausible scenario for the strong variability of blazars. The relative strength of the core emission and the extended radio emission depends heavily on the viewing direction. In blazars, a dominance of the core emission is expected, which is exactly what we observe.

5.5.3 Beaming on large scales

A consequence of this model is that the jets on kpc scales, which are mainly observed by the VLA, also need to be at least semi-relativistic: kpc-scale jets are in most cases also one-sided, and they are always on the same side of the core as the VLBI jet on pc scales. Thus, if the one-sidedness of the VLBI jet is caused by beaming and the corresponding Doppler favoritism of an otherwise intrinsically symmetric source, the one-sidedness of large-scale jets should have the same explanation, implying relativistic velocities for them as well. These do not need to be as close to c as those of the components that show superluminal motion, but their velocity should also be at least a few tenths of the speed of light. In addition, it follows that the kpc-scale jet is moving towards us and is therefore closer to us than the core of the AGN; for the counter-jet we have the opposite case. This prediction can be tested empirically, and it was confirmed in polarization measurements. Radiation from the counter-jet crosses the ISM of the host galaxy, where it experiences additional Faraday rotation (see Sect. 2.3.4). It is in fact observed that the Faraday rotation of counter-jets is systematically larger than that of jets. This can be explained by the fact that the counter-jet is located behind the host galaxy and we are thus observing it through the gas of that galaxy.

5.5.4 Jets at higher frequencies

Optical jets. In Sect. 5.1.2, we discussed the radio emission of jets, and Sect. 5.3.3 described how their relativistic motion is detected from their structural changes, i.e., superluminal motion. However, jets are not only observable at radio frequencies; they also emit at much shorter wavelengths. Indeed, the first two jets were detected in optical observations, namely in QSO 3C273 (Fig. 5.46) and in the radio galaxy M87 (Fig. 5.47), as a linear source structure pointing radially away from the core of the respective galaxy. With the commissioning of the VLA (Fig. 1.26) as a sensitive and high-resolution radio interferometer, the discovery and examination of hundreds of jets at radio frequencies became possible.

The HST, with its unique angular resolution, has detected numerous jets in the optical (see also Fig. 5.17). They are situated on the same side of the corresponding AGNs as the main radio jet. Optical counterparts of radio counter-jets have not been detected thus far. Optical jets are always shorter, narrower, and show more structure than the corresponding radio jets. The spectrum of optical jets follows a power law (5.2) similar to that in the radio domain, with an index α that describes, in general, a slightly steeper spectrum. In some cases, linear polarization in the optical jet radiation of $\sim 10\,\%$ was also detected. If we also take into account that the positions of the knots in the optical and in the radio jets agree very well, we inevitably come to the conclusion that the optical radiation is also synchrotron emission. This

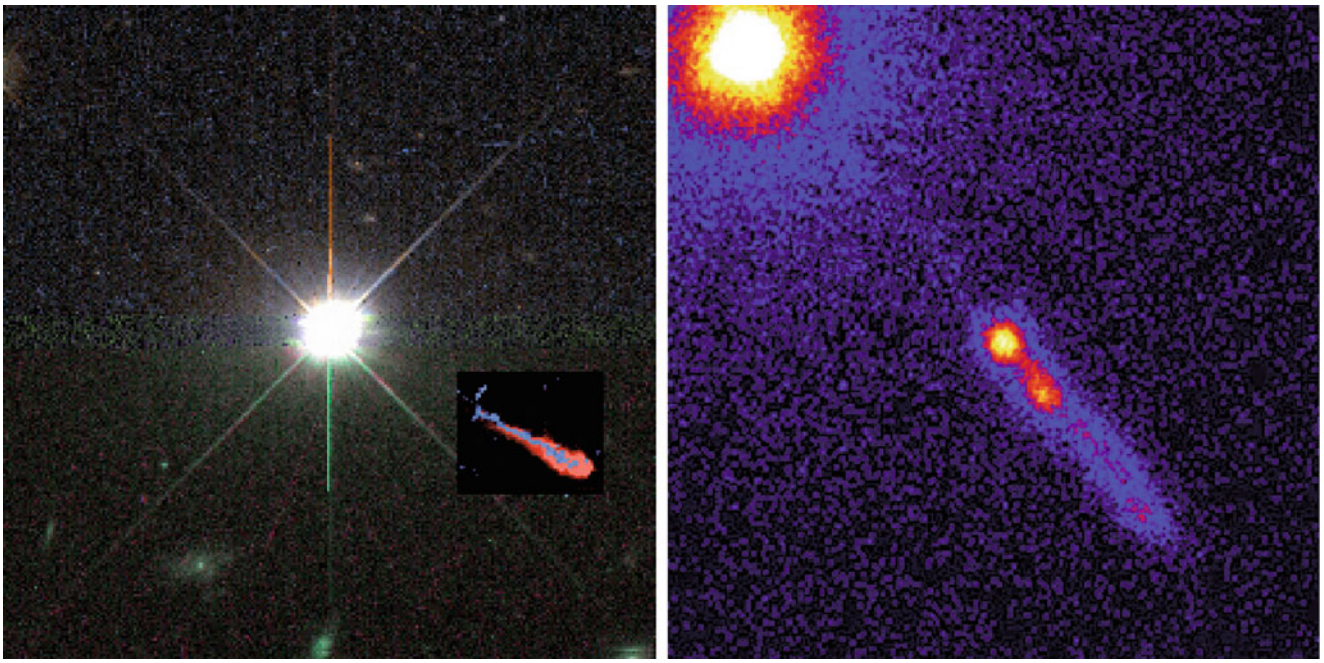

Fig. 5.46 Jets are visible not only in the radio domain but in some cases also at other wavelengths. *On the left*, an HST image of the quasar 3C273 is shown, with the point-like quasar in the center and jet-shaped optical emission (shown in *blue*) that spatially coincides with the radio jet (displayed in *red*). *On the right*, an X-ray image of this quasar taken by the Chandra satellite. The jet is also visible at very high energies. Credit: J. Bahcall/IAS Princeton and NASA; NASA/CXC/SAO/H. Marshall et al.

conclusion is further supported by a nearly constant flux ratio of radio and optical radiation along the jets.

As was mentioned in Sect. 5.1.3, the relativistic electrons that produce the synchrotron radiation lose energy by emission. In many cases, the cooling time (5.6) of the electrons responsible for the radio emission is longer than the time of flow of the material from the central core along the jet, in particular if the flow is (semi-)relativistic. It is thus possible that relativistic electrons are produced or accelerated in the immediate vicinity of the AGN and are then transported away by the jet. This is not the case for those electrons producing the optical synchrotron radiation, however, because the cooling time for emission at optical wavelengths is only $t_{\rm cool} \sim 10^3\,(B/10^{-4}\,{\rm G})^{-3/2}\,{\rm yr}$.[14] Even if the electrons responsible for the optical emission are transported in a (semi-)relativistic jet, they cannot travel more than a distance of ~ 1 kpc before losing their energy. The observed length of optical jets is much larger, though. For this reason, the corresponding electrons cannot be originating in the AGN itself but instead must be produced locally in the jet. The knots in the jets, which are probably shock fronts in the outflow, are thought to represent the location of the acceleration of relativistic particles. Quantitative estimates of the cooling time are hampered by the unknown beaming factor (5.37). Since optical jets are all one-sided, and in most cases observed in radio sources with a flat spectrum, a very large beaming factor is generally assumed. Transforming back into the rest-frame of the electrons yields a lower frequency and a lower luminosity. Since the latter is utilized for estimating the strength of the magnetic fields (by assuming equipartition of energy, for instance), this also changes the estimated cooling time.

The cooling time of electrons emitting at radio frequencies is much longer. The energy-dependent cooling time causes a spectral break in the electron distribution, which shows up accordingly in the synchrotron spectrum, as can be seen in Fig. 5.48 for the case of Centaurus A. The break frequency is highest close to the nucleus, and decreases as one moves away from it.

X-ray radiation of jets. The Chandra satellite discovered that many of the jets which had been identified in the radio are also visible in X-ray light (see Figs. 5.46, 5.48 and 5.50); in fact, currently about 50 X-ray jets are known which are spatially related to corresponding radio structures. This discovery came as a real surprise, since the strong correlation of the spatial distribution of radio, optical, and X-ray emission implies that they must all originate from the same regions in the jets, i.e., that the origins of the emission must be linked to each other. As we have discussed, radio and optical radiation originate from synchrotron emission, the

[14] This dependence of the cooling time $t_{\rm cool} = E/\dot{E}$ on the magnetic field strength follows from (5.3), which at a fixed frequency yields $\gamma = E/(m_{\rm e}c^2) \propto B^{-1/2}$, and the energy loss (5.5), which reads $\dot{E} \propto \gamma^2 B^2 \propto B$ at fixed frequency.

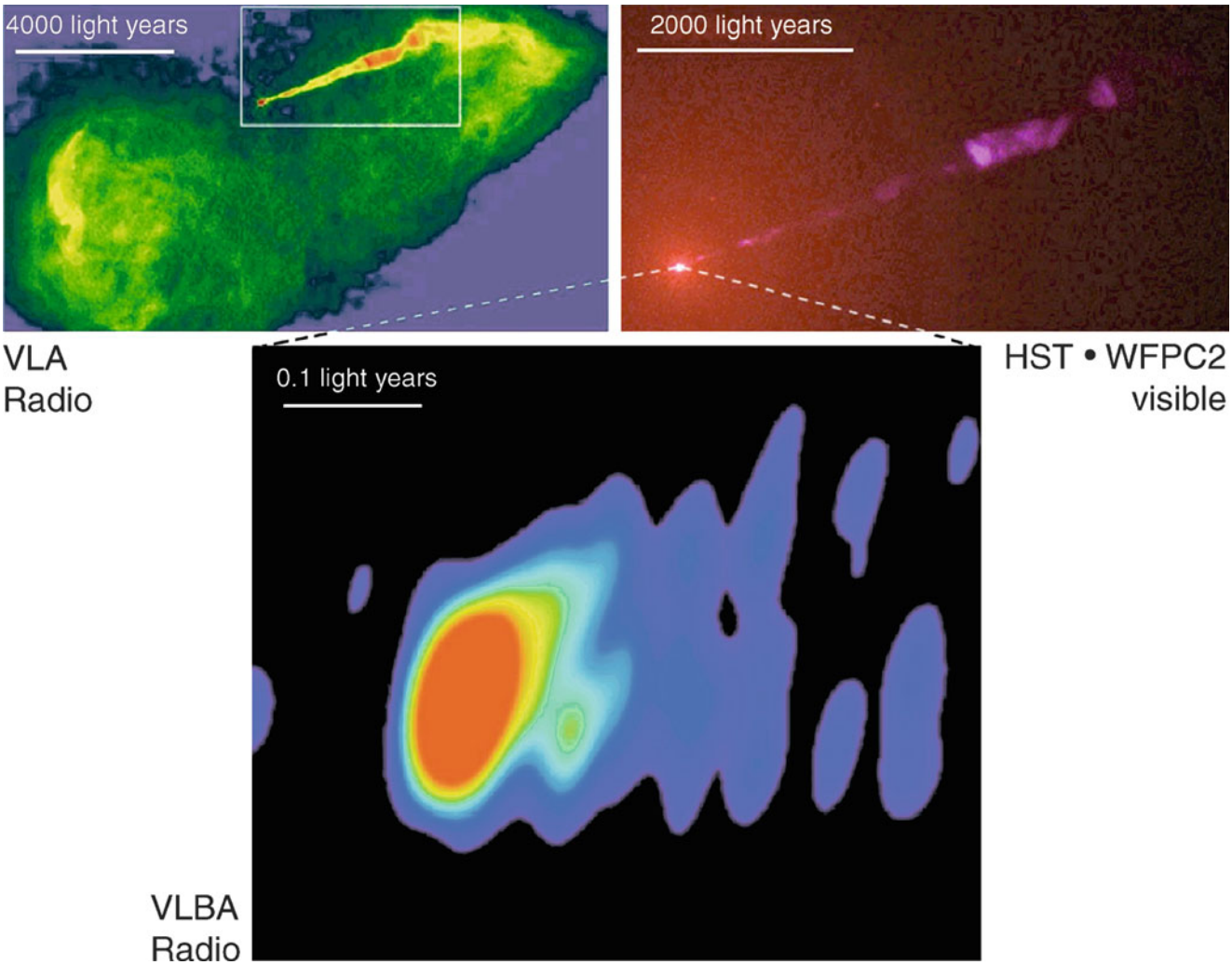

Fig. 5.47 *Top left*: A radio map of M87, the central galaxy in the Virgo cluster of galaxies. *Top right*: An HST image of the region shown in the *inset* of the *left-hand panel*. The radio jet is also visible at optical wavelengths. The *lower image* shows a VLBI map of the region around the galaxy core; the jet is formed within a few 10^{17} cm from the core of the galaxy, which contains a black hole of $M_\bullet \sim 6 \times 10^9 M_\odot$. Very close to the center the opening angle of the jet is significantly larger than further out. This indicates that the jet only becomes collimated at a larger distance. With VLBI observations at 1.3 mm, it was shown that the base of the jet is as compact as 40 μarcsec—corresponding to about $6r_S$! Credit: NASA, National Radio Astronomy Observatory/National Science Foundation, John Biretta (STScI/JHU), and Associated Universities, Inc.

emission by relativistic electrons moving in a magnetic field. This electron population can also produce X-ray photons. However, it is less clear which emission process generates the X-ray emission. The two principal possibilities are (i) synchrotron emission as well, or (ii) inverse Compton scattering. In the first case, the spectrum of the electrons must extend to extremely high energies: If the radio emission is due to synchrotron emission of electrons with $\gamma \sim 10^4$ in a magnetic field with $\sim 100\,\mu$G [cf. (5.3)], then in order to generate X-rays in the same magnetic field, the Lorentz factor must be $\sim 10^8$—corresponding to an electron energy of $E = \gamma m_e c^2 \sim 100$ TeV. It is currently unclear which acceleration processes may account for these high energies.

In the case of inverse Compton scattering, low-energy photons are scattered to much higher energies by collisions with relativistic electrons—a photon of frequency ν typically has a frequency $\nu' \approx \gamma^2 \nu$ after being scattered by an electron of energy $\gamma m_e c^2$ (see also Sect. 5.4.4 for the case of inverse Compton scattering by a thermal distribution of electrons). Since the characteristic Lorentz factors of electrons causing the synchrotron radiation of radio jets may reach values of $\gamma \sim 10^4$, these electrons may scatter, by inverse Compton scattering, radio photons into the X-ray domain of the spectrum. If the radio photons are those produced in situ by synchrotron emission, this effect is also called synchrotron self-Compton radiation. In particular, if the low-frequency spectrum is a power law, then the inverse

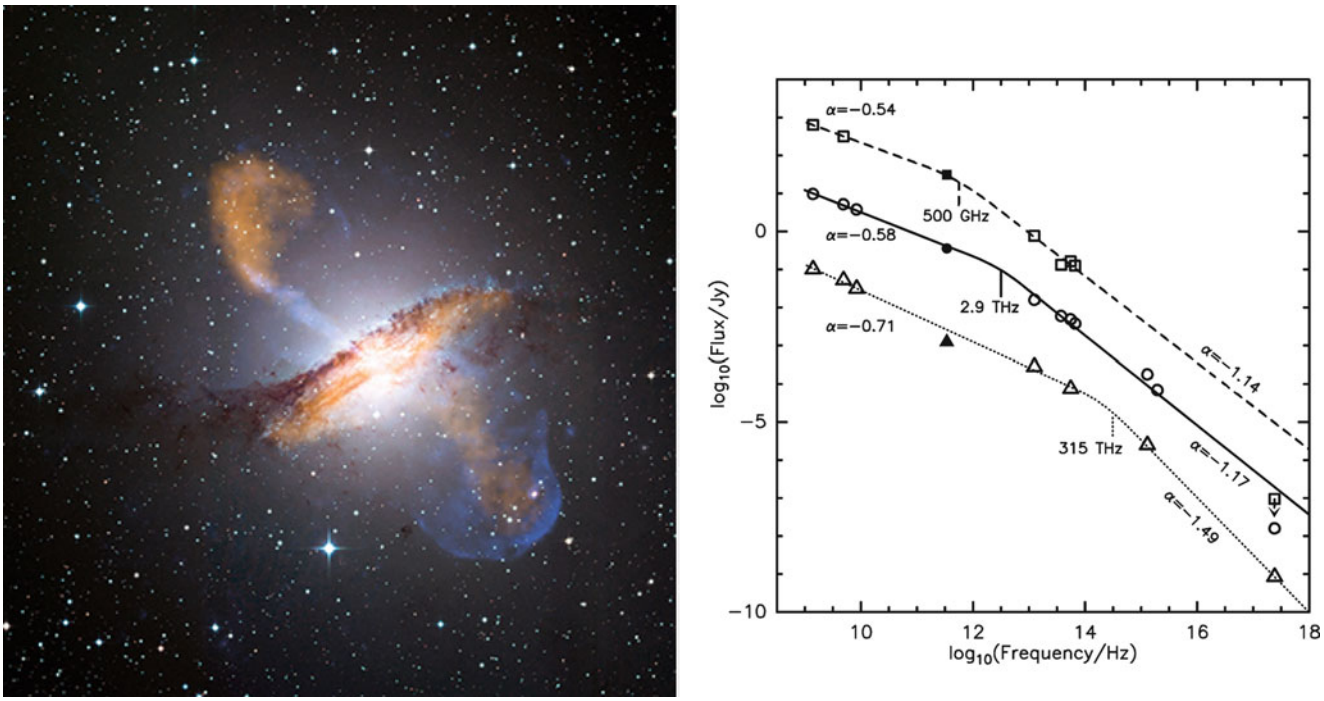

Fig. 5.48 *Left panel*: A composite image of the radio galaxy Centaurus A (NGC 5128), one of the first active galaxies discovered. This image is 16′ across, and combines optical data from the Wide Field Imager at the MPG/ESO 2.2 m telescope at La Silla (shown in *white*), the X-ray data from the Chandra observatory (*blue*), and the sub-millimeter data from the APEX telescope in Chile (*orange*). The dust lane, clearly seen in the optical image, emits strongly in the sub-millimeter regime, due to dust heated by starlight. Furthermore, APEX and Chandra display the AGN activity in this source, due to jets and lobes, which are located in a direction roughly perpendicular to the dust lane. Thus, the origin of the sub-millimeter emission from the dust lane and the jet/lobe regions is quite different: in the former case it is due to warm dust with a temperature of $T_{\rm d} \sim 30$ K, as obtained by combining the APEX data with far-IR data from the ISO satellite, whereas in the latter case it is due to synchrotron emission. *Right panel*: By combining the data shown in the image with radio and optical data of the jet region, one finds that the spectrum can be well described by a broken power law, from the radio to the X-rays, which is the expected spectral behavior of a synchrotron source where the population of relativistic electrons cools due to emission. The three sets of data, and the corresponding lines, are the spectra at different locations along the jet/lobe, with the lowest one being closest to the galaxy nucleus, and the *dashed* one being furthest away. As can be seen, the break frequency, i.e., the frequency where the spectrum steepens, decreases as one gets further out into the lobe, again as expected from a cooling electron distribution. Credit: *Left*: X-ray: NASA/CXC/CfA/R. Kraft et al.; Sub-millimeter: MPIfR/ESO/APEX/A. Weiss et al.; optical: ESO/WFI. *Right*: A. Weiss et al. 2008, *LABOCA observations of nearby, active galaxies*, A&A 490, 77, p. 85, Fig. 11. ©ESO. Reproduced with permission

Compton spectrum will be a power law as well. Alternatively, relativistic electrons can also scatter optical photons from the AGN, for which less energetic electrons are required. The omnipresent CMB may also be considered as a photon source for the inverse Compton effect, and in many cases the observed X-ray radiation is probably Compton-scattered CMB radiation.

It is usually very difficult to distinguish between these alternatives, and it is by no mean clear that the high-energy emission has the same origin in all sources. For at least one source, there is good evidence that the synchrotron self-Compton effect is at play. For that, we start with a simple thought experiment: suppose that in a synchrotron source the number density of relativistic electron would suddenly double. In this case, the synchrotron emission would simply double as well, since it is linear in the electron number. However, the synchrotron self-Compton emission would increase by a factor of four, since both the number of low-energy (synchrotron) photons, as well as the number of electrons doubles, and hence there are four times as many scatterings between low-energy photons and electrons. Hence, for synchrotron self-Compton, one expects that there is a quadratic dependence of the Comptonized flux on the synchrotron flux.

This argument, however, assumes that the source is optically thin with respect to Compton scattering. In the most compact sources, this may not be the case, and the Comptonized photons may be upscattered again. This higher-order scattering emission would then, to first order, have a cubic dependence on the electron density—and hence a cubic dependence on the synchrotron emissivity.

The BL Lac object PKS 2155−304 was observed simultaneously in the optical, X-ray and TeV gamma regime when it had a major flare; this flare, which happened on a time-scale of ~ 1 h, led to flux variations by more than a factor of 20 in the TeV region, whereas the variations were merely a factor

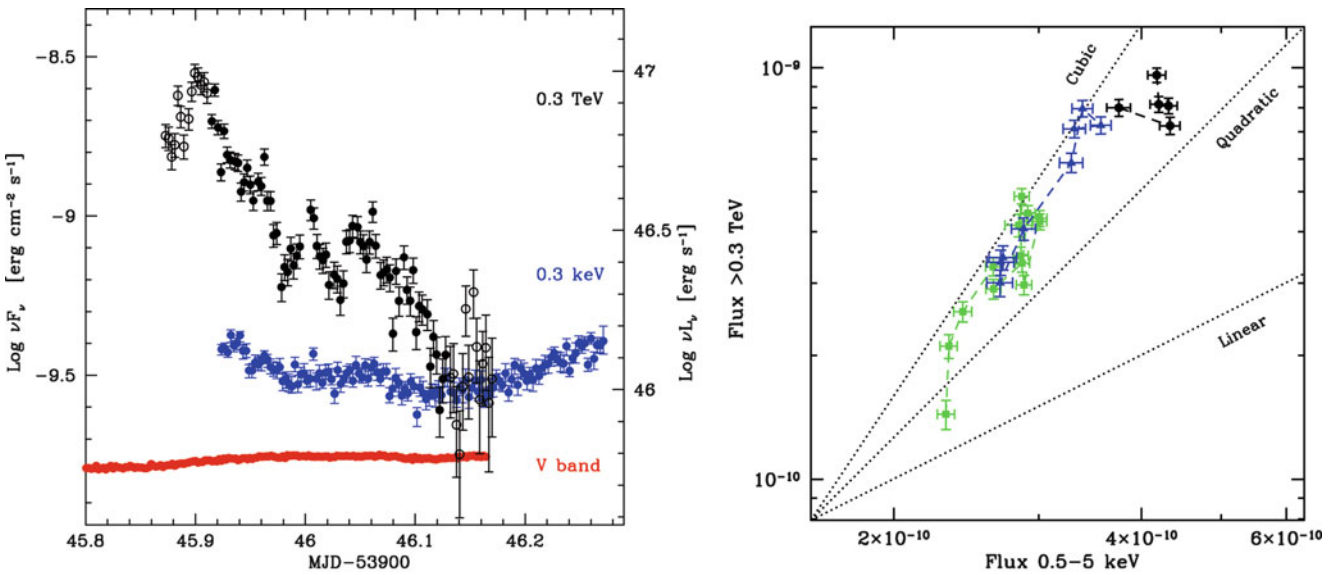

Fig. 5.49 A spectacular flare of the BL Lac object PKS 2155−304. In the *left panel*, the light curves of the source in the optical (*red*), X-ray (*blue*) and high-energy gamma radiation (*black*) at $E = 0.3$ TeV are shown. Note the very much enhanced range of fluxes in the gamma rays. In the *right panel*, the TeV flux is plotted against the X-ray flux, for several time intervals during the flare. As can be seen, the TeV flux varies approximately as the cube of the X-ray flux. Source: F. Aharonian et al. 2009, *Simultaneous multiwavelength observations of the second exceptional γ-ray flare of PKS 2155−304 in July 2006*, A&A 502, 749, p. 754, 762, Figs. 3, 13. ©ESO. Reproduced with permission

of 2 in the X-ray flux (and 15 % in the optical)—see the left panel of Fig. 5.49. The TeV light curve showed similar features as the X-ray (at 0.3 keV) light curve, however with quite a different scaling: correlating the TeV flux with that of the X-ray flux, a behavior $S_{\mathrm{TeV}} \propto S_{\mathrm{X}}^3$ was found (right panel of Fig. 5.49)—i.e., a behavior that is predicted for the case that the TeV radiation is second-order inverse Compton scattering and the X-ray emission is due to synchrotron radiation. Although the actual processes in the source are most likely much more complicated than outlined in this simple thought experiment, it nevertheless indicates that the synchrotron self-Compton process is acting in this source.

The inverse Compton model cannot, however, be applied to all X-ray jets without serious problems occurring. For instance, variability in X-ray emission was observed in the knots of M87, indicating a very short cooling time for the electrons. Since the electrons must have a much larger Lorentz factor γ if the radiation, at a given frequency, originates from synchrotron emission, compared to the case that the X-rays are produced by inverse Compton scattering, their cooling time t_{cool} (5.6) would be much shorter as well. In such sources, which are typically FR I radio sources, the variability therefore argues for the synchrotron process to be responsible for the X-ray emission. The implied very short cooling time-scales then leads to an increased necessity for a local acceleration of the electrons. On the other hand, the required energies for the electrons are very high, ~ 100 TeV.

Detecting radio jets at X-ray frequencies seems to be a frequent phenomenon: about half of the flat-spectrum radio QSO with jet-like extended radio emission also show an X-ray jet. All of those are one-sided, although the corresponding radio images often show lobes also on the other side of the X-ray jets, reinforcing the necessity for Doppler favoritism also in the X-ray waveband (Fig. 5.50).

TeV emission from blazars. Blazars emit at the highest photon energies yet observed, in the TeV $= 10^{12}$ eV spectral range, as can be observed with ground-based Cherenkov telescopes (see Sect. 1.3.6). Furthermore, they are the dominant population of extragalactic sources observed in the GeV-range. At TeV energies, they can exhibit strong flux variations on time-scales of minutes (see Fig. 5.49 for an example). The broad-band spectral energy distribution of these TeV blazars show two dominant broad peaks, one located in the X-ray regime of the spectrum, the other at TeV energies, with approximately equal energy output in both. As mentioned before, the correlated variability in both spectral ranges clearly argues for a synchrotron self-Compton origin of the high-energy gamma radiation.

The fact that we can see this emission is another clear argument for the highly relativistic nature of the source. If we ignore relativistic effects for a moment, the small time-scale of variability yields an upper bound on the source size. Together with the observed flux, and thus an estimated luminosity, one can calculate the energy density inside the source. This density is so high that no TeV photon could escape the emitting region—they would scatter with lower energy photons to produce electron-positron pairs. In order to avoid

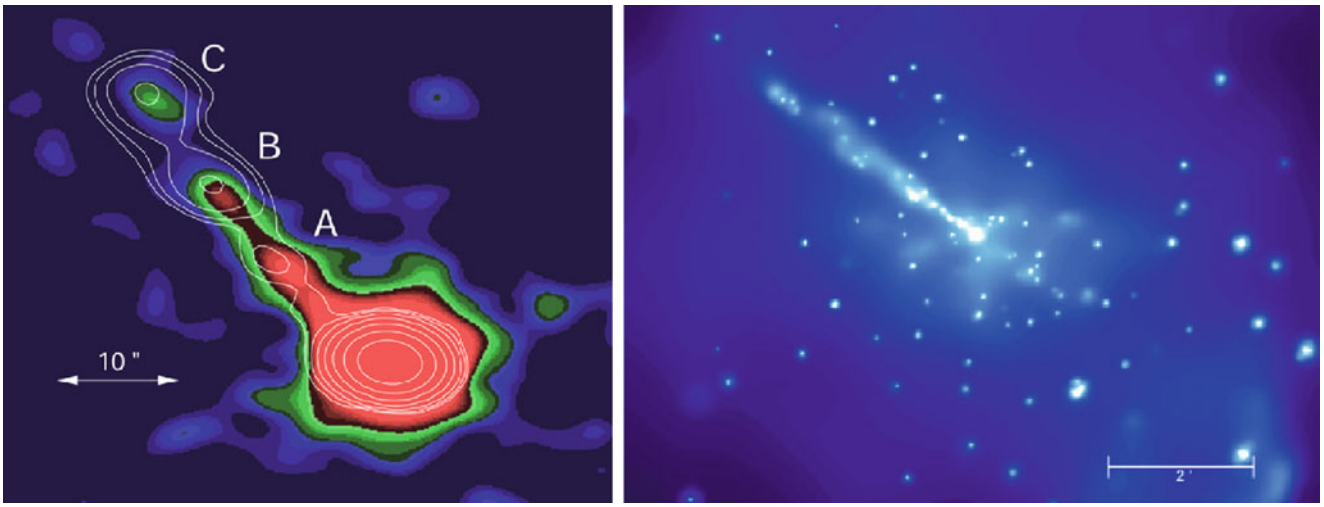

Fig. 5.50 X-ray images of AGN jets. *On the left*, a Chandra image of the jet in the QSO PKS 1127−145, with overlaid contours of radio emission (1.4 cm, VLA). The direction of the jet and its substructure are very similar at both wavelengths, suggesting an interpretation in which the radiation is caused by the same population of relativistic electrons. *On the right*, a Chandra image of the active galaxy Centaurus A. Here the jet is visible, as well as a large number of compact sources interpreted to be X-ray binaries. Credit: *Left*: X-ray: NASA/CXC/A. Siemiginowska/CfA & J. Bechtold/University of Arizona; radio: Siemiginowska et al./VLA. *Right*: NASA/SAO/R. Kraft et al.

this conclusion, the intrinsic luminosity must be smaller and the source size larger. Both is the case if we allow for a relativistic source component for which significant beaming occurs, with a Doppler factor $\mathcal{D} = [\gamma(1 - \beta \cos\phi)]^{-1}$ of order 10 or larger. This argument also implies that we are located in a direction very close to the jet axis.

5.5.5 Unified models—summary

After discussing the various components of the unified model, we summarize here our current picture of the AGN phenomenon, referring back to Fig. 5.12.

In common of all AGNs is a central SMBH, powered by an accretion disk. Surrounding gas is photoionized and radiates emission lines, with the broad lines coming from a region with a size of $\sim 10^3$ Schwarzschild radii around the black hole (and that of the highest ionization species from an even smaller region), and the narrow lines from a much more extended region. The strong infrared emission in all AGNs argues for the presence of obscuring matter, which is concentrated towards the plane in which the accretion occurs—the torus. There is a large spread in the ratio of radio-to-optical luminosity of AGNs, which presumably is related to the different efficiency of jet formation.

Our classification of an AGN depends strongly on the viewing direction, or the inclination angle. Type 1 AGNs are those where our view to the central engine is not blocked by the absorbing material of the torus, and we can see the BLR. At higher inclination, the view towards the BLR is no longer possible, and the object appears as a Type 2 AGN. Depending on the luminosity of the AGN and its radio-to-optical ratio, the Type 1's are either QSOs, Seyfert 1 galaxies or broad line radio galaxies, whereas for Type 2's, they are Type 2 QSOs, Seyfert 2 galaxies or narrow line radio galaxies. Blazars are sources in which we look directly into the relativistic jet, to within a few degrees of the jet axis, where phenomena of Special Relativity (like beaming) can explain the unusual spectral properties and rapid variability of these sources.

Refinement. The fact that the torus is probably a collection of optically thick clouds modifies this simple picture in a slight way. The question of whether a sight-line to the central AGN is absorbed is now one of probability: With the torus being in the equatorial plane, it is much more likely that our view to the AGN is absorbed if we are in this plane, compared to being close to the symmetry axis. However, there is a finite probability that we see through the torus (and thus classify the source as Type 1) even if we are close to the equatorial plane, and conversely, there is a finite probability to have an absorbed sight-line to the AGN even if being close to the symmetry axis. These refinements of the unification scheme are needed to understand some objects with observed properties which can place them in either (or neither) category.

Quasar mode vs. radio mode. Geometry and orientation are not the only effects which determine the appearance of an AGN. For example, the spin of the black hole may play an important role in the behavior of an accretion flow. Furthermore, the efficiency of an accretion flow to launch a jet probably depends on the accretion rate. If this is

sufficiently low, the accretion disk will become optically thin, and can not longer radiate efficiently (see Sect. 5.3.2). In this case, most of the energy is advected inwards, and there are strong indications that this provides a very favorable situation for launching a jet. Indeed, there is a very strong correlation between the ratio of the emitted fluxes in the radio and optical spectral range, and the Eddington ratio $L/L_{\rm edd}$. Sources for which the latter is high, i.e., where the accretion occurs through an optically thick disk, emit only a small fraction of their luminosity at radio wavelengths, indicating rather inefficient jet production. This can also be seen in Fig. 5.13, in which the radio-to-optical luminosity increases from the bottom right to the upper left corner. The luminous QSOs (star symbols) are located towards the bottom right corner, whereas the FR I galaxies are near the upper left one.

It thus seems that, depending on the accretion rate, a black hole can either shine through the 'QSO mode', where the luminosity is dominated by quasi-thermal radiation from the accretion disk, or through a 'radio mode', in which the disk does not radiate efficiently, and a large fraction of the energy is channeled into an outflow, visible in form of a jet.

Our attempts at finding a unification scheme for the different classes of AGNs have been quite successful. The scheme of unification is generally accepted, even though some aspects are still subject to discussion and require further studies.

5.5.6 Tidal disruption events

Our AGN model connects the activity with the accretion of gas onto a SMBH in the center of a galaxy. Despite the fact that all galaxies (with a spheroidal stellar component) host a SMBH, only those where gas can flow inwards and accrete show AGN activity.

Disruption of a star. Situated in a galactic nucleus, from time to time a star on its orbit may come close to the SMBH—perhaps even too close for its survival. We have seen in Sect. 2.3.6 that tidal gravitational forces can disrupt a system of particles—or a star. The condition for this to happen is [cf. (2.47)]

$$\frac{M_*}{r_*^3} \lesssim \frac{M_\bullet}{R^3} \,, \tag{5.40}$$

where M_* and r_* denote the stellar mass and radius, and R is its distance to the SMBH. Once a star satisfies this condition, it will be disrupted in the tidal field of the SMBH.

Expectations. The consequences of such an event were studied theoretically already in the 1980s: Following this process, about half of the stellar mass will be forced into a bound orbit around the black hole, while the other half is ejected. The bound mass will then be accreted onto the black hole; the initial phase of this accretion process proceeds rapidly, whereas the accretion rate at later times is expected to decreases with time as $t^{-5/3}$, if t is measured from the time of disruption.

This process is expected to have clear observational signatures: The initial accretion event should show up as a bright flare, perhaps even close to the Eddington luminosity of the SMBH. At later times, the luminosity should decrease roughly in proportion to the accretion rate, i.e., $\propto t^{-5/3}$. This flare should occur in the center of a galaxy, even if that galaxy has shown no sign of nuclear activity before.

These events should be quite rare; from the density of stars in the center of galaxies and stellar dynamics, typical event rate are estimated to be of order $\sim 10^{-5}$ per galaxy per year. Hence, they may only be detected by monitoring a large number of galaxies.

Detection of stellar disruption events. Not surprisingly, the first such tidal disruption events (TDEs) were found in the X-rays—since in this spectral regime, radiation from the AGN sticks out most clearly from the emission of the host galaxy. Repeated observations by the ROSAT satellite led to the discovery of the first TDEs in the late 1990s. The soft X-ray luminosity at peak brightness is huge, reaching $10^{44}\,{\rm erg/s}$ or even higher. Long-term monitoring shows that the decline of the luminosity is consistent with the $t^{-5/3}$-law predicted from the accretion of tidal debris—for the first events, the decline could be followed over more than a decade, showing a decrease from the peak flux by more than three orders of magnitude. From modeling the events, typical SMBH masses of 10^6 to $10^8\,M_\odot$ are derived, as expected from the scaling relation between $M_\bullet$ and the properties of the stellar population of the host galaxy. Despite the small number statistics of events (of order a dozen have been observed by now), the event rate is compatible with the theoretical expectations.

One of the most recent events was discovered by the Swift satellite, both in soft and hard X-rays. The peak flux was at least 10^4 times larger than the X-ray flux 20 years before the TDE, and at least a factor of 100 larger than a year before the event, as obtained from upper flux limits derived from previous X-ray observations of the galaxy (at $z = 0.35$). In fact, the peak flux corresponds to a luminosity of $\sim 10^{48}\,{\rm erg/s}$—corresponding to the Eddington luminosity of a SMBH with $M_\bullet \sim 10^{10} M_\odot$. However, the actual black hole mass must be far smaller: strong variability of the X-ray flux was seen on a time-scale of 100 s, which implies that the light-travel time across the Schwarzschild radius cannot be larger than this value. This argument yields an upper bound of $M_\bullet \leq 7 \times 10^6 M_\odot$. Hence, during its peak the apparent

luminosity of the flare was much larger than the Eddington luminosity of the SMBH.

The solution of this apparent discrepancy is provided from radio observations of this event; from the temporal and spectral properties seen at radio frequencies, it is suspected that the TDE launched a relativistic jet. The associated beaming can explain the huge apparent X-ray luminosity. Curiously, integrating the X-ray luminosity over the first 50 days of the flare yields a total energy of $\sim 10^{53}$ erg—which is close to the energy generated by accreting $1 M_\odot$ with an efficiency of 10 %.

TDEs not only provide a clear confirmation that inactive SMBHs can be revived, but they will allow us to study details of the accretion physics, provided they are followed with sufficient time coverage and depth. The onset of the formation of a jet is just one example of what we can learn from these events. The future eROSITA survey will obtain several all-sky surveys and is ideally placed to discover many such events.

5.6 Properties of the AGN population

AGNs, and QSOs in particular, are visible out to very high redshifts. Since their discovery in 1963, QSOs have held the redshift record most of the time. Only in recent years have QSOs and galaxies been taking turns in holding the record. Today, several hundred QSOs are known with $z \geq 4$, and the number of those with $z > 5$ continues to grow since a criterion was found to identify these objects. This leads to the possibility that QSOs could be used as cosmological probes, and thus to the question of what we can learn about the Universe from QSOs. For example, one of the most exciting questions is how does the QSO population evolve with redshift—was the abundance of QSO at high redshifts, i.e., at early epochs of the cosmos, similar to that today, or does it evolve over time?

5.6.1 The K-correction

To answer this question, we must know the luminosity function of QSOs, along with its redshift dependence. As we did for galaxies, we define the luminosity function $\Phi(L, z)\,\mathrm{d}L$ as the spatial number density of QSOs with luminosity between L and $L + \mathrm{d}L$. Φ normally refers to a comoving volume element, so that a non-evolving QSO population would correspond to a z-independent Φ. One of the problems in determining Φ is related to the question of which kind of luminosity is meant here. For a given observed frequency band, the corresponding rest-frame radiation of the sources depends on their redshift. For optical observations, the measured flux of nearby QSOs corresponds to the rest-frame optical luminosity, whereas it corresponds to the UV luminosity for higher-redshift QSOs. In principle, using the bolometric luminosity would be a possible solution; however, this is not feasible since it is *very* difficult to measure the bolometric luminosity (if at all possible) due to the very broad spectral energy distribution of AGNs. Observations at all frequencies, from the radio to the gamma domain, would be required, and obviously, such observations can only be obtained for selected individual sources.

Of course, the same problem occurs for all sources at high redshift. In comparing the luminosity of galaxies at high redshift with that of nearby galaxies, for instance, it must always be taken into account that, at given observed wavelength, different spectral ranges in the galaxies' rest-frames are measured. This means that in order to investigate the optical emission of galaxies at $z \sim 1$, observations in the NIR region of the spectrum are necessary.

Frequently the only possibility is to use the luminosity in some spectral band and to compensate for the above effect as well as possible by performing observations in several bands. For instance, one picks as a reference the blue filter which has its maximum efficiency at ~ 4500 Å and measures the blue luminosity for nearby objects in this filter, whereas for objects at redshift $z \sim 0.7$ the intrinsic blue luminosity is obtained by observing with the I-band filter, and for even larger redshifts observations need to be extended into the near-IR. The observational problems with this strategy, and the corresponding corrections for the different sensitivity profiles of the filters, must not be underestimated and are always a source of systematic uncertainties. An alternative is to perform the observation in only one (or a few) filters and to approximately correct for the redshift effect.

In Sect. 4.3.3, we defined various distance measures in cosmology. In particular, the relation $S = L/(4\pi D_\mathrm{L}^2)$ between the observed flux S and the luminosity L of a source defines the luminosity distance D_L. Here both the flux and the luminosity refer to bolometric quantities, i.e., flux and luminosity integrated over all frequencies. Due to the redshift, the measured spectral flux S_ν is related to the spectral luminosity $L_{\nu'}$ at a frequency $\nu' = \nu(1 + z)$, where one finds

$$S_\nu = \frac{(1+z) L_{\nu'}}{4\pi D_\mathrm{L}^2} \,. \tag{5.41}$$

We write this relation in a slightly different form,

$$S_\nu = \frac{L_\nu}{4\pi D_\mathrm{L}^2} \left[\frac{L_{\nu'}}{L_\nu} (1+z) \right] \,, \tag{5.42}$$

where the first factor is of the same form as in the relation between the bolometric quantities while the second factor corrects for the frequency shift. This factor is denoted the

K-correction. It obviously depends on the spectrum of the source, i.e., to determine the K-correction for a source its spectrum needs to be known. Furthermore, this factor depends on the filter used. Since in optical astronomy magnitudes are used as a measure for brightness, (5.42) is usually written in the form

$$m_{\rm int} = m_{\rm obs} + K(z) \quad \text{with} \quad K(z) = -2.5 \log \left[\frac{L_{\nu'}}{L_\nu}(1+z) \right] , \tag{5.43}$$

where $m_{\rm int}$ is the magnitude that would be measured if the spectrum of the source would not be shifted in wavelength by redshift, and $m_{\rm obs}$ describes the brightness actually observed. The K-correction is not only relevant for QSOs but for all objects at high redshift, in particular also for galaxies.

5.6.2 The luminosity function of QSOs

Construction of the luminosity function. By counting QSOs, we obtain the number density $N(>S)$ of QSOs with a flux larger than S. We find a relation of roughly $N(>S) \propto S^{-2}$ for large optical fluxes S, whereas the source counts are considerably flatter for smaller fluxes. The flux at which the transition from steep counts to flatter ones occurs corresponds to an apparent magnitude of about $B \sim 19.5$. About 10 QSOs per square degree are found brighter than this break magnitude.

From QSO number counts, combined with measurements of QSO redshifts, the luminosity function $\Phi(L, z)$ can be determined. As already defined above, $\Phi(L, z)\,{\rm d}L$ is the number density *in a comoving volume element* of QSOs at redshift z with a luminosity between L and $L + {\rm d}L$.

Two fundamental problems exist in determining the luminosity function. The first is related to the above discussion of wavelength shift due to cosmological redshift: a fixed wavelength range in which the brightness is observed corresponds to different wavelength intervals in the intrinsic QSO spectra, depending on their redshift. We need to correct for this effect if the number density of QSOs above a given luminosity in a certain frequency interval is to be compared for local and distant QSOs. One way of achieving this is by assuming a universal spectral shape for QSOs; over a limited spectral range (e.g., in the optical and the UV ranges), this assumption is indeed quite well satisfied. This universal spectrum is obtained by averaging over the spectra of a larger number of QSOs (Fig. 5.2). By this means, a useful K-correction of QSOs as a function of redshift can then be derived.

The second difficulty in determining $\Phi(L, z)$ is to construct QSO samples that are 'complete'. Since QSOs are point-like they cannot be distinguished from stars by morphology on optical images, but rather only by their color properties and subsequent spectroscopy. However, with the star density being much higher than that of QSOs, this selection of QSO candidates by color criteria, and subsequent spectroscopic verification, is very time-consuming. Only more recent surveys, which image large areas of the sky in several filters, were sufficiently successful in their color selection and subsequent spectroscopic verification, so that very large QSO samples could be compiled. An enormous increase in statistically well-defined QSO samples was achieved by two large surveys with the 2dF spectrograph and the Sloan Digital Sky Survey which we discuss in the context of galaxy redshift surveys in Sect. 8.1.2.

The optical QSO luminosity function. The luminosity function that results from such analyses is typically parametrized as

$$\Phi(L, z) = \frac{\Phi^*}{L^*(z)} \left[\left(\frac{L}{L^*(z)} \right)^{\gamma_1} + \left(\frac{L}{L^*(z)} \right)^{\gamma_2} \right]^{-1} ; \tag{5.44}$$

i.e., for fixed z, Φ is a double power law in L. At $L \gg L^*(z)$, the second term in the square brackets in (5.44) dominates if $\gamma_2 > \gamma_1$, yielding $\Phi \propto L^{-\gamma_2}$. On the other hand, the first term dominates for $L \ll L^*(z)$, so that $\Phi \propto L^{-\gamma_1}$. Typical values for the exponents are $\gamma_1 \approx 1.5$, $\gamma_2 \approx 3.5$. The characteristic luminosity $L^*(z)$ where the L-dependence changes, strongly depends on redshift. A good fit to the data for $z \lesssim 2$ is achieved by

$$L^*(z) = L_0^*(1+z)^k , \tag{5.45}$$

with $k \approx 3.45$, where the value of k depends on the assumed density parameters $\Omega_{\rm m}$ and Ω_Λ. This approximation is valid for $z \lesssim 2$, whereas for larger redshifts $L^*(z)$ seems to vary less with z. The normalization constant is determined to be $\Phi^* \approx 5.2 \times 10^3\, h^3\, {\rm Gpc}^{-3}$, and L_0^* corresponds to roughly $M_{\rm B} = -20.9 + 5 \log h$. The luminosity function as determined from the combined 2dF and SDSS surveys is plotted in Fig. 5.51.

From this luminosity function, a number of conclusions can be drawn. The luminosity function of QSOs is considerably broader than that of galaxies, which we found to decrease exponentially for large L, compared to the power-law behavior we see here. The strong dependence of the characteristic luminosity $L^*(z)$ on redshift, which is seen in Fig. 5.51 as a systematic shift of the turnover luminosity towards fainter values as z decreases, clearly shows a very significant cosmological evolution of the QSO luminosity function. For example, at $z \sim 2$, $L^*(z)$ is about 50 times larger than today. Furthermore, for high luminosities, $\Phi \propto [L^*(z)]^{\gamma_2 - 1}\, L^{-\gamma_2}$. This means that the spatial number density of luminous QSOs was more than 1000 times larger at $z \sim 2$ than it is today, which can also be seen directly by

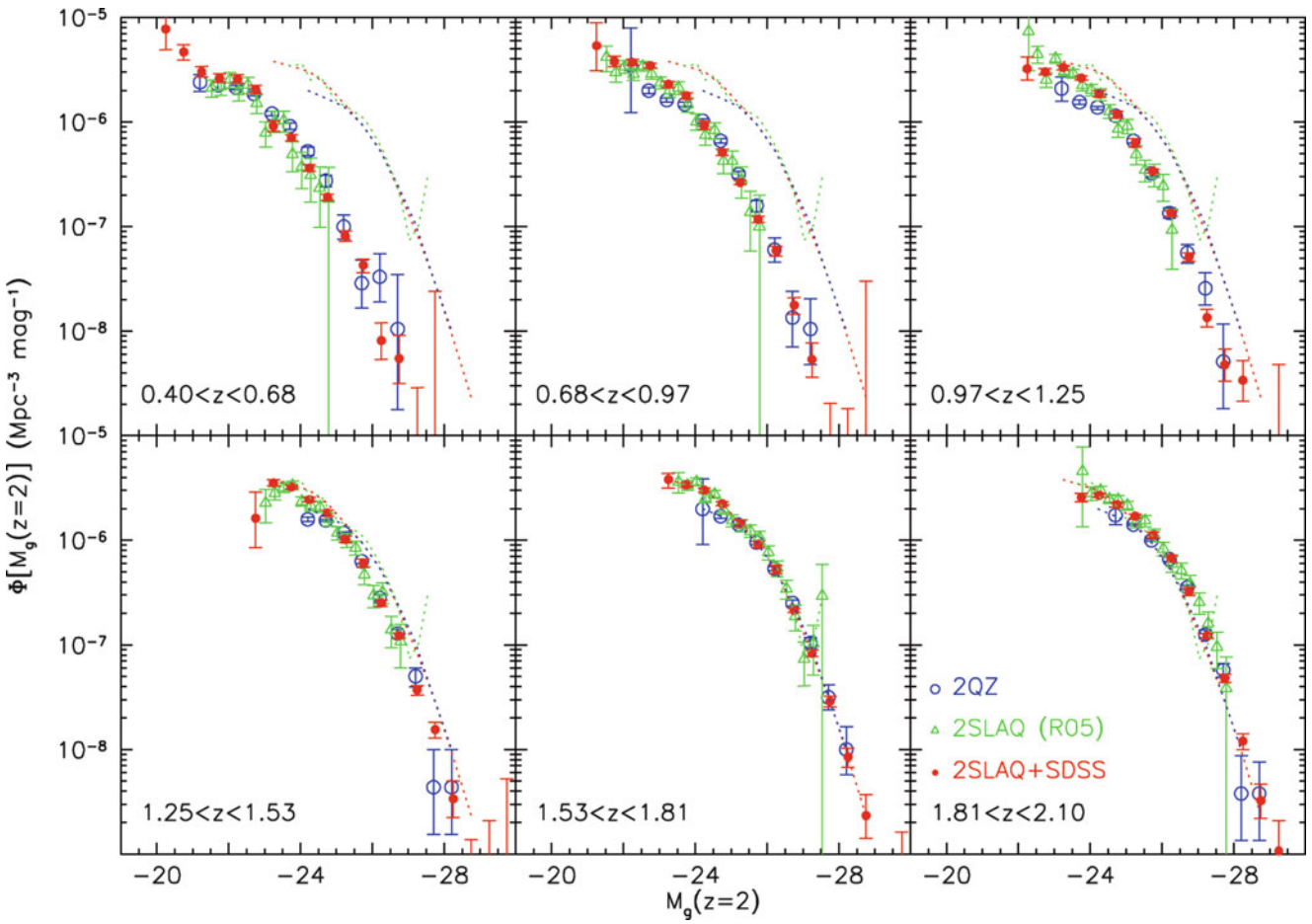

Fig. 5.51 The luminosity function from the combined 2dF and SDSS QSO surveys (*red points*), compared to earlier surveys, plotted for six different redshift intervals as indicated. The *dotted curves* in each panel show the fit to the data of the redshift bin $1.53 < z < 1.81$. For obtaining these results, several corrections were necessary, including K-correction and to account for the light from the host galaxy. Source: S.M. Croom et al. 2009, *The 2dF-SDSS LRG and QSO survey: the QSO luminosity function at* $0.4 < z < 2.6$, MNRAS 399, 1755, p. 1764, Fig. 11. Reproduced by permission of Oxford University Press on behalf of the Royal Astronomical Society

comparing the data points in the lowest-redshift panel with the dotted curve in Fig. 5.51—the low-redshift luminosity function does not extend to the very bright luminosities for which the luminosity function at high redshifts was measured. The number density of very luminous QSOs at low redshifts is so small that essentially none of them are contained in the survey volume from which the results in Fig. 5.51 were derived. This point is better illustrated in Fig. 5.52 which shows the evolution of the number density of QSOs as a function of redshift, for different bins in luminosity. High-L QSOs are very rare in the current Universe, and in particular, the ratio of the number density of high- to-low luminosity QSOs strongly decreases towards lower redshifts.

AGN selection in X-rays. Finding QSOs at higher redshifts with optical methods is more difficult, as we discussed before—the observed optical colors change as the strong emission lines move into the optical bands. Furthermore, optical surveys are not best suited for finding low-luminosity AGNs; for them, the optical light of the host galaxy renders the AGN less pronounced, and thus it is more difficult to identify it as such without spectroscopy.

Obtaining a complete census of the AGN population is much easier by X-ray selection, for a number of reasons. First, all known AGN-types emit X-rays; furthermore, the fraction of the total energy that is emitted in the form of X-rays is less dependent on the AGN type than is the case for the optical emission. Second, the X-ray emission from galaxies is weak, and thus the AGN sticks out clearly. In fact, in a high Galactic latitude field, $\sim$ 90 % of all X-ray sources are AGNs, the rest being galaxy clusters. The fact that far more optically-selected AGNs are known than X-ray-selected ones is due to the small sky areas over which deep X-ray surveys have been carried out.

Indeed, one can estimate that optical surveys miss $\sim$ 80 % of the AGNs at any fixed bolometric luminosity, whereas X-ray surveys are more efficient. Observing at hard X-rays (2–20 keV), one obtains essentially a complete census of AGNs for bolometric luminosities $L_{\rm bol} \gtrsim 10^{45}$ erg/s, and

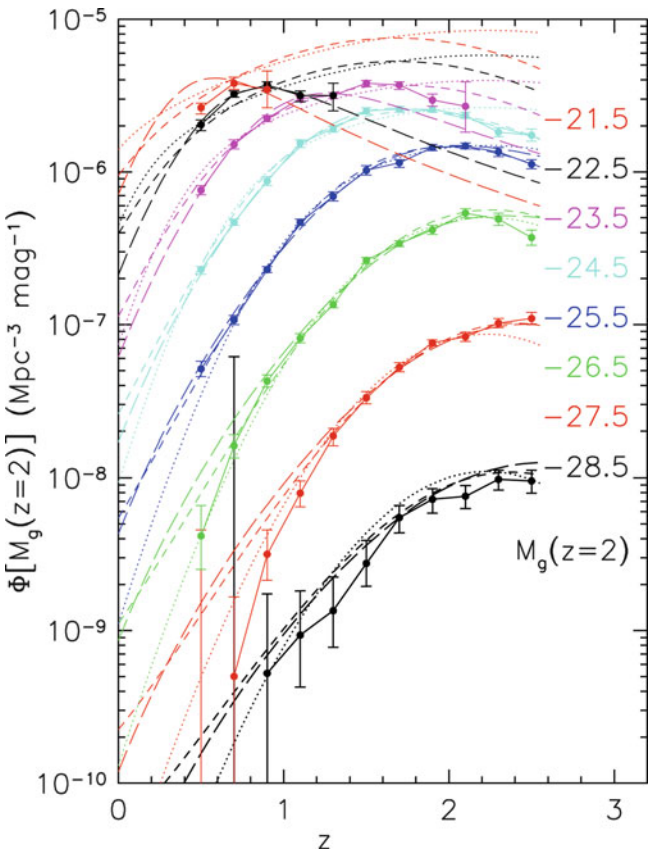

Fig. 5.52 The comoving spatial density of QSOs per magnitude interval, as obtained from the combined 2dF and SDSS QSO surveys, plotted for eight different bins in absolute magnitude. The *solid curves* connect the data points, whereas the other *curves* show various evolution models. As already seen from Fig. 5.51 the density of luminous QSOs steadily decreases from $z \sim 2$ until today. This decrease is also seen for somewhat less luminous QSOs, however it is less steep. In particular, for the less luminous ones, it appears as if the maximal space density is at redshifts smaller than 2. The redshift of the peak density of QSOs decreases towards lower luminosity sources. Source: S.M. Croom et al. 2009, *The 2dF-SDSS LRG and QSO survey: the QSO luminosity function at* $0.4 < z < 2.6$, MNRAS 399, 1755, p. 1764, Fig. 10. Reproduced by permission of Oxford University Press on behalf of the Royal Astronomical Society

a moderate incompleteness of $\sim$ 30 % for a factor of 100 less luminous sources. Despite the higher sensitivity of soft X-ray (0.5–2 keV) detectors, the incompleteness here is larger, of order 50 % (again depending on luminosity). The reason for this behavior is the photoelectric absorption in the source, perhaps from the gas in the torus, combined with the frequency dependence of this effect. Thus, the emission in the soft X-ray band is more sensitive to this absorption, whereas the hard X-rays are much less affected. Only if the column density of hydrogen approaches $N_{\rm H} \sim 10^{24}\,{\rm cm}^{-2}$, and the source starts to become optically thick with respect to Compton scattering ('Compton thick AGN'), are the sources difficult to detect in hard X-rays.

We thus conclude that the high AGN completeness of X-ray observations makes this the preferred band for selection. As a drawback, however, one requires in addition an optical identification and spectrum, to obtain the redshift of the source.

Bolometric luminosity function. Given the broad-band spectral energy distribution of AGNs, together with observational constraints on the distribution of absorber column densities, the luminosity in one spectral band can be used to estimate the bolometric luminosity of sources. These estimates can be cross-checked by requiring that the bolometric luminosity determined, say, from the X-ray flux agrees with that obtained from the optical data, and that the distribution of the bolometric luminosity functions obtained from both band mutually agree. In this way, it is possible to construct the bolometric luminosity function of AGNs.

In Fig. 5.53 the parameters of the double power-law fit (5.44) to the luminosity function are displayed, now for the bolometric luminosity function. The bright- and faint-end slopes γ_i were allowed to depend on redshift as well. As the upper left and middle panel of Fig. 5.53 show, the data are not compatible with constant slopes; in particular, towards very high redshifts, the bright end of the luminosity function flattens. The overall normalization Φ^* is essentially constant.

There is a dramatic evolution in the break luminosity L^* with redshift, as seen in the lower left panel, increasing by a factor $\sim$ 30 from today to redshift $z \sim 2$, and then decreasing towards even higher redshifts. This drastic change is accompanied by a strong evolution of the bolometric luminosity density of AGNs, obtained by integrating the luminosity function (5.44) over L, as shown in the lower middle panel: The energy output from AGNs at redshift $z \sim 2$ was an order of magnitude larger than it is today, and decreases towards very high redshift. The drastic decrease in the space density of luminous optical QSOs, seen in Fig. 5.52, is matched by the model: the abundance of the most luminous AGNs shows a very strong drop from its maximum at $z \sim 2.5$ towards lower and higher redshifts.

Hence we conclude that the AGN population displays a remarkable cosmic evolution. AGN activity was peaked at redshifts $z \sim 2.5$, and the decrease towards earlier and later times is very pronounced. Given that the width of the peak, $\delta z \sim 1$, corresponds to a cosmologically relatively short time-scale, a kind of 'quasar epoch' happened in our Universe, in the sense that the (luminous) QSO population seems to have quickly formed and then largely became extinct again. For less luminous AGNs, the peak density occurs at lower redshifts, as can be seen also from Fig. 5.52. It thus appears that the most luminous AGNs formed first, and the less luminous ones at later epochs.

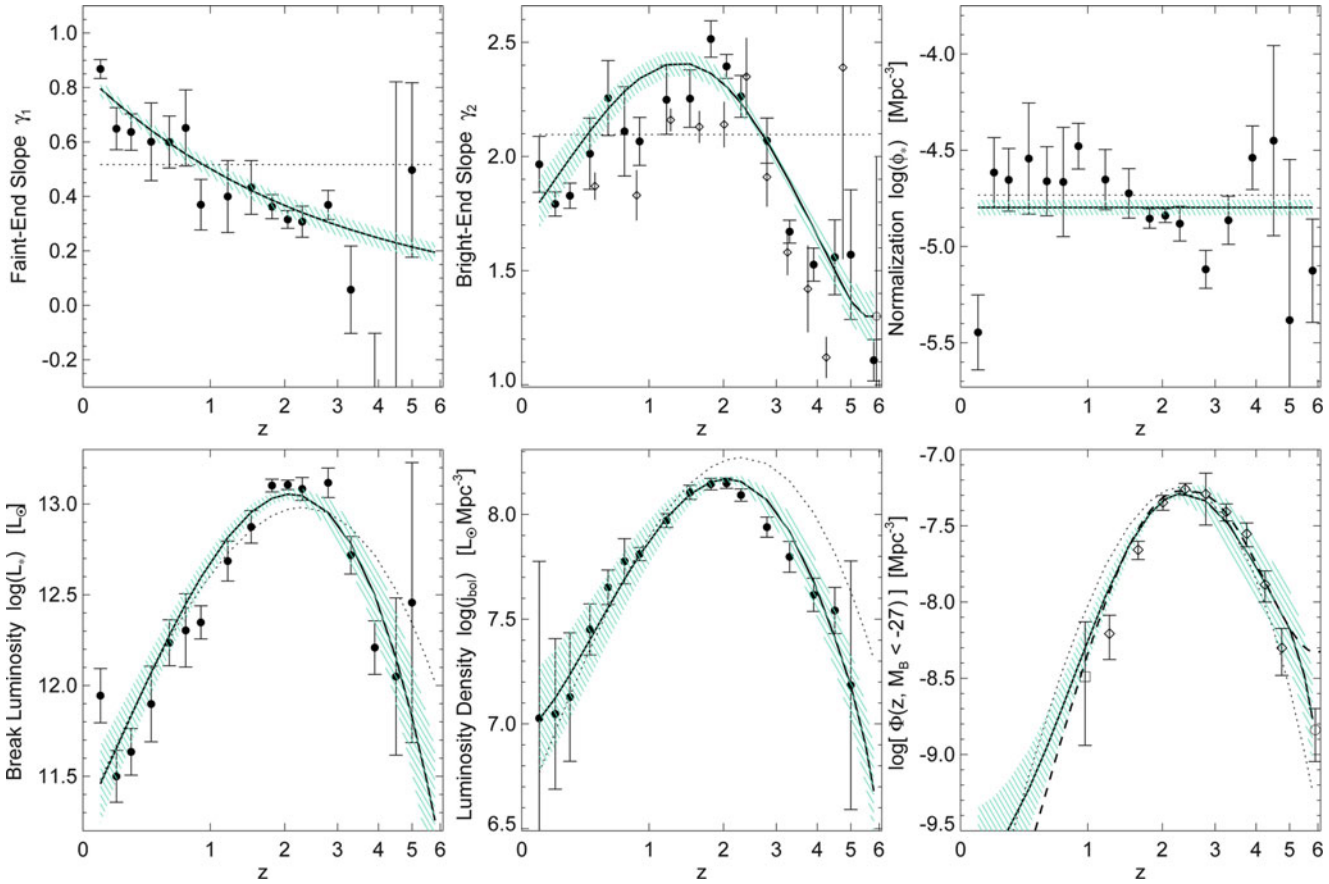

Fig. 5.53 Redshift-dependent parameters of the double power-law fit (5.44) to the bolometric luminosity function of AGNs. The *top row* shows the faint- and bright-end slopes γ_1, γ_2, and the normalization Φ^*. The *bottom row* shows the break luminosity L^*, the luminosity density and the abundance of optically luminous AGNs. The *solid curves* show the parameter fit to the data, the *dotted curves* correspond to a simple model in which the abundance is kept fixed, but the luminosity of each source evolves in redshift. This simple model (pure luminosity evolution) does not yield an acceptable fit. Source: P.F. Hopkins et al. 2007, *An Observational Determination of the Bolometric Quasar Luminosity Function*, ApJ 654, 731, p. 742, Fig. 8. ©AAS. Reproduced with permission

Interpretation. There are several possible interpretations of the QSO luminosity function and its redshift dependence. One of them is that the luminosity of any one QSO varies in time, parallel to the evolution of $L^*(z)$—this would correspond to a pure luminosity evolution model, indicated by the dotted curves in Fig. 5.53, which is seen to provide a rather poor fit to the data. Most likely this interpretation is wrong, also because it implies that a luminous QSO will always remain luminous. Although the efficiency of energy conversion into radiation is much higher for accretion than for thermonuclear burning, an extremely high mass would nevertheless accumulate in this case. This would then be present as the mass of the SMBH in local QSOs.[15] However, estimates of $M_\bullet$ in QSOs rarely yield values larger than $\sim 3 \times 10^9 M_\odot$.

However, it is by no means clear that a given source will be a QSO throughout its lifetime: a source may be active as a QSO for a limited time, and later appear as a normal galaxy again. It is likely that virtually any massive galaxy hosts a potential AGN. This is clearly supported by the fact that apparently all massive galaxies harbor a central SMBH. If the SMBH is fed by accreting matter, this galaxy will then host an AGN. However, if no more mass is provided, the nucleus will cease to radiate and the galaxy will no longer be active. Our Milky Way may serve as an example of this effect, since although the mass of the SMBH in the center of the Galaxy would be sufficient to power an AGN luminosity of more than 10^{44} erg/s considering its Eddington luminosity (5.25), the observed luminosity is lower by many orders of magnitude.

AGNs are often found in the vicinity of other galaxies. One possible interpretation is that the neighboring galaxy disturbs the gravitational field of the QSO's host, such that the flow of matter towards its central regions is favored where it is accreted onto the central black hole—and 'the

[15] Compare the mass estimate in Sect. 5.3.1 where, instead of 10^7 yr, the lifetime to be inserted here is the age of the Universe, $\sim 10^{10}$ yr.

monster starts to shine'. If this is the case, the luminosity function (5.44) does not provide information about individual AGNs, but only about the population as a whole.

Interpreting the redshift evolution then becomes obvious. The increase in QSO density with redshift in the scenario described above originates from the fact that at earlier times in the Universe, interactions between galaxies and merger processes were significantly more frequent than today. On the other hand, the decrease at very high z is to be expected because the SMBHs in the center of galaxies first need to form, and this obviously happens in the first $\sim 10^9$ yr after the Big Bang. We will see later (Sect. 9.6.2) that the star-formation history of the Universe displays a similar behavior as that of AGNs. In Chap. 10, we will consider models how this behavior can be understood in terms of the evolution of galaxies and their central SMBHs.

Black hole demography. Supermassive black holes grow in mass by accretion. Whereas the population of supermassive black holes can also be changed by merging processes, i.e., in the aftermath of galaxy mergers, the corresponding central black holes will merge as well, the total black hole mass is largely conserved in this case, modulo some general relativistic effects. The accretion is related to the energy release in AGNs; thus one might ask whether the total mass density of black holes at the present epoch is compatible with the integrated AGN luminosity. In other words, can the mass density of black holes be accounted for by the total accretion luminosity over cosmic time, as seen in the AGN population?

The first of these numbers is obtained from the scaling relation between SMBH mass and the properties of the spheroidal components in galaxies, as discussed in Sect. 3.8.3. This yields a value of the spatial mass density of SMBHs in the mass range $10^6 \le M_\bullet/M_\odot \le 5 \times 10^9$ of $\sim 4 \times 10^5 M_\odot/\mathrm{Mpc}^3$, with about a 30 % uncertainty. About a quarter of this mass is contributed by SMBHs in the bulges of late-type galaxies; hence, the total SMBH mass density is dominated by ellipticals.

The overall accreted mass is obtained from the redshift-dependent luminosity function of AGNs, by assuming an efficiency ϵ of the conversion of mass into energy. Indeed, the local mass density of SMBHs is matched if the accretion efficiency is $\epsilon \sim 0.10$, as is expected from standard accretion disk models. It therefore seems that the population of SMBHs located in normal galaxies at the present epoch have undergone an active phase in their past, causing their mass growth.

This argument may be slightly incomplete, in that some fraction of the energy released during the accretion process is converted into kinetic energy, as seen by powerful jets in AGN. This fraction is largely undetermined at present, but may not be negligible. In this case, the true ϵ needs to be somewhat higher than 0.1, which is only possible for black holes which rotate rapidly. In fact, the observed profile of the iron emission line from AGNs indicates black hole rotation. On the other hand, the mass growth of black holes is dominated by AGNs with a high Eddington ratio L/L_{edd}, and we have argued that most of them emit a relatively small fraction of their luminosity in the radio regime—and thus their jet power is relatively low.

A more detailed comparison between the SMBH and AGN populations reveals that the characteristic Eddington ratio is $\lambda_{\mathrm{Edd}} \sim 0.3$. With this value, combined with (5.30), one can estimate the mean time-scale over which a typical SMBH was active in the past, yielding $t_{\mathrm{act}} \sim 2 \times 10^8$ yr. Hence, the SMBH of a current day massive galaxy was active during about 2 % of its lifetime.

5.7 Quasar absorption lines

The optical/UV spectra of QSOs are characterized by strong emission lines. In addition, they also show absorption lines, which we have not mentioned thus far. Depending on the redshift of the QSOs, the wavelength range of the spectrum, and the spectral resolution, QSO spectra may contain a large variety of absorption lines. In principle, several different explanations for their occurrence exist. They may be caused by absorbing material in the AGN itself or in its host galaxy, so they have an intrinsic origin. Alternatively, they may arise during the long journey between the QSO and us due to intervening gas along the line-of-sight. We will see that different kinds of absorption lines exist, and that both of these possibilities indeed occur. The analysis of those absorption lines which do not have their origin in the QSO itself provides information about the gas in the Universe. For this purpose, a QSO is basically a very distant bright light source used for probing the intervening gas.

This gas can either be in intergalactic space or being correlated with foreground galaxies. In the first case, we expect that this gas is metal-poor and thus consists mainly of hydrogen and helium. Furthermore, in order to cause absorption, the intergalactic medium must not be fully ionized, but needs to contain a fraction of neutral hydrogen. Gas located closer to galaxies may be expected to also contain appreciable amounts of metals which can give rise to absorption lines.

The identification of a spectral line with a specific line transition and a corresponding redshift is, in general, possible only if at least two lines occur at the same redshift. For this reason, doublet transitions are particularly valuable, such as those of MgII ($\lambda = 2795$ Å and $\lambda = 2802$ Å), and CIV ($\lambda = 1548$ Å and $\lambda = 1551$ Å). The spectrum of virtually any QSO at high redshift z_{em} shows narrow absorption lines of CIV and MgII at redshifts $z_{\mathrm{abs}} < z_{\mathrm{em}}$. If the spectral coverage extends to shorter wavelengths than the observed Lyα emission line of the QSO, numerous narrow absorption lines show up at $\lambda_{\mathrm{obs}} \lesssim \lambda_{\mathrm{obs}}(\mathrm{Ly}\alpha) = (1 + z_{\mathrm{em}})1216$ Å.

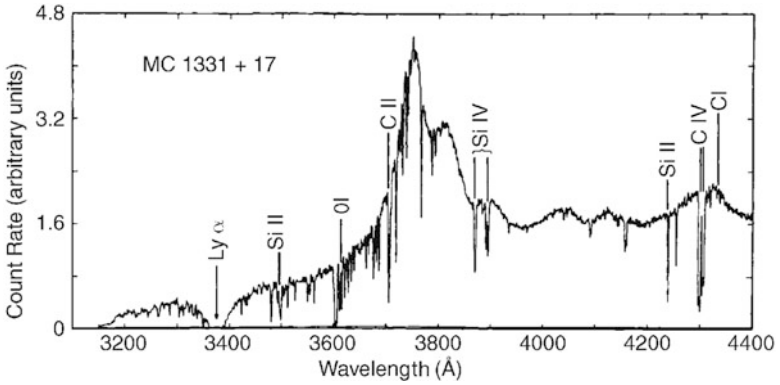

Fig. 5.54 Spectrum of the QSO 1331 + 17 at $z_{em} = 2.081$ observed by the Multi-Mirror Telescope in Arizona. In the spectrum, a whole series of absorption lines can be seen which have all been identified with gas at $z_{abs} = 1.776$. The corresponding Lyα line at $\lambda \approx 3400$ Å is very broad; it belongs to the damped Lyα lines. Source: F.H. Chaffee et al. 1988, *Molecules at early epochs. III - The Lyman-alpha disk system toward 1331 + 170*, ApJ 335, 584, p. 586, Fig. 1. ©AAS. Reproduced with permission

This set of absorption lines is denoted as the *Lyman-α forest*. In about 15 % of all QSOs, very broad absorption lines are found, the width of which may even considerably exceed that of the broad emission lines.

Classification of QSO absorption lines. The different absorption lines in QSOs are distinguished by classes according to their wavelength and width.

- *Metal systems:* In general these are narrow absorption lines, of which MgII and CIV most frequently occur (and which are the easiest to identify). However, in addition, a number of lines of other elements exist (Fig. 5.54). The redshift of these absorption lines is $0 < z_{abs} < z_{em}$; therefore they are caused by intervening matter along the line-of-sight and are not associated with the QSO. Normally a metal system consists of many different lines of different ions, all at the same redshift. From the line strength, the column density of the absorbing ions can be derived. For an assumed chemical composition and degree of ionization of the gas, the corresponding column density of hydrogen can then be determined. Estimates for such metal systems yield typical values of $10^{17}\,\mathrm{cm}^{-2} \lesssim N_H \lesssim 10^{21}\,\mathrm{cm}^{-2}$, where the lower limit depends on the sensitivity of the spectral observation.
- *Associated metal systems:* These systems have characteristics very similar to those of the aforementioned intervening metal systems, but their redshift is $z_{abs} \sim z_{em}$. Since such systems are over-abundant compared to a statistical z-distribution of the metal systems, these systems are interpreted as being related to the QSO itself. Thus the absorber is physically associated with the QSO and may be due, for example, to absorption in the QSO host galaxy or in a companion galaxy.
- *Lyα forest:* The large set of lines at $\lambda < (1 + z_{em})\,1216$ Å, as shown in Fig. 5.55, is interpreted to be Lyα absorption by hydrogen along the line-of-sight to the QSO. The statistical properties of these lines are essentially the same for all QSOs and seem to depend only on the redshift of the Lyα lines, but not on z_{em}. This interpretation is confirmed by the fact that for nearly any line in the Lyα forest, the corresponding Lyβ line is found if the quality and the wavelength range of the observed spectra permit this. The Lyα forest is further subdivided, according to the strength of the absorption, into narrow lines, Lyman-limit systems, and damped Lyα systems. Narrow Lyα lines are caused by absorbing gas of neutral hydrogen column densities of $N_H \lesssim 10^{17}\,\mathrm{cm}^{-2}$. Lyman-limit systems derive their name from the fact that at column densities of $N_H \gtrsim 10^{17}\,\mathrm{cm}^{-2}$, neutral hydrogen almost totally absorbs all radiation with $\lambda \lesssim 912$ Å (in the rest-frame of the gas), i.e., those photons which can ionize hydrogen (Fig. 5.56). If such a system is located at z_{limit} in the spectrum of a QSO, the observed spectrum at $\lambda < (1 + z_{limit})\,912$ Å is almost completely suppressed. Damped Lyα systems occur if the column density of neutral hydrogen is $N_H \gtrsim 2 \times 10^{20}\,\mathrm{cm}^{-2}$. In this case, the absorption line becomes very broad due to the extended damping wings of the Voigt profile.[16]

[16]The Voigt profile $\phi(\nu)$ of a line, which specifies the spectral energy distribution of the photons around the central frequency ν_0 of the line, is the convolution of the intrinsic line profile, described by a Lorentz profile,

$$\phi_L(\nu) = \frac{\Gamma/4\pi^2}{(\nu-\nu_0)^2 + (\Gamma/4\pi)^2}\,,$$

and the Maxwellian velocity distribution of atoms in a thermal gas of temperature T. From this, the Voigt profile follows,

$$\phi(\nu) = \frac{\Gamma}{4\pi^2}\int_{-\infty}^{\infty} \mathrm{d}v\, \frac{\sqrt{m/2\pi k_B T}\,\exp\left(-mv^2/2k_B T\right)}{(\nu-\nu_0-\nu_0 v/c)^2 + (\Gamma/4\pi)^2}\,, \qquad (5.46)$$

where the integral extends over the velocity component along the line-of-sight. In these equations, Γ is the intrinsic line width which results from the natural line width (related to the lifetime of the atomic states)

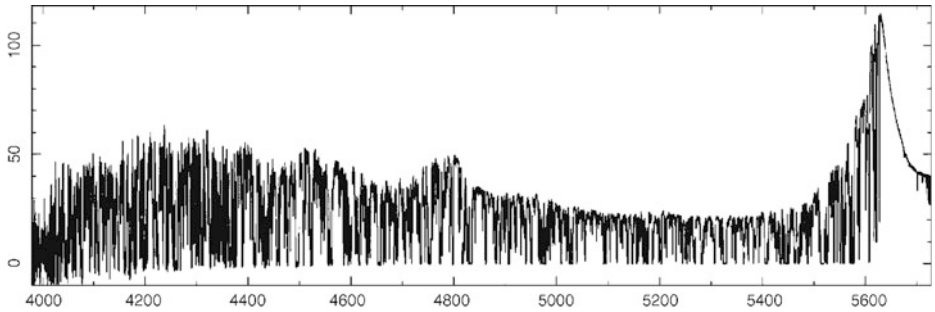

Fig. 5.55 Keck spectrum of the Lyman-α forest towards QSO $1422+231$, a QSO at $z = 3.62$. As an aside, this is a quadruply-imaged lensed QSO; it is strongly magnified by the gravitational lensing effect, so that this source is one of the brightest high-redshift QSOs—which eases obtaining high-quality spectra. The wavelength resolution is about 7 km/s. On the blue side of the Lyα emission line, a large variety of narrow absorption lines of neutral hydrogen in the intergalactic medium is visible. The statistical analysis of these lines provides information on the gas distribution in the Universe (see Sect. 8.5). Source: M. Rauch 1998, *The Lyman Alpha Forest in the Spectra of QSOs*, ARA&A 36, 267, Fig. 1, p. 268. Reprinted, with permission, from the *Annual Review of Astronomy & Astrophysics*, Volume 36 ©1998 by Annual Reviews www.annualreviews.org

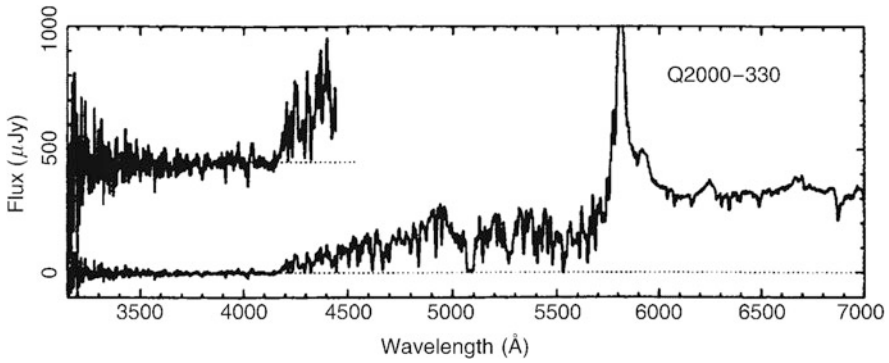

Fig. 5.56 A Lyman-limit system along the line-of-sight towards the QSO 2000−330 is absorbing virtually all radiation at wavelengths $\lambda \leq 912$ Å in the rest-frame of the absorber, here redshifted to about 4150 Å. Source: W.L.W. Sargent et al. 1989, *A survey of Lyman-limit absorption in the spectra of 59 high-redshift QSOs*, ApJS 69, 703, p. 706, Fig. 1. ©AAS. Reproduced with permission

- *Broad absorption lines:* For about 15 % of the QSOs, very broad absorption lines are found in the spectrum at redshifts slightly below $z_{\rm em}$ (Fig. 5.57). The lines show a profile which is typical for sources with outflowing material, as seen, for instance, in stars with stellar winds. However, in contrast to the latter, the Doppler width of the lines in the *broad absorption line* (BAL) QSOs is a significant fraction of the speed of light.

Interpretation. The metal systems with a redshift significantly smaller than $z_{\rm em}$ originate either in overdense regions in intergalactic space or they are associated with galaxies (or more specifically, galaxy halos) located along the line-of-sight. In fact, MgII systems always seem to be correlated with a galaxy at the same redshift as the absorbing gas. From the statistics of the angular separations of these associated galaxies to the QSO sight-line and from their redshifts, we obtain a characteristic extent of the gaseous halos of such galaxies of $\sim 25h^{-1}$ kpc. For CIV systems, the extent seems to be even larger, $\sim 40h^{-1}$ kpc.

The Lyα forest is caused by the diffuse intergalactic distribution of gas. In Sect. 8.5, we will discuss models of the Lyα forest and its relevance for cosmology more thoroughly (see also Fig. 5.58).

Broad absorption lines originate from material in the AGN itself, as follows immediately from their redshift and their enormous width. Since the redshift of the broad absorption lines is slightly lower than that of the corresponding emission lines, the absorbing gas must be moving towards us. The idea is that this is material flowing out at a very high velocity. BAL-QSOs (broad absorption line QSOs) are virtually always radio-quiet. The role of BAL-QSOs in the AGN family is unclear. A plausible interpretation is that the BAL property also depends on the orientation of the QSO.

and pressure broadening. m is the mass of the atom, which defines, together with the temperature T of the gas, the Maxwellian velocity distribution. If the natural line width is small compared to the thermal width, the Doppler profile dominates in the center of the line, that is for frequencies close to ν_0. The line profile is then well approximated by a Gaussian. In the wings of the line, the Lorentz profile dominates. For the wings of the line, where $\phi(\nu)$ is small, to become observable the optical depth needs to be high. This is the case in damped Lyα systems.

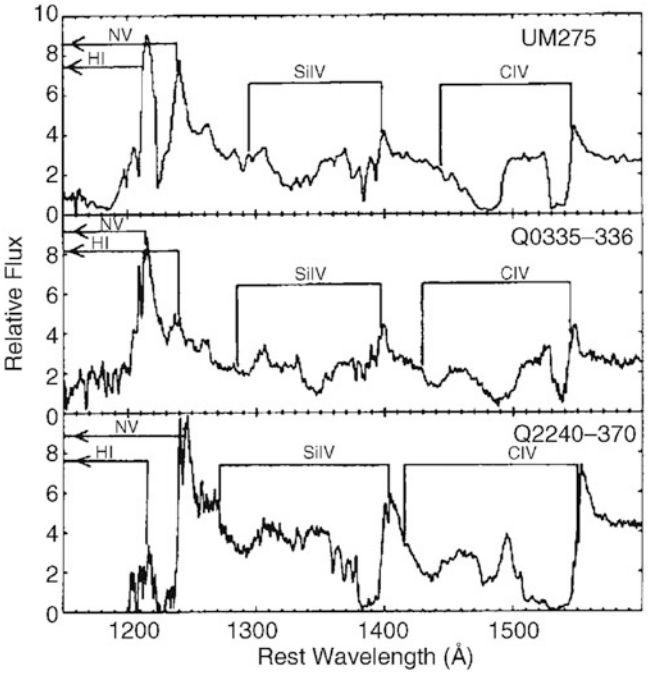

Fig. 5.57 Spectra of three BAL-QSOs, QSOs with broad absorption lines. On the blue side of every strong emission line very broad absorption is visible, such as can be caused by outflowing material. The wavelength range over which the absorption by a given line occurs is indicated by the *bars*. Such line shapes, with much lower width (of course) are also found in the spectra of stars with strong stellar winds. Source: D.A. Turnshek 1988, *BAL QSOs - Observations, models and implications for narrow absorption line systems*, in: QSO absorption lines: Probing the universe; Proceedings of the QSO Absorption Line Meeting, Baltimore, MD, Cambridge University Press, 1988, p. 17

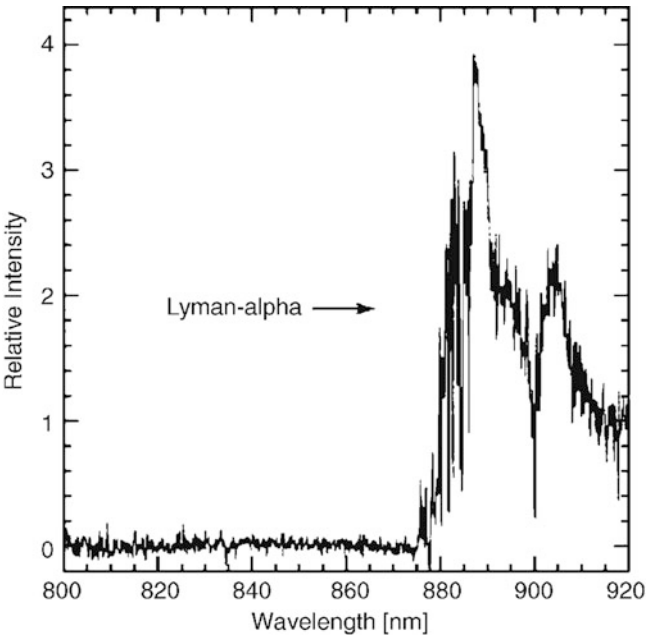

Fig. 5.58 A VLT spectrum of the QSO SDSS 1030 + 0524 at $z = 6.28$, currently one of the highest known QSO redshifts. The blue side of the Lyα emission line and the adjacent continuum are almost completely devoured by the dense Lyα forest. Credit: Laura Pentericci and Hans-Walter Rix; Max-Planck Institut für Astronomie, ESO

In this case, any QSO would be a BAL if observed from the direction into which the absorbing material streams out.

Discussion. Most absorption lines in QSO spectra are not physically related to the AGN phenomenon. Rather, they provide us with an opportunity to probe the matter along the line-of-sight to the QSO. The Lyα forest will be discussed in relation to this aspect in Sect. 8.5. Furthermore, absorption line spectroscopy of QSOs carried out with UV satellites has proven the existence of very hot gas in the halo of our Milky Way. Such UV spectroscopy provides one of the very few opportunities to analyze the intergalactic medium if its temperature is of the order of $\sim 10^6$ K—gas at this temperature is very difficult to detect since it emits in the extreme UV which is unobservable from our location inside the Milky Way, and since almost all atoms are fully ionized and therefore cause no absorption. Only absorption lines from very highly ionized metals (such as the five times ionized oxygen) can still be observed. Since the majority of the baryons should be found in this hot gas phase today, this test is of great interest for cosmology.

5.8 Problems

5.1. The spectral index of synchrotron radiation. Assume that a synchrotron source contains a population of relativistic electrons, with a power-law energy distribution, $N(E)\,\mathrm{d}E = A\,E^{-s}\,\mathrm{d}E$, where A is a constant of proportionality. The synchrotron emissivity of a single electron is a function $f(\nu/\nu_\mathrm{c})$ that depends only on the ratio of emitted frequency and the frequency ν_c given by (5.3).

1. Check that this last statement is compatible with the emitted power (5.5).
2. Calculate the shape of the synchrotron spectrum from the electron population.

5.2. Energy loss of electrons. The energy loss of an electron due to synchrotron radiation is given by (5.9).

1. Show that the energy loss due to synchrotron emission can be written as

$$\frac{\mathrm{d}E}{\mathrm{d}t} = -\frac{4}{3}\sigma_\mathrm{T} c \gamma^2 U_B \ ,$$

   where σ_T is the Thomson cross section, and $U_\mathrm{B} = B^2/(8\pi)$ is the energy density in the magnetic field.
2. If a low-energy photon with frequency ν is Compton scattered by a relativistic electron of Lorentz factor γ, its frequency after scattering is on average $\nu' = (4/3)\gamma^2\nu$.

Show that the energy loss of an electron due to inverse Compton scattering is

$$\frac{\mathrm{d}E}{\mathrm{d}t} = -\frac{4}{3}\sigma_{\mathrm{T}} c \gamma^2 U_\gamma \ ,$$

where U_γ is the energy density in the radiation field. Hence, the relative efficiency of relativistic electron cooling due to synchrotron emission and Compton scattering depends on the ratio of magnetic to photon energy density.

5.3. Spectrum of an optically thick accretion disk. Consider an optically thick accretion disk with a temperature profile given by (5.13). Neglecting any boundary effect (i.e., the fact that a real accretion disk extends only over a finite range in radii), show that the emitted spectrum is a power law, $L_\nu \propto \nu^{1/3}$. Comment: The true spectrum of an accretion disk deviates from this simple power law, mainly due to the existence of an inner boundary.

5.4. Mass growth of a black hole. Suppose that the black hole at some initial time $t = 0$ has mass $M_\bullet(0)$, and then accretes at constant efficiency ϵ as fixed Eddington ratio L/L_{edd}.

1. Show that its mass after some time t has grown to

$$M_\bullet(t) = M_\bullet(0) \exp\left(\frac{1-\epsilon}{\epsilon}\frac{L}{L_{\mathrm{edd}}}\frac{t}{t_{\mathrm{gr}}}\right) \ , \qquad (5.47)$$

where $t_{\mathrm{gr}} = M_\bullet c^2/L_{\mathrm{edd}} \approx 5\times10^8$ yr, independent of $M_\bullet$.

2. Suppose the initial (seed) mass is $M_\bullet(0) = 10M_\odot$. If the efficiency is $\epsilon = 0.1$, and the accretion occurs with Eddington luminosity, what is the black hole mass after 10^9 yr?

5.5. Properties of the BLR. Assume that the BLR is a spherical shell with characteristic radius r and thickness $\delta r \approx r$. Furthermore, assume that it consists of N_{c} clouds of radius r_{c} and electron number density n_{e}.

1. What is the covering factor of the BLR clouds as seen from the continuum source, i.e., which fraction of lines-of-sight from the center of the BLR intersect a cloud, in terms of the model parameters?
2. Calculate the filling factor, i.e., the volume fraction of the BLR that is filled with clouds.
3. Assume that the covering factor is 0.1, and that the filling factor is 10^{-6}. For a BLR radius of $r = 10^{16}$ cm and $n_{\mathrm{e}} = 10^{10}\,\mathrm{cm}^{-3}$, determine r_{c} and N_c. What is the total mass of the gas in the clouds in the BLR?

Comment: Given the uncertainty with which quantities like the covering factor can be determined, it is legitimate to neglect factors of order unity in the calculation.

5.6. Relative luminosity of AGN and host galaxy. Assume that the SMBH mass in an AGN host galaxy is 10^{-3} times the stellar mass of its spheroidal component, as found for nearby galaxies. Furthermore, assume that the spheroidal component contains a fraction f_{sph} of the total stellar mass of the galaxy. Let the AGN radiate with an Eddington ratio L/L_{edd}, and assume that 10% of the radiation comes out in the optical waveband.

1. Calculate the ratio of the optical AGN luminosity and the stellar luminosity, as a function of Eddington ratio, mass-to-light ratio of the stellar population, and the spheroidal fraction f_{sph}.
2. Discuss your result in terms of the detectability of the AGN, assuming $L/L_{\mathrm{edd}} \sim 0.1$.

5.7. Tidal disruption of a star. Show that a star of mass M_* and radius R_* can be disrupted by a SMBH only if the black hole mass $M_\bullet$ is not too large. Calculate this limiting mass for a Solar-like star.