1

Black Holes and Galaxies

1.1 Basic Properties

A black hole is an object so dense that light cannot escape its gravity. The proper description of black holes requires general relativity (GR), and we discuss this in Chapter 2. But simple Newtonian ideas already give us some insight into their properties if we bear in mind the restrictions to velocities below the speed of light and energies smaller than the rest-mass value.

The escape velocity from the surface of a star of mass M and radius R is $v = (2GM/R)^{1/2}$, where G is the gravitational constant. This velocity reaches the speed of light, c, for a radius

$$R = \frac{2GM}{c^2}. \tag{1.1}$$

We see that for $M = 1\mathrm{M}_\odot$ (the solar mass), the 'star' must have a tiny radius $R \lesssim 3\,\mathrm{km}$. The characteristic property of a black hole is that it is small for a given mass, making the gravitational field very strong in its immediate vicinity. But it is important to remember that at large distances from a black hole, the gravitational field strength is the same as for any gravitating object of the same mass. Black holes have their distinctive properties only because they are small enough to allow matter to get very close, so that orbital speeds approach that of light. The characteristic size of a black hole is its gravitational radius

$$R_g = \frac{GM}{c^2} \simeq 1.5 \times 10^{14} M_8\,\mathrm{cm}, \tag{1.2}$$

where we have parametrized the black hole mass $M = 10^8 M_8 \mathrm{M}_\odot$, as this is a typical SMBH mass.

Matter falling radially towards an object like this acquires very high speeds because of the large gravitational potential energy available near the black hole. If nothing intervenes to stop it getting very close to the hole the matter eventually gains gravitational energy approaching $\sim 0.5c^2$ times its rest mass. If the matter is gaseous, it is very likely that the process is sufficiently untidy that much of this energy is dissipated as radiation. The luminosity this releases is far greater than would happen if the same mass was consumed in nuclear burning. Transmuting hydrogen to helium or heavier elements, which powers most stars, releases energy only $0.007c^2$ times the rest mass.

In reality two things complicate this comparison a little. First, in reality matter is almost certain to orbit the black hole with some angular momentum, and fall towards it more slowly, as it gradually loses this angular momentum to matter further out. This kind of configuration is called an *accretion disc*, and will appear everywhere in this book. Accretion through a disc gives up the gravitational binding energy of orbits close to the black hole, which is somewhat less than the radial infall kinetic energy (a factor of two for circular Newtonian orbits).

Second, GR changes these binding energies slightly from Newtonian values. A full GR calculation (see Section 4.1) refines the estimate of the infall energy $\sim 0.5c^2$ to a value ~ 0.1–$0.4c^2$, but this slightly reduced ‘accretion yield’ is still far higher than the nuclear yield $0.007c^2$. Matter–antimatter annihilation releases the full rest-mass energy, but is very unlikely on any scale larger than an atomic nucleus, so that accretion on to a black hole is the most efficient way of getting energy from normal matter. We conclude that

> *accretion on to black holes must power the most luminous objects in the Universe.*

The obvious candidates here are quasars and active galactic nuclei, collectively called AGN, which harbour the most massive black holes. Their luminosities can reach 10^{46}–10^{48} erg s^{-1} or even more.[1] At typical distances of Mpc, the angular sizes of their gravitational radii R_g are extremely small. But radio interferometry with extremely long baselines is now able to resolve some of the nearer AGN, giving spectacular images (e.g. Figure 1.1).

At a smaller mass scale, the same argument tells us that binary systems where a stellar-mass black hole accretes gas at a high rate from a companion star are good candidates for explaining some of the stellar-mass X-ray binaries, with luminosities up to 10^{38}–10^{39} erg s^{-1}. Of course there are no AGN analogues of neutron stars, which cannot have masses larger than about $3\mathrm{M}_{\odot}$.[2]

[1] Gamma-ray bursts can for extremely short times exceed this luminosity, and reach $\sim 10^{53}$ erg s^{-1}. This approaches the total luminosity L_* of all the stars in the observable Universe, which contains $\sim 10^{11}$ galaxies, each with $\sim 10^{11}$ stars emitting roughly solar luminosities $\sim \mathrm{L}_{\odot} \sim 10^{33}$ erg s^{-1}. This gives $L_* \sim 10^{55}$ erg s^{-1}. See Problem 1.1 at the end of the book which investigates if this is a coincidence.

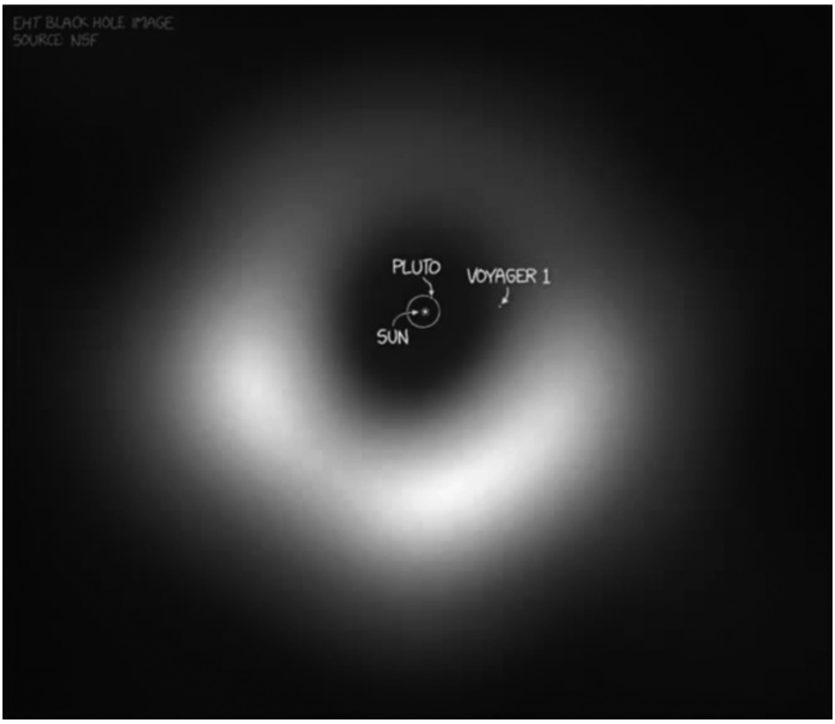

Figure 1.1 The immediate surroundings of the supermassive black hole in the galaxy M87 as imaged in the radio by the Event Horizon Telescope (Event Horizon Telescope Collaboration, 2019). The scale of the solar system is shown for comparison. Credit: Randall Munro (2019).

Observations clearly distinguish between stellar-mass accretors and SMBH. X-ray binaries are in spatially extended populations in their host galaxies, while AGN are point sources close to the centres of their hosts, and are generally intrinsically far brighter than X-ray binaries. They cannot simply be unresolved collections of X-ray binaries because they are often observed to vary by factors of $\gtrsim 2$.

1.2 The Eddington Limit

The accretion luminosity L of a black hole must be related to its mass accretion rate $\dot{M}$ by

$$L = \eta \dot{M} c^2, \tag{1.3}$$

[2] The minor complication here is that neutron stars are almost as compact as black holes – they have radii ~ 10 km for masses ~ 1–$3\mathrm{M}_\odot$. Both black hole and neutron star X-ray binaries are fairly common, although there are probably far more neutron star systems in total.

where we see from the discussions of the last previous section that the *accretion efficiency* η is a dimensionless quantity of order at most a few times 0.1. But there is a limit to the luminosity that any gravitating object, accreting or otherwise, can emit, since radiation produces a pressure force which tends to disperse the matter producing the luminosity. This force acts on electrons because they scatter electromagnetic radiation, which carries momentum $(1/c)$ times its energy flux $L/4\pi r^2$. Protons have little effect on radiation, but make up most of the mass of the gas. For simplicity we consider a spherically symmetric situation, so that the radiation pressure force acts radially outwards. Its magnitude at radius r from the centre is

$$F_{\rm rad} = \frac{L\sigma_T}{4\pi cr^2}, \tag{1.4}$$

where $\sigma_T \simeq 6.65 \times 10^{-25}\,{\rm cm}^2$ is the Thomson cross-section, the effective blocking area of an electron in a beam of radiation. The electron is not free to move in response to this outward force, since charge neutrality means that it is strongly bound by Coulomb attraction to a mass of gas carrying one proton charge. Most astrophysical gases are largely hydrogen, so this mass is of order the proton mass m_p. Then the gravity force resisting the radiation pressure is

$$F_{\rm grav} \simeq \frac{G(m_p + m_e)}{r^2} \simeq \frac{GMm_p}{r^2}, \tag{1.5}$$

since the electron mass m_e is much smaller than m_p. Both $F_{\rm rad}$ and $F_{\rm grav}$ vary as r^{-2}, so we see that if one of them exceeds the other at any one radius, it does so at all radii. Then in spherical symmetry, accretion must be at least inhibited once L is large enough to make $F_{\rm rad} = F_{\rm grav}$. This defines the *Eddington luminosity* or *Eddington limit* as

$$L_{\rm Edd} = \frac{4\pi GMc}{\kappa}, \tag{1.6}$$

where we have used the electron opacity $\kappa \simeq \sigma_T/m_p \simeq 0.34\,{\rm cm}^2\,{\rm g}^{-1}$ for an astrophysical gas of typical composition, giving

$$L_{\rm Edd} \simeq 1.3 \times 10^{46} M_8\,{\rm erg\ s}^{-1}. \tag{1.7}$$

(The opacity κ is roughly halved, so $L_{\rm Edd}$ doubled, if the accreting gas is hydrogen-poor.) This result, and the huge difference in the mass that is accreting gas, explains why AGN can be far more intrinsically luminous than X-ray binaries.

Although we have derived it here for spherically symmetric systems, the limit (1.6) holds to factors of order unity for almost any other geometry. In particular, this is true even for matter falling in through a sequence of orbits of decreasing angular momentum – that is, in an *accretion disc*. The appropriate form of the Eddington limit applies to any luminous object, whatever powers its luminosity.

The importance of the Eddington luminosity was first realized a century ago in the context of stellar structure, where the radiation comes from nuclear burning. Massive hot stars radiate luminosities close to the limit (1.7). A nuclear luminosity even slightly above $L_{\rm Edd}$ would make them expand a little, lowering the density ρ in the central nuclear-burning core. The nuclear luminosity $L_{\rm nuc}$ varies as ρ^2 and so drops below $L_{\rm Edd}$. This self-limiting property means that hot stars can remain stably in equilibrium very close to $L_{\rm Edd}$. The source of the luminosity reacts sensitively – and negatively – to the luminosity itself, rather like a thermostat.

But this kind of self-limiting behaviour does not apply to accretion-powered objects. The mass supply rate driving accretion is in general given by some process totally unaffected by changes in the accretion luminosity, and so is unlikely to adjust to respect the Eddington limit. For example, in close binary systems there is no reason why the evolution of the donor star, and so the resulting mass transfer rate $\dot{M}_{\rm supp}$ it supplies to a companion black hole, should know or care about the possibility that $\dot{M}_{\rm supp}$ might exceed the rate

$$\dot{M}_{\rm Edd} = \frac{L_{\rm Edd}}{\eta c^2} = \frac{4\pi GM}{\eta\kappa c} \tag{1.8}$$

that would make the accretion luminosity $L_{\rm acc} = \eta c^2 \dot{M}_{\rm acc} = \eta c^2 \dot{M}_{\rm trans}$ greater than $L_{\rm Edd}$. This possibility was already recognized in the very first papers discussing realistic accretion processes: it is entirely possible for a black hole (or any other accreting object) to be supplied with mass at rates $\dot{M} > \dot{M}_{\rm Edd}$, and for this situation to persist over significant timescales. We will discuss in detail what happens in such cases in Section 4.6, but it already is clear that there are only two routes to dealing with the mismatch – either

(a) preventing much of the matter getting too close to the hole, where it would gain and then radiate the full accretion energy, or
(b) ensuring that the matter close to the hole has unusually low accretion efficiency and so does not radiate a strongly super-Eddington luminosity.

The mild complexity of these two possibilities has generated a cloud of confused and confusing language in the astrophysical literature. The phrase ‘super-Eddington accretion’ is deeply ambiguous if not carefully qualified, as it is often used to denote either one of the outcomes (a, b) (or in the worst cases, both simultaneously!). To avoid this ambiguity, this book uses the description super-Eddington *mass supply* (or *feeding*) to describe cases where $\dot{M}_{\rm supp} > \dot{M}_{\rm Edd}$. In treating these, it is vital to distinguish between the outcomes (a) and (b), which differ markedly.[3]

[3] This muddle is maximal in discussions of the (stellar-mass) ultraluminous X-ray sources (ULXs). ULXs have very anisotropic radiation patterns (see Section 4.10), and when viewed from tightly defined directions appear

In case (a), the black hole mass cannot grow faster than the rate $\dot{M}_{\rm Edd}$. Then the shortest e-folding timescale for mass growth is the Salpeter timescale

$$t_{\rm Sal} = \frac{M}{\dot{M}_{\rm Edd}} \simeq 5 \times 10^7 \eta_{0.1}\ {\rm yr} \tag{1.9}$$

(Salpeter, 1964), where $\eta_{0.1}$ is the efficiency of conversion of rest-mass energy to radiation in units of $0.1c^2$. We see that high radiative efficiency implies slower mass growth, as the limiting luminosity $L_{\rm Edd}$ is produced by a smaller accretion rate.

If instead (b) holds, the accretor mass grows faster than the Eddington rate, that is, on a timescale $< t_{\rm Sal}$.

In both cases (a) and (b) it is difficult for the total luminosity (correctly evaluated over all directions in the case of anisotropy – see footnote 3) of an accreting object to exceed the Eddington luminosity $L_{\rm Edd}$ by large amounts, except in impulsive or explosive situations such as supernovae or gamma-ray bursts. If we are confident that a given object is not of this type, and its luminosity is not markedly anisotropic, its luminosity gives us a lower limit to its mass through (1.7).

1.3 SMBH Accretion

By the argument detailed previous section, observation places tight constraints on the total mass in SMBHs in the local (low-redshift) Universe. AGN spectra (cf. Figure 1.2) typically peak in the soft X-ray–far UV region, with almost always a significant component in the medium-energy X-rays. This latter component is relatively easy to observe, as it is often fairly immune to interstellar absorption or scattering. In addition, very few astronomical objects other than those that accrete produce substantial X-ray emission, so it is usually safe to assume that all the detected emission comes from the AGN itself.

Medium-energy cosmic X-ray detectors find non-zero fluxes even when not observing specific point sources – this is called the X-ray background. Since the emission from AGN in this spectral band is far more powerful than from anything else, such as X-ray binaries or supernovae, we can identify this background flux as the result of the collective emission from AGN – that is, growing SMBH – in the local Universe. From the typical X-ray spectrum (cf. Figure 1.2) this gives us the total SMBH growth in this region, and so a lower limit on the total SMBH mass there. This is in outline the *Soltan argument* (Soltan, 1982). It tells us that the

to have luminosities $\gg L_{\rm Edd}$, if the observed flux is (wrongly) assumed to be isotropic. Confusingly, the indirect cause of the anisotropic emission, and so of the apparent super-Eddington luminosity, is super-Eddington *feeding*, since radiation pressure blows matter away from the accretion disc (see (a)) except near the rotational axes of the accretion flow, so collimating the escaping radiation tightly. The accretors in ULXs do not gain mass at significantly super-Eddington rates, even though the mass transfer rates from their companions are super-Eddington. Their luminosities are *apparently* super-Eddington, but in reality are not.

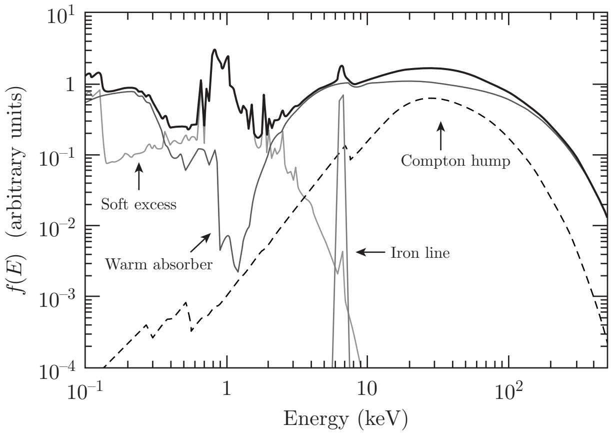

Figure 1.2 Average total spectrum (thick line) and main components (thin lines) in the X-ray spectrum of a type I AGN. The main primary continuum component is a power law with a high energy cut-off at $E \sim$ 100–300 keV, absorbed at soft energies by warm gas with $N_H \sim 10^{21}$–10^{23} cm^{-2}. A cold reflection component is also shown. The most relevant narrow feature is the iron Kα emission line at 6.4 keV. Finally, a 'soft excess' is shown, resulting from thermal emission of a Compton thin plasma with temperature $kT \sim$ 0.1–1 keV. Credit: Risaliti and Elvis (2004).

average ration of SMBH mass is $\gtrsim 10^8 \mathrm{M}_\odot$ per medium-size galaxy. This mass scale agrees with the values we can deduce from the Eddington limit (1.7), and we shall see that it is similar to the masses found by various kinds of direct observation, so it cannot be concentrated in just a minority of galaxies. We conclude that

> *the centre of almost every medium to large-mass galaxy in the local Universe must host a supermassive black hole, whose mass grew via luminous accretion of gas.*

Despite this, observations show that only a minority (less than 1%) of low-redshift galaxies have active nuclei, where the SMBH are currently growing their masses. But since we have just concluded that almost every galaxy has an SMBH, this must mean that accretion and SMBH growth occur only in short-lived phases – AGN must be strongly variable, not simply on the timescales we observe directly, but on longer ones also.

1.4 SMBH Locations

The reasoning of Section 1.3 implies that almost all low-redshift galaxies must host SMBH. Observations of the active minority where the SMBH is caught in the act of accreting always find the AGN close to the dynamical centre (hence the 'nucleus' part of 'AGN'). This is no accident, but results because the gravity of an SMBH moving through a galaxy makes it pull large numbers of stars along behind it, in a kind of gravitational wake (see Figure 1.3). This process is called 'dynamical friction', and is directly analogous to the way the Coulomb attraction of a charged particle slows its motion through a plasma. The result (Chandrasekhar, 1943) is a drag force – the speed v of an SMBH of mass M moving through a galaxy with stellar mass density ρ obeys

$$\frac{\mathrm{d}v}{\mathrm{d}t} = -\frac{4\pi CG^2M\rho}{v^2}. \tag{1.10}$$

Here $C \simeq 10$ is a constant (the Coulomb logarithm, measuring the cumulative effect of the weak drag forces of many distant stars ('small-angle scattering') compared with the individually stronger drag forces ('large-angle scattering') of a few nearby stars). If ρ is constant this equation integrates as

$$v^3 = v_0^3\left(1 - \frac{t}{t_{\mathrm{fric}}}\right) \tag{1.11}$$

where

$$t_{\mathrm{fric}} = \frac{v_0^3}{12\pi CG^2M\rho}, \tag{1.12}$$

with v_0 the initial velocity. In a time $\sim t_{\mathrm{fric}}$ the SMBH is reduced to rest, which is only possible if it has spiralled in to the dynamical centre of the galaxy. Crudely modelling the central region of a galaxy as a uniform sphere of stars with total mass $M_* = 10^{11}\mathrm{M}_\odot$ and radius R_* = a few kpc, a $10^8\mathrm{M}_\odot$ SMBH with speed $v_0 \sim (2GM_*/R_*)^{1/2}$ has $t_{\mathrm{fric}} \lesssim 10^8$ yr, much smaller than the age $\sim 10^{10}$ yr of a low-redshift galaxy.

So unless a galaxy has recently been disturbed, which is often obvious because it has an irregular shape, its SMBH is very likely to be at its dynamical centre. If it somehow has more than one SMBH, dynamical friction makes them collect rapidly in its centre, and then orbit under their mutual gravitational attraction. Here they lose energy and spiral inwards as they emit gravitational radiation, eventually merging. The merger may eject the lightest hole(s) if there are more than two, since these must be moving fastest if the holes have similar orbital energies, as is likely. So in most galaxies we expect to find just one SMBH, in its centre. (In very small galaxies the shallow gravitational potential may be unable to retain a merging pair of black holes as they recoil under anisotropic gravitational wave emission, so some may have no central black hole). By the same reasoning, a merger of two galaxies is likely to produce a single larger galaxy with a merged SMBH in its centre (see Figure 1.4).

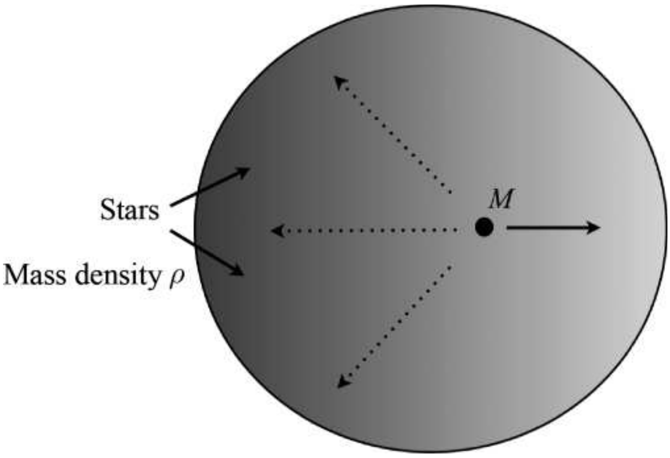

Figure 1.3 dynamical friction: a massive object moving through a collection of stars raises a gravitational ‘wake’ and slows.

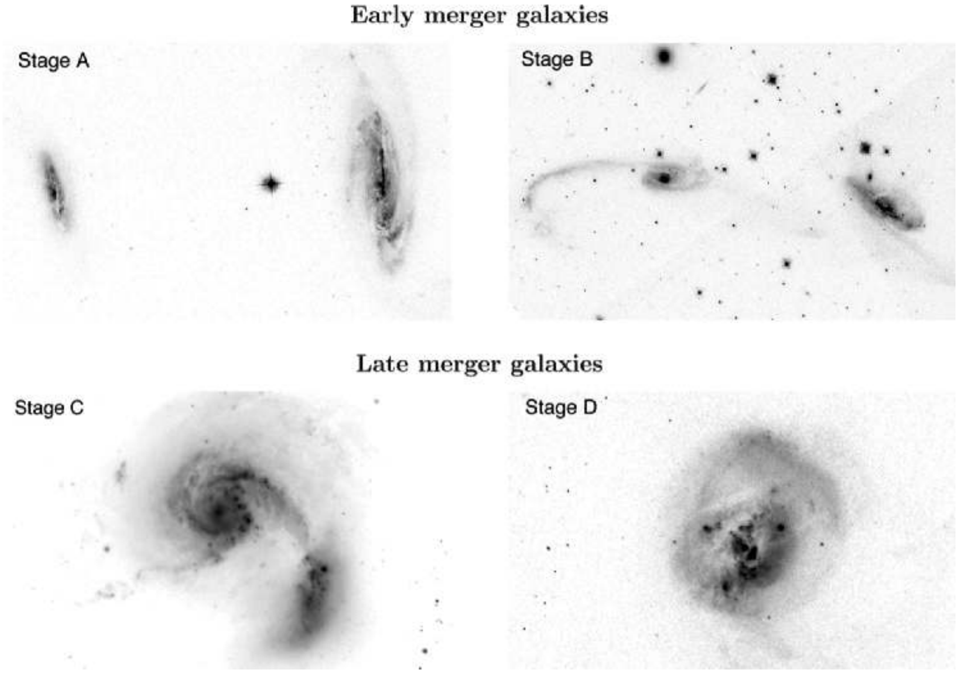

Figure 1.4 Galaxies in various stages of merging. Stages A and B are early mergers and stages C and D are late mergers. Credit: Ricci et al. (2017).

In summary, we expect almost every galaxy, except possibly a few dwarfs, to have a single SMBH at its centre. Observations appear to agree with this very simple one-to-one relation. Our own Galaxy’s central region is well studied. Infrared

Cambridge University Press & Assessment
978-1-108-48805-1 — Supermassive Black Holes
Andrew King
Excerpt
More Information

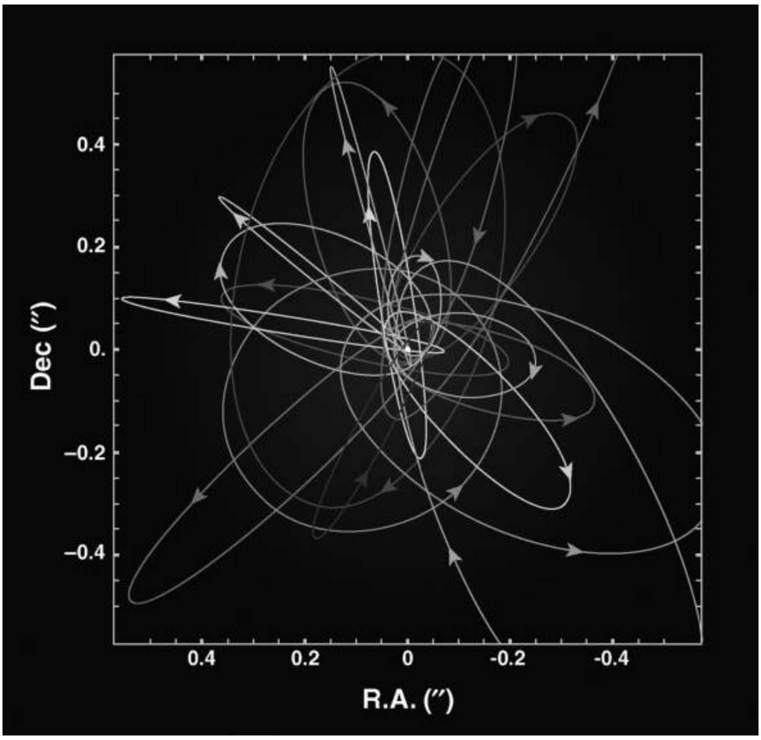

Figure 1.5 Orbits of stars around the Galactic Centre supermassive black hole Sgr A* (position given by the white dot in the centre of the figure). All of these orbits, observed with great precision over decades, require a unique black hole mass $4 \times 10^6 M_\odot$. The closest orbits now show evidence for the Einstein precession (advance of the pericentre, as predicted by general relativity). Credit: S. Gillessen.

observations by groups in Germany and the United States have followed the proper motions of stars around it in exquisite detail (see Figure 1.5) for more than 25 years.[4] Interpreting these motions as Kepler orbits shows that the moving stars orbit a central mass of order $4.5 \times 10^6 M_\odot$. Constraints on its size leave no room for doubt that this object (Sgr A*) must be the Galaxy's own SMBH.

This SMBH is remarkably inactive at the current epoch, but there is indirect evidence of activity in the past. In particular, two large gamma-ray-emitting lobes (the 'Fermi bubbles' – see Figures 1.6 and 1.7) are symmetrically placed each side of the Galactic plane with Sgr A* at the centre of symmetry. These are probably the result of an energetic outflow event from the SMBH about 6 Myr ago – the event was probably roughly isotropic, but the greater density of the Galactic plane means that propagation only occurs along the axis of the Galaxy.

[4] Reinhard Genzel and Andrea Ghez shared half the 2020 Nobel Prize in Physics for leading this work.