

1 Historical background

1.1 Introduction: the energy problem

The discovery of quasi-stellar objects (QSOs or quasars) in 1963 represents a landmark in observational astronomy. Thanks to a coordination between optical and radio astronomers, it was possible to discover a new and important class of astronomical objects. Because this text book is all about quasars and related phenomena, it will not be out of place to begin at the beginning of the subject and to review briefly how these remarkable objects were first discovered.

The science of radio astronomy really began after the end of World War II, when some of the scientists and engineers engaged in wartime radar projects used their know-how to follow up the pioneering works of Karl Jansky in the 1930s and Grote Reber in the early 1940s. Thus radio dishes and interferometers appeared in England and Australia, at Jodrell Bank, Cambridge, Sydney and Parkes.

The early observations revealed the existence of cosmic radio sources and by the mid-1950s it became an accepted fact that radio galaxies exist. The nature of their radiation was non-thermal, and its polarization properties indicated that its origin lay in the synchrotron process. As we will discuss in Chapter 3, in this process radiation comes from electrons accelerated by a magnetic field. Thus a typical radio source has as its energy reservoir the dynamical energy of relativistic particles and magnetic field energy.

In 1958 Geoffrey Burbidge drew attention to the enormous size of this energy reservoir. Since we will go through his argument in Chapter 3, we simply state the result here. He found that the *minimum* energy available is of the order of 10^{60} erg! This estimate would go up further if one included the energy of the protons as well as that of the electrons. Indeed, so large was this estimate that it was the last nail in the coffin of the collision hypothesis, which sought to explain the radio source phenomenon as the outcome of colliding galaxies. The typical gravitational potential energy of a pair of colliding galaxies of masses M_1 and M_2 separated by a distance R is $-GM_1M_2/R$. For typical galactic mass $\sim 10^{11}M_\odot$, where M_\odot is the mass of the Sun, and $R \sim 10$ kpc, we get an answer $\sim 10^{59}$ erg. This calculation also assumes an optimistic situation where one energy reservoir can be converted to another with perfect efficiency, whereas astrophysical processes seldom exhibit high efficiency. The mass-energy conversion in the thermonuclear fusion of hydrogen to helium operates with an efficiency of 0.007. Thus even a 10 per cent efficiency would require the primary source of energy in a radio source to be of the order of 10^{62} erg. What can such a source be like?

2 Historical background

Burbidge himself had suggested that the source could lie in a chain reaction that triggers off one supernova after another in the nuclear region of a galaxy. Since the nuclear region is expected to be much denser than average, and supernovae do pour out a lot of energy in an explosive fashion, this idea seemed plausible. However, it was not pursued in detail.

Instead, in 1962, Fred Hoyle and William A. Fowler proposed that the nuclear region of a galaxy may permit the formation of a supermassive star having, say, a million or more solar masses which would evolve to become a ‘super-supernova’ thus generating the requisite energy from a thermonuclear source. However, by this time, Hoyle and Fowler had come to realize that a thermonuclear source of energy will not be as efficient as a gravitational one for masses of this order. This is clear from a qualitative argument to start with: the nuclear energy increases in proportion with the mass whereas the gravitational potential energy increases as the *square* of the mass. Thus while at the solar-mass level the former wins over the latter (*vide* the historical argument that dethroned the Kelvin–Helmholtz contraction hypothesis in favour of thermonuclear energy), the situation is the exact opposite for supermassive stars. But how can one tap the gravitational energy?

In a pioneering paper in *Nature* in early 1963, Hoyle and Fowler proposed that the energy for such sources was of gravitational origin, being derived from the collapse of very massive objects under their own strong gravitational fields. They pointed out that for very massive objects the internal pressures are inadequate to withstand this force of gravity. The objects therefore start contracting and as they contract the force of gravity grows, thus widening the gap between the inward and outward pressures. The situation becomes unstable, leading to the phenomenon of *gravitational collapse*. Thus the stage was set on theoretical grounds for highly collapsed supermassive objects as sources of high energy.

1.2 The discovery

We now return to the observational aspects once again. Even in the early days of radio astronomy, the importance of optical identification of a radio source was appreciated. The identification of Cygnus A had been the key observation to underscore the extragalactic nature of radio sources. Following the optical identification of a suspected extragalactic source, one can hope to do spectroscopy of the optical object and, if possible, determine its redshift. Then, by Hubble’s law (see Chapter 2) one can infer the source’s distance and hence its radio and optical luminosity.

One of the early catalogues of radio sources was the *Third Cambridge Catalogue* prepared by the Mullard Radio Astronomy Observatory at the University of Cambridge. Its sources were picked on the basis of their falling within a declination band and being brighter than 9 Jy.¹ The sources were listed in increasing order of right ascension

¹ 1 Jy (jansky) = 10^{-26} W m⁻² Hz.

with the prefix 3C. Of these, two sources, 3C 273 and 3C 48, were to play a major role in the developments of the early 1960s.

Radio astronomers at Jodrell Bank were interested in looking at the angular sizes of radio sources. In their first survey they looked at about 300 sources and found that their average size was around 30 arcsec. The size of most sources was in the range of 5 arcsec to a few arcminutes. However, a set of some 10 sources were extremely small, less than 1 arcsec in size. What could they be?

One of them was 3C 48, a source that was optically identified with a star-like object. Allan Sandage found its spectrum to be very unusual, one peculiarity being that it had strong emission lines. Also, Matthews and Sandage (1963) found the light from the object to be variable. These considerations led, in 1962, to this object's being labelled a radio star.

Meanwhile, Cyril Hazard was trying out a new method of fixing the position of a radio source very accurately in the sky. The method, which involved observing the occultation of the source by the Moon, when used on the source 3C 212 seemed to be very promising. And so he wished to try it for the compact source 3C 273, which was also in the Moon's path. Hazard proposed to observe it from Australia along with M.B. Mackey and A.J. Shimmins. The observation was successfully carried out in 1962 and the position of 3C 273 (a source with two components separated by about 20 arcseconds) was obtained with an accuracy of ~ 1 arcsec (Hazard, Mackay and Shimmins 1963).

This positional accuracy was sufficient for astronomers to identify the optical counterpart of the source. It turned out to be a star-like object of some thirteenth magnitude. However, its spectrum, taken by Maarten Schmidt, was peculiar: it had four emission lines. It turned out that they were one of the doublets of [O III], as well as the three hydrogen lines $H\alpha$, $H\beta$ and $H\gamma$. However, all four were appearing with their wavelengths increased by about 16 per cent. Later, using a spectrum scanner, Beverley Oke also found the expected line $H\alpha$ with a similar shift.

Thus the object had a redshift of 16 per cent (see the next chapter for the interpretation of redshift); and as such its original classification as a star in our galaxy went overboard! Instead of being one of the 100 billion or so stars in our galaxy, it had to be an extragalactic object whose redshift-related distance was so large that it was a hundred times brighter than an entire galaxy. Yet, it had to be compact enough to be mistaken for a star. Further, examination of old plates showed that the light from this object had also fluctuated significantly with time. Figure 1.1 illustrates the optical object with the optical jet clearly seen at its lower right-hand side.

The radio source structure of 3C 273 determined by lunar occultation showed it to be a two-component system with the components separated by 19.5 arcsec. Of the two components A and B, the optically identified object sits on component B and shows a jet directed towards the other component, A. The jet is another indication that the object is not a star but is probably a much more violent system. The radio structure on various angular scales can be seen in Figure 9.13.

Cambridge University Press

978-0-521-47989-9 - Quasars and Active Galactic Nuclei: An Introduction

Ajit K. Kembhavi and Jayant V. Narlikar

Excerpt

[More information](#)

4 Historical background

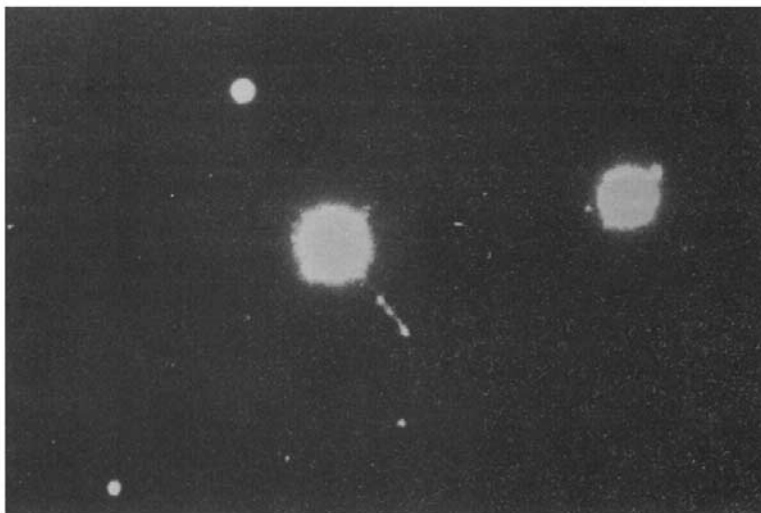


Fig. 1.1. The quasar 3C 273 in the optical band. Reproduced from Narlikar (N93).

Alerted by these unusual features in 3C 273, astronomers took a second look at the object 3C 48 and found that its spectrum had looked unusual (or, unfamiliar) because it too was redshifted, and by an even larger degree, around 36.7 per cent.

It should be recalled that in the early 1960s the measured redshifts of galaxies usually ranged up to about 0.2, i.e., 20 per cent. The galaxy identified with the radio source 3C 295 had a redshift of 0.46, which was then the record! Thus the large redshifts of these two unusually compact objects immediately drew the attention of theoreticians. Further, this discovery, coming as it did in early 1963, coincided with the theoretical expectations of Hoyle and Fowler that highly compact massive objects could serve as high energy reservoirs.

The role of theoretical inputs was considered so important that an international symposium was convened in Dallas, Texas in December 1963 to bring together general relativists, theoretical astrophysicists and observers to discuss the implications of the discovery of these remarkable objects. This was to be the first of a continuing series of biennial symposia under the name 'Texas Symposia'. The objects became known as *quasi-stellar objects (QSO)* or *quasars*. We will use either of these names in this text.¹

1.3 The nature of redshift

The quasars 3C 48 and 3C 273 had large redshifts in the context of the galactic redshifts known in the early 1960s. Subsequently, however, quasars with even higher redshifts began to be discovered. The source 3C 9 was the first very large redshift object to

¹ See the discussion in Section 6.1.

be found, with redshift $z = 2$. In fact, a large number of quasars with redshifts in this neighbourhood became known during the first decade of the discovery of quasars. These discoveries were also accompanied by theoretical discussions on their nature.

In the 1960s the astronomer was familiar with three kinds of redshift: (1) the Doppler shift arising from the relative motion between the source of light and the observer, (2) the gravitational redshift of light travelling from a strong to a weak gravitational field and (3) the redshift arising from the expansion of the universe. In the early days of quasar astronomy, all three interpretations were tried, before the consensus settled in favour of the last of these. It may be worth looking at the pros and cons of these different hypotheses briefly. We will return to this issue in a more detailed fashion in Chapter 15.

1.3.1 Doppler shift

If a source of light moves with velocity \mathbf{v} relative to the observer so that the velocity vector makes an angle θ with respect to the radial vector from the observer to the source, the spectral shift of the light from the source measured by the observer will be given by

$$1 + z = \frac{1 + (v/c) \cos \theta}{\sqrt{1 - (v/c)^2}}, \quad (1.1)$$

where $v = |\mathbf{v}|$ and c is the speed of light. For small velocities, this formula reduces to the Newtonian limit $z = v/c$ for radially outward motion.

For stellar motion in our galaxy it is common to find small Doppler shifts both positive and negative and these are interpreted in terms of stars moving away from or towards us. James Terrell in 1966 interpreted the quasar redshifts as arising from fast ejection of quasars from the galactic centre. The problem with this concept was that it endowed the galactic centre with violent activity, which was not consistent with the undisturbed motions of stars in the area. This idea was taken further by Hoyle and Burbidge (1966), who suggested that, unlike the case for our galaxy, ejection from a galactic nucleus that shows violent activity would appear normal. In particular, they suggested NGC 5128 as a likely site for ejection.

The Doppler hypothesis thus delinks redshift from distance; that is, a large redshift does not imply that the source is very distant. This eases the energy budget of a typical source. However, if several such sources are to be ejected from a single site (such as NGC 5128) the energetics of that site becomes problematical. Moreover, there is the problem of blueshifts.

The problem was first highlighted by P. Strittmatter in 1966 (unpublished). A site of explosion will eject quasars in all directions. Thus, unless the explosion occurred a long time ago, some quasars would still be travelling *towards* the observer and these should show blueshifts. Moreover, these quasars would appear brighter due to the blueshift effect and would tend to dominate in a sample that is complete with respect to all quasars brighter than a specified flux level. Detailed calculations based on isotropic

Cambridge University Press

978-0-521-47989-9 - Quasars and Active Galactic Nuclei: An Introduction

Ajit K. Kembhavi and Jayant V. Narlikar

Excerpt

[More information](#)

6 Historical background

ejection of all such quasars show that the ratio of blueshifted (N^-) quasars to redshifted (N^+) quasars in a flux-limited sample is

$$\frac{N^-}{N^+} = (1 + z_m)^{2.5+1.5\alpha}, \quad (1.2)$$

where z_m is the maximum redshift in the sample, and α is the spectral index of a typical source. Thus, for a maximum redshift of 2, say, and a spectral index unity, we may expect 81 times as many blueshifted sources as redshifted ones!

We will return to this problem in Chapter 15. For the time being we leave the topic of Doppler shifts as it was perceived in the early sixties, and move on to the next alternative.

1.3.2 Gravitational redshift

Einstein's general theory of relativity drew the attention of astronomers to this source of redshift: indeed, the early demonstrations of this effect were for the white dwarf stars Sirius B and δ Eridani B. In either case the effect was very small, being less than 0.001. Theoretically one could, on the basis of Schwarzschild's solution, define the surface redshift of a mass M with radius R by the formula

$$1 + z = (1 - 2GM/Rc^2)^{-1/2}. \quad (1.3)$$

Thus, by having R arbitrarily close to the critical value $2GM/c^2$, it may seem possible to have large redshifts. This is not, however, the case.

In 1964 Hermann Bondi showed that whatever equation of state is chosen, provided it is physically realistic (i.e., with sound speed in the material not exceeding the speed of light), the surface redshift cannot exceed 0.62. Thus surface gravitational redshift proved inadequate to explain the large redshifts of quasars. There was also an observational objection put forward by J. Greenstein and M. Schmidt (1964). We briefly reproduce their elegant argument as applied to the quasar 3C 273.

For small redshifts we get from the above formula

$$z \simeq \frac{GM}{Rc^2} \simeq 1.47 \times 10^5 \frac{M}{M_\odot} \frac{R}{1 \text{ cm}}. \quad (1.4)$$

Consider a thin shell of radius R and thickness ΔR emitting radiation in a line of width w . For line widths in 3C 273 arising from the slight change of redshift over ΔR we get

$$\frac{\Delta R}{R} = \frac{w}{\Delta \lambda} = 0.07. \quad (1.5)$$

Now suppose that the object is of stellar type with mass $\sim M_\odot$. For such an object a redshift 0.158 as in 3C 273 gives a radius of $\sim 10^6$ cm, and so $\Delta R \sim 7 \times 10^4$ cm.

Next we determine the volume of the shell and use the volume emissivity of the H β line to estimate the flux of radiation in that line. For the source 3C 273 the volume emissivity at a temperature $\sim 10^4$ K is $\sim 10^{-25} n_e^2$ erg sec $^{-1}$ cm $^{-3}$, where n_e is the electron

1.3 The nature of redshift

7

number density. Multiplying this with the volume of the shell and equating the result to the observed luminosity we get

$$10^{-25} R^2 \Delta R n_e^2 = 3.4 \times 10^{-12} d^2, \quad (1.6)$$

where d is the expected distance of the source measured in centimetres. This gives

$$n_e^2 = 5.10 \times 10^{-4} d^2. \quad (1.7)$$

Since the searches for proper motion of 3C 273 proved negative, Greenstein and Schmidt assumed that the source could not be nearer than 100 pc. This gave

$$n_e > 6 \times 10^{18} \text{ cm}^{-3}. \quad (1.8)$$

This density is too high, and is precluded by the appearance of the forbidden line [O III] λ 5007 in the source spectrum. Thus the stellar possibility is ruled out. Similarly Greenstein and Schmidt were able to reject the alternative hypothesis that the source is much more massive, extragalactic and located within, say, 25 Mpc.

To get round this argument and the Bondi limit, Hoyle and Fowler (1967) proposed a rather unusual model of a quasar in which the redshift arose from light coming from the interior. In this model the gravitational potential well at the centre of the object is provided by a distribution of a large number of very compact objects (such as, say, neutron stars). Any gas present in the system would descend to the centre and settle down there and form the emission line region. It is not difficult to see that the gravitating compact objects would leave a sufficient gap for the central radiation to emerge with redshifts much higher than the Bondi limit. The interior Schwarzschild solution used by Hoyle and Fowler to demonstrate this effect gives arbitrarily high central redshifts. However, more realistic models discussed by P.K. Das (1977) yielded redshifts of the order of 2–3. These models also get around the Greenstein–Schmidt criticism provided the masses are of galactic order.

We will take a second look at the gravitational redshift option in Chapter 15, when considering other evidence that has a bearing on the nature of redshifts.

1.3.3 Cosmological redshift

The third alternative, which is often referred to as the *cosmological hypothesis*, has occupied the central place in this field. In this hypothesis our Galaxy, along with other galaxies and any other extragalactic objects, is located in an expanding universe and the redshift of any such object is explained as arising from the time-dilation produced in the curved space–time of such a universe. This hypothesis explains the Hubble law (see Chapter 2) observed for galaxies, namely that the redshift of a source is proportional to its distance from us. Right from the early days this has been the favourite hypothesis for quasars also, for the following reasons.

- The redshifts of quasars, although by and large greater than those of galaxies, seem to arise as a natural consequence of the expanding universe hypothesis.

Cambridge University Press

978-0-521-47989-9 - Quasars and Active Galactic Nuclei: An Introduction

Ajit K. Kembhavi and Jayant V. Narlikar

Excerpt

[More information](#)

8 *Historical background*

If this is true, no detailed structural scenarios are needed (as in the other two options discussed above) to understand the redshift.

- Quasar properties are in many respects similar to those of the nuclei of Seyfert galaxies, and of active galactic nuclei in general. Since the redshifts of these galaxies and nuclei are believed to be cosmological, the redshifts of quasars can likewise assumed to be so.

All the same there still persist some doubts about the universality of the cosmological hypothesis and we will return to the contentious issues in Chapter 15. For the time being, and in the following chapters, we will assume that this hypothesis holds for the quasars.

1.4 **Physical characteristics: quasars**

Keeping in mind the serendipitous discovery of the first two quasars, we need to specify the broad criteria that would help distinguish a quasar from a star or a galaxy. In the classic book *Quasi-Stellar Objects* by Geoffrey and Margaret Burbidge, written in the early years of quasar searches, the following criteria specified by Maarten Schmidt were given:

- star-like object identified with a radio source;
- variable light;
- large ultraviolet flux of radiation;
- broad emission lines in the spectra with absorption lines in some cases;
- large redshift.

Later searches showed that the radio source property is not generic and only some 10 per cent of quasars may be radio-loud. We shall discuss this property of quasars in Chapter 9. In fact, X-ray emission seems more common with quasars, as became clear with the studies of the *EINSTEIN* observatory in 1980. Time variability has been found in several wavebands, e.g. visible, radio, X-ray etc. In the 1970s very long baseline interferometry revealed several cases of apparent superluminal separation of the components of radio-loud quasars on the scale of parsecs. We shall consider again in Section 6.1 the defining characteristics of quasars and other related objects.

The spectra of quasars have been used to get information on the physical environment in the universe at large redshifts. This environment can be in the quasar itself or in its vicinity or in the intermediate region between it and us (the observers). In particular, the absorption lines may arise from absorbing material in the quasar itself or in the intervening medium. With the discovery of high redshift galaxies this property is no longer the prerogative of quasars.

Finally, in 1979 the first case of the gravitational lensing of quasars came to light. Two quasar images named PHL 0957+561 A and B seemed so similar in physical properties that it became plausible to argue that they were virtual images of only one real object. The multiple imaging could be caused by the bending of light from

Cambridge University Press

978-0-521-47989-9 - Quasars and Active Galactic Nuclei: An Introduction

Ajit K. Kembhavi and Jayant V. Narlikar

Excerpt

[More information](#)

1.5 Physical characteristics: active galactic nuclei

9

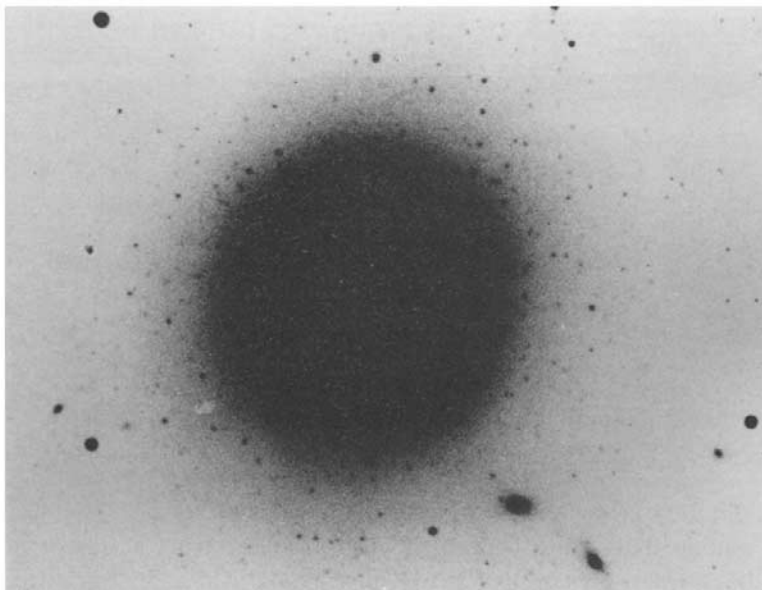


Fig. 1.2. The galaxy M87 at optical wavelengths. Reproduced from *The Cambridge Atlas of Astronomy*, second edition (1988).

the source to the observer by an intervening galaxy (or a cluster of galaxies, or some massive lump of dark matter), allowing more than one path. Gravitational lensing has since been proposed for several multiple quasar systems with members having very close angular separations ($\lesssim 5\text{--}6$ arcsec).

We will discuss the details and implications of all these physical characteristics in later chapters.

1.5 Physical characteristics: active galactic nuclei

In Figure 1.2 we see an optical photograph of the galaxy M87. This galaxy is of Hubble type E0, with the peculiar feature that again a jet appears to come out of it. The emergence of the jet is an indication of some violent activity in the nucleus, which, as mentioned above, is of an exceptional nature in the context of galaxies. Normally one would expect that more stars would congregate towards the galactic centre, which might lead to greater velocity dispersion and greater brightness compared to the outer regions. However, when in the late 1970s, with one of the early uses of charge coupled device (CCD) cameras, the nucleus of M87 was looked at more closely, it showed a sharp rise in the luminosity towards the centre (see Figure 1.3). This rise could be explained by supposing that there is a massive attractor at the centre, which pulls a large population of stars into a compact region, thereby pushing up the luminosity of the region.

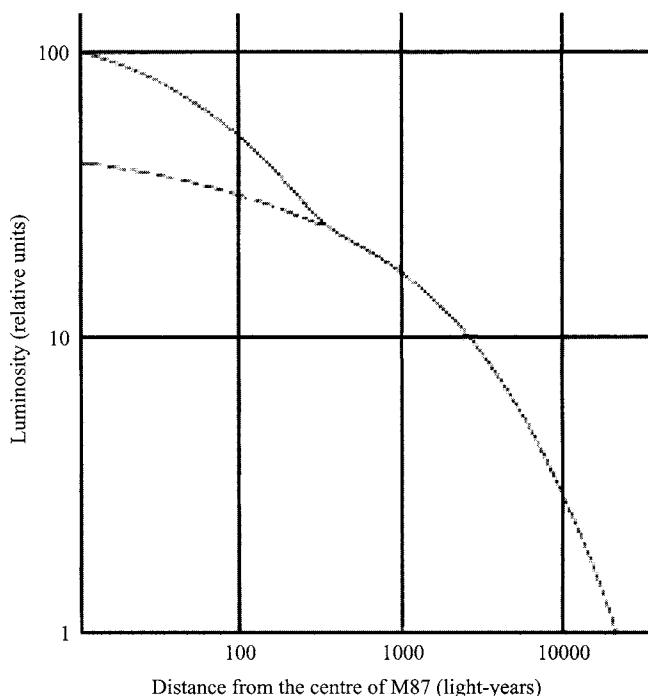
10 *Historical background*

Fig. 1.3. The luminosity profile of M87. Reproduced from Narlikar (N96).

Around the same time the velocity dispersion of stars in the nuclear region was also measured and it also seemed to rise rapidly inwards. This too suggested the presence of a massive object in the nucleus. The conclusion from both these studies was that this central attractor may well be a massive black hole of $\sim 5 \times 10^9 M_{\odot}$. The galaxy M87 is also a strong emitter of X-rays and thus requires an energy production mechanism that goes beyond the typical stellar energy production found in normal galaxies.

There is thus a similarity between the nucleus of M87 and a quasar-like 3C 273, in that both are seats of intense energy production, both require a collapsed massive object, both are X-ray sources and, in this particular case, both have jets issuing from the central region. As has become clearer over the years, there is a whole class of galactic nuclei of this kind that show signs of energetic activity. These are collectively referred to as *active galactic nuclei* (AGN).

Indeed, a class of galaxies originally studied by C. Seyfert show the unusual quasar-like feature of emission lines from their centres. Seyfert galaxies, as these galaxies are called, probably form no more than ~ 2 per cent of the class of spiral galaxies. The Seyferts are further subdivided into two classes, Seyfert 1 and Seyfert 2, depending on the width of the emission lines found in them. The distinction made in the literature between a quasar and a Seyfert 1 nucleus is largely morphological: in the latter a galactic envelope is seen whereas a quasar is generally star-like, without a nebulosity.