Supermassive black holes and the evolution of galaxies


Black holes—an extreme consequence of the mathematics of General Relativity—have long been suspected of being the main energy source for quasars, which emit more energy than any other objects in the Universe. Recent evidence indicates that supermassive black holes reside at the centres of most galaxies, where they are probably the remnants of quasars that have been starved of fuel. As our knowledge of the demographics of supermassive black holes improves, we see clues that they participated in the formation of galaxies, and strongly influenced the evolution towards the present-day structure of their surrounding hosts.

Black holes are a prediction of Einstein’s theory of gravity, fore-shadowed by the work of Michell and later Laplace in the late eighteenth century. K. Schwarzschild discovered the simplest kind of black hole in the first solution of Einstein’s equations of General Relativity, and Oppenheimer was among the first to consider the possibility that black holes might actually form in nature. The subject gained life in the 1960s and 1970s, when supermassive black holes were implicated as the powerhouses for quasars, and stellar-mass black holes were touted as the engines for many galactic X-ray sources. In the past decade, we have progressed from seeking supermassive black holes in only the most energetic astrophysical contexts, to suspecting that they might be routinely present at the centres of galaxies.

The defining property of a black hole is its event horizon, the boundary of the region surrounding the black hole from which no matter or photons can escape. Because the horizon itself is invisible, we must often settle for evidence of mass without light. All dynamical techniques for finding supermassive black holes at the centres of galaxies rely on a determination of mass enclosed within a radius $r$ from the velocity $v$ of test particles. In newtonian physics, this mass is $M = av^2/rG$. Determining $a$ requires a detailed dynamical analysis, but it is often of order 1. In cases where there is extra mass above that associated with starlight, we refer to the object as a ‘massive dark object’ (MDO). In most of the cases discussed in this paper, it is likely that the MDO is a supermassive black hole (MBH), but in only a few cases have plausible alternatives to a black hole been ruled out. These are important because they establish the reality of MBHs and justify the interpretation of less compelling objects as MBHs.

Black holes as energy sources of quasars

Black holes are thought to exist in two mass ranges. Small ones of $\sim 10^{M_\odot}$ are the evolutionary end points of some massive stars. This paper discusses the much more massive black holes that might power quasars and their weaker kin, active galactic nuclei (AGN). Quasars produce luminosities of $L \sim 10^{46}$ erg s$^{-1}$ ($\sim 10^{11}L_\odot$). Where they power double-lobed radio sources, the minimum energy stored in the lobes is $E \sim 10^{50}$–$10^{47}$ erg. The mass equivalent of this energy is $M = E/c^2 \sim 10^{10}$–$10^{5} M_\odot$, and the horizon scale associated with that mass is $R_h = GM/c^2 \sim 10^{15}$–$10^{13}$ cm. Although most quasars do not vary much at visual wavelengths, a few objects change their luminosity in minutes at high energies. Because an object cannot causally vary faster than the light-travel time $\tau$ across it, such objects must be smaller than $R \sim c\tau \sim 10^{13}$ cm. Although relativistic corrections can alter this limit somewhat in either direction via Doppler boosting or gravitational redshift, there is no escaping the conclusion that many quasars are prodigiously luminous yet tiny, outshining a galaxy in a volume smaller than the Solar System.

The small size, together with the enormous energy output of quasars, mandates black-hole accretion as the energy source. Most investigators believe that quasars and AGNs are MBHs accreting mass from their environment, nearly always at the centre of a galaxy$^{9,10}$. Black holes of mass $\sim 10^7 M_\odot$ must normally lie at the centre because dynamical friction drags them to the bottom of the potential well. This location is now clearly established for low-redshift ($z \leq 0.3$) quasars. The connection between MBHs and quasars was first made by Zeldovich$^{9,10}$ and Salpeter$^{11}$. Lynden-Bell sharpened the argument by computing the ratio of gravitational energy to nuclear energy:

$$\frac{E}{E_n} = \frac{\left(\frac{\dot{E}}{\epsilon M c^2}\right) R_h}{\left(\frac{\dot{E}_n}{\epsilon_n M c^2}\right)} \sim 100 \epsilon$$

where $R_h$ is the Schwarzschild radius of a black hole of mass $M$, $R$ is the size of the quasar, and $\epsilon$ and $\epsilon_n$ are gravitational and nuclear energy conversion efficiencies; the last equality follows from the typical astrophysical thermonuclear efficiency of $\sim 1\%$ and the size scale from variability noted above.

Because quasars were populous in the youthful Universe, but have mostly died out, the Universe should be populated with relic black holes whose average mass density $\rho_\odot$ matches or exceeds the mass-equivalent of the energy density $u$ emitted by them$^{12}$. The integrated co-moving energy density in quasar light (as emitted) is:

$$u = \int_0^\infty L(\nu) \Phi(\nu) \nu d\nu = 1.3 \times 10^{-15} \text{ erg cm}^{-3}$$

where $\Phi$ is the co-moving density of quasars of luminosity $L$, and $t$ is cosmic time. The corresponding present-day mass density for a radiative efficiency $\epsilon$ is $\rho_\odot = u/(\epsilon c^2) = 2 \times 10^9(0.1/\epsilon) M_\odot$ Mpc$^{-3}$. This density can be compared with the luminous density in galaxies, $\rho_j = 1.1 \times 10^5 L_\odot$ Mpc$^{-3}$ (ref. 14), to obtain the ratio of the mass in relic MBHs to the light of galaxies:

$$Y = \frac{\rho_\odot}{\rho_j} = 1.8 \times 10^{-2} \left(\frac{0.1}{\epsilon}\right) \left(\frac{M_j}{L_\odot}\right).$$

Dynamical evidence for massive black holes

First steps. The first dynamical evidence for black holes in galactic centres was the measurement of a rising central velocity dispersion, reaching $\sim 400$ km s$^{-1}$, in the giant elliptical galaxy M87. This object is a prime site to prospect for an MBH by virtue of its AGN features—non-thermal radio emission, broad nuclear emission lines, and a ‘jet’ of collimated relativistic particles being ejected from the nucleus. Isotropic models of the stellar kinematics, when
combined with photometry, implied an MDO of $5 \times 10^3 M_\odot$. The result was criticized because the data were also matched by a model with radially anisotropic stellar orbits and no black hole. Thus, the importance of understanding the stellar orbital structure of the centres of galaxies was obvious at the very beginning, and this subject has developed in parallel with the search for MBHs.

More convincing evidence was found in the 1980s for MDOs in M31 and M32 (the Andromeda galaxy and its satellite), which are nearby and hence observable at high spatial resolution. Rapid rotation near their centres reduces the danger of confusing a central mass with radial orbits, and Schwarzchild's method of locating orbits in a galaxy potential was used to eliminate models contrary to physics. Modern methods use Schwarzchild's method to fit the entire line-of-sight velocity distribution for axisymmetric models.

The stellar velocity work on M87 was largely vindicated two decades later by the Hubble Space Telescope (HST), which revealed a small gas disk at the centre. The gas is plausibly in circular motion, so the MDO mass estimate is straightforward. These and later data provide strong evidence for an MDO of $3 \times 10^6 M_\odot$ (refs 20, 22), a value similar to but slightly smaller than the one derived in ref. 23.

Two remarkable examples. The work described above revealed strong examples of MDOs but no iron-clad evidence for MBHs. This gap has now been partly closed in two remarkable objects. The mild AGN NGC 4258 was shown to possess a tiny annular gaseous disk near the nucleus, populated by water masers whose Doppler velocities can be observed with exquisite precision. The rotation curve is keplerian to high accuracy over the annulus width (0.13–0.26 pc). The very small velocity residuals of $\lesssim 1\%$ inspire confidence in the derived mass of $3.6 \times 10^6 M_\odot$ (ref. 25). The extraordinary high implied density of $>10^5 M_\odot$ pc$^{-3}$ (10$^8$ if one takes the limits on departures from keplerian motion as a constraint on the concentration of the mass) permits the use of astrophysical arguments to rule out most other explanations for the dark mass (see below). This is a firm link from MDOs to MBHs.

The centre of our Galaxy holds the second confirmed MBH. Near-infrared observations detect proper motions of stars in orbit about the galactic centre and indicate a rising stellar velocity dispersion down to distances of 0.01 pc. For the first time stars are being observed to orbit an MDO, year by year, with impressive accuracy that will steadily improve. The density of $>10^{12} M_\odot$ pc$^{-3}$ within the resolved region is again extraordinarily high, ruling out most alternatives to an MBH.

In the Galaxy and NGC 4258, the MDOs are almost surely MBHs rather than clusters of smaller masses. A cluster of radius $r$ and total mass $M$ of $N$ self-gravitating point masses will collapse or evaporate on a timescale of a few hundred two-body relaxation times, $t_r = 0.14N(r^2/GM)^{1/2}\ln(0.4N)$ $^{-1}$ (some important caveats were noted in ref. 28). At a chosen non-relativistic density, the lifetime of the cluster can be made longer than the age of the Universe by making the point masses sufficiently light (and therefore numerous).

In NGC 4258 and the Galaxy, this constraint requires constituents with masses $\lesssim 0.1 M_\odot$. Brown dwarfs and white dwarfs with this mass have large radii and would rapidly collide and merge. Thus the remaining candidate components for cluster models are low-mass black holes or non-interacting elementary particles. However, there is no known way to make $0.1 M_\odot$ black holes, and, because non-interacting elementary particles do not radiate energy efficiently, they do not settle into a compact configuration. The cases for black holes in both NGC 4258 and the Galaxy therefore seem very strong.

Potential new tools. The techniques above are difficult to apply to galaxies containing AGNs. The bright nucleus renders the stellar absorption lines close to the centre nearly invisible, and in many cases the nebular emission lines are influenced by non-gravitational forces. Nevertheless, in AGNs that emit broad emission lines originating from gas near the central engine ($\Delta v \approx 10,000$ km s$^{-1}$), one can attempt to estimate a mass from the average velocity of the gas and the radius of the emitting region. The velocity comes from the widths of the lines, but the radius is harder to measure. It can be estimated either from photoionization models of the gas or by ‘reverberation mapping’. In the latter method, the radius is inferred from the time interval (due to light travel) between fluctuations in the continuum radiation and the changes that these induce in the emission lines.

Masses obtained from reverberation mapping for nearby Seyfert nuclei range from $10^5$ to $10^8 M_\odot$ (ref. 30), roughly consistent with (but somewhat smaller than) the dynamical results of the previous subsection. However, the identifications of line width with orbital velocity and time delay with radius are problematic, given the absence of any correlation between line width and radius of the form $v \propto r^{-1/2}$ within the same object. Evidently some essential component of the model is still missing. A proper understanding of this technique would give us a powerful tool for working with more luminous and distant objects.

Recent advances in X-ray astronomy have furnished dramatic new evidence for MBHs in AGNs. It had been known for some time that the X-ray spectra of many AGNs show an iron K$_{\alpha}$ emission line at a rest energy of 6.4 eV, thought to arise from X-ray fluorescence of cold, neutral material in an accretion disk. Until recently, the

![Figure 1](https://example.com/figure1.png)

**Figure 1** Mass estimates of the candidate MBHs in galaxies with dynamical information plotted against the bulge luminosity of their host galaxy. The labelled points are the results of painstaking observation and detailed modelling. The symbols indicate how $M_\bullet$ was derived: triangles, kinematics of gas; filled circles, dynamics of stars; diamonds, masers; small squares, two-integral modelling using ground-based stellar kinematics. Arrows indicate upper limits on $M_\bullet$. The solid line is a model with $M_\bullet = 0.005 M_{\text{bulge}}$ and $M_{\text{bulge}} = 5L_X(0.1 L_\odot)^{-1/2}$. The distribution of $M_\bullet$ is roughly gaussian in $\log(M_\bullet/M_{\text{bulge}})$ with mean $-2.27$ ($M_{\text{bulge}}$=0.005) and standard deviation 0.5. The broken line is the quasar light prediction of equation (3) apportioned according to the bulge mass: $M_\bullet = 2 \times 10^7 L_\odot$, $6 \times 10^7 (L_\odot)^{1/2}$. The small offset from the observed black-hole/bulge-mass relation indicates that the present integrated density in MBHs is broadly consistent with the integrated luminosity produced by AGNs over the lifetime of the Universe. This offset might reflect a radiative efficiency of average quasar accretion less than 0.10.
available spectral resolution was insufficient to test the predicted line profile, but ASCA (Advanced Satellite for Cosmology and Astrophysics) provided the much-anticipated breakthrough in the Seyfert 1 galaxy MCG-6-30-15 (ref. 32). The Kα line exhibits relativistic Doppler motions of nearly 100,000 km s⁻¹, as well as an asymmetric red wing consistent with gravitational redshift. The best-fitting disk has an inner radius of only a few Schwarzschild radii. The Fe Kα line profile has now been seen in many objects33,34, and data of better quality might eventually allow measurement of the spin of the black hole34,35.

### Demographics of supermassive central black holes

We now turn to the question of the number of MBHs in the Universe. Figure 1 illustrates the relationship between black-hole mass and host spheroid luminosity from the data in Table 1 (labelled points). The labelled MDOs seem to correlate with spheroid luminosity (solid line); the upper limits are also consistent with this relation30,35. However, the number of points is small, and further progress requires more objects. Few objects at present have been studied with the detailed spatial resolution and/or modelling of the labelled points.

At the risk of greater uncertainty, more galaxies can be included by combining ground-based stellar kinematics with HST central light profiles. A simple modelling recipe based on two-integral axisymmetric models has been used on such data for a further 25 E and S0 galaxies37. This procedure assumes that the phase-space density is only a function of the energy and one component of angular momentum. MDO masses from this technique can be checked against galaxies with HST spectroscopy. The results for five low-mass galaxies with steep inner light profiles show good agreement, but the two-integral method might overestimate masses severalfold for massive ellipticals such as M87, which rotate slowly and have shallow central light profiles. A bayesian analysis of this sample indicates that MDOs are in fact very common features of normal bright galaxy centres37.

All known MDOs with measured masses so far are in galaxies with identifiable spheroidal components, suggesting that black-hole formation is exclusively linked to spheroid formation. However, several well-imaged low-z quasars do not seem to be associated with spheroids3.

It is clearly important to survey more late-type spirals without bulges. So far, we have little dynamical evidence on the centres of such galaxies, but AGN activity might perhaps be used as a proxy for a black hole. Seyferts are generally not found in late-type spirals38, but a single dwarf Seyfert nucleus (out of hundreds surveyed) has been discovered in NGC 4395, a nearby bulgeless Sd IV galaxy. The bolometric Seyfert luminosity of this nucleus is 1.4 × 10⁶⁸ erg s⁻¹ (ref. 39), and the Eddington black-hole mass is only 110M☉, small enough to have been produced by stellar evolution.

There is no evidence for MDOs in low-surface-brightness galaxies, although existing studies do not set compelling limits40. Our view of these observational results, largely developed over the past decade, is as follows: (1) MBHs are a normal feature of the central regions of bright galaxies, particularly those with spheroids; (2) their masses scale in rough proportion to host-galaxy spheroid mass; and (3) the total mass density in black holes is broadly consistent with the mass-equivalent energy density in the quasar light background. We therefore believe that the black-hole fossils of the quasar era have been found.

### Co-evolution of galaxies and black holes

#### The era of quasars

The improving statistics on local MBHs can be compared with their properties and distribution during the quasar era at z ~ 3. Were today’s MBHs already fully formed by that time, or was the average MBH smaller then, having grown by later accretion or mergers to form the present population? The evidence is not conclusive but seems to favour some growth.

The epoch of maximal activity in the Universe peaked at the same time as, or slightly before, the epoch of maximal star formation, and MBHs must have formed before this time to be available to power quasars. Figure 2 illustrates this point by plotting the history of the rate of observed star formation in the Universe together with the density of luminous quasars (those with L > 3 × 10⁶⁸ erg s⁻¹ (ref. 44). The rise in starbirth is tracked closely by the rise of luminous quasars. However, the bright quasars reach their peak at z ≳ 2 (t ≲ 1.6 × 10⁹ yr), and then proceed to die off nearly 10⁹ yr before the peak in star formation, which occurs at z ~ 1.2 (t = 2.6 × 10⁹ yr) (ref. 45). The application of extinction corrections to the cosmic star formation rate before z ~ 2 is a controversial subject that might evolve rapidly as better infrared data become available.

This chronology favours models in which the black hole forms before, or in close association with, the densest parts of galaxies (see, for example, ref. 46), as opposed to models in which the galaxy

---

**Table 1 Dynamically identified MDOs**

<table>
<thead>
<tr>
<th>Galaxy Name</th>
<th>Type</th>
<th>Distance (Mpc)</th>
<th>M_8 (Bulge)</th>
<th>M_8/M_0</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>NGC0221 M32</td>
<td>E2</td>
<td>0.0085</td>
<td>-17.0</td>
<td>2.8 × 10⁸</td>
<td>See 26, 27</td>
</tr>
<tr>
<td>NGC0223 M31</td>
<td>Sb</td>
<td>0.7</td>
<td>-15.6</td>
<td>3.4 × 10⁸</td>
<td>16, 73, 74, 75</td>
</tr>
<tr>
<td>NGC3115</td>
<td></td>
<td>0.7</td>
<td>-18.8</td>
<td>3.0 × 10⁸</td>
<td>16, 17</td>
</tr>
<tr>
<td>NGC4258 M106</td>
<td>Sbc</td>
<td>7.5</td>
<td></td>
<td>5.7 × 10⁶</td>
<td></td>
</tr>
<tr>
<td>NGC7052 E4</td>
<td></td>
<td>58.7</td>
<td>-20.9</td>
<td>1.0 × 10⁹</td>
<td></td>
</tr>
<tr>
<td>NGC4374 MB4</td>
<td>E1</td>
<td>15</td>
<td>-20.96</td>
<td>1.4 × 10⁸</td>
<td>83</td>
</tr>
<tr>
<td>NGC4449 MB7</td>
<td>E0</td>
<td>15.3</td>
<td>-21.4</td>
<td>3.3 × 10⁸</td>
<td>20, 21</td>
</tr>
<tr>
<td>NGC4584</td>
<td></td>
<td>27.4</td>
<td>-20.8</td>
<td>4.5 × 10⁸</td>
<td>84</td>
</tr>
<tr>
<td>NGC7052 E4</td>
<td></td>
<td>58.7</td>
<td>-21.31</td>
<td>3.3 × 10⁹</td>
<td>85</td>
</tr>
<tr>
<td>NGC1068 M77</td>
<td>Sb</td>
<td>7.5</td>
<td>-17.28</td>
<td>4.2 × 10⁸</td>
<td>24</td>
</tr>
<tr>
<td>NGC4258 M106</td>
<td>Sbc</td>
<td>7.5</td>
<td>-17.28</td>
<td>4.2 × 10⁸</td>
<td>24</td>
</tr>
<tr>
<td>NGC3115</td>
<td></td>
<td>0.7</td>
<td>-15.15</td>
<td>1.4 × 10⁸</td>
<td>87</td>
</tr>
<tr>
<td>NGC0205 Sph</td>
<td></td>
<td>0.72</td>
<td>-9.02</td>
<td>&lt; 9 × 10⁸</td>
<td>88</td>
</tr>
<tr>
<td>NGC0598 M33</td>
<td>Scd</td>
<td>0.795</td>
<td>-10.21</td>
<td>&lt; 5 × 10⁸</td>
<td>89</td>
</tr>
<tr>
<td>NGC4395 Sm</td>
<td></td>
<td>0.772</td>
<td>-7.27</td>
<td>&lt; 8 × 10⁸</td>
<td>30</td>
</tr>
</tbody>
</table>

* All quantities in this paper are computed for a Friedman–Robertson–Walker universe with Ω = 1 and H₀ = 80 km s⁻¹ Mpc⁻¹. Distances to nearby MBHs come from many sources, but are always rescaled to this Hubble constant.
forms first and later spawns the quasar. For this reason we associate the birth of quasars with spheroid formation, a process also identified with dense regions that collapse early\(^4\). The collapse of these regions would pre-date the average rise of star formation illustrated in Fig. 2.

The decline of bright quasars at \(z < 2\) seems likely to reflect a loss of fuel. Galaxy mergers, an effective gas transport process, are less frequent as time passes and involve a lower mean density and slower dynamical clock. There is also less gas in galaxies overall (especially protogalaxies\(^6\)) and those of fossil MBHs today. To make this comparison, we must identify a quasar of specified luminosity with its fossil MBH mass. We do this under the assumption that the brightest quasars are Eddington-limited; that is, that their luminosity is so great that radiation pressure on nearby electrons balances the gravitational force on associated protons. In this situation, the ‘Eddington luminosity’ is

\[
L_e = 1.3 \times 10^{36} M_8 \text{ erg s}^{-1}
\]

where \(M_8\) is the mass in units of \(10^8 M_\odot\).

The co-moving density of luminous quasars and galaxies at \(z \approx 2\) and those of fossil MBHs today. To make this comparison, we must identify a quasar of specified luminosity with its fossil MBH mass. We do this under the assumption that the brightest quasars are Eddington-limited; that is, that their luminosity is so great that radiation pressure on nearby electrons balances the gravitational force on associated protons. In this situation, the ‘Eddington luminosity’ is

\[
L_e = 1.3 \times 10^{36} M_8 \text{ erg s}^{-1}
\]

where \(M_8\) is the mass in units of \(10^8 M_\odot\).

The co-moving density of luminous quasars with \(L \geq 6 \times 10^{46} \text{ erg s}^{-1}\) reached its peak value of \(10^{-5} \text{ Mpc}^{-3}\) at \(z \approx 3\) (ref. 44). We assume for the moment that the MBHs underlying these bright quasars have not grown substantially since then. To estimate the corresponding mass range, we first correct the absolute magnitude limit in the quasar surveys \((M_B < -26.0 \text{ (refs 44, 50)})\) downward to our \(H_0\) and upward by a bolometric correction factor of 10 (ref. 51) to get a lower-limit luminosity of \(6 \times 10^{46} \text{ erg s}^{-1}\). This luminosity implies an MBH mass of \(M_8^{-1} = 4 \times 10^{9} M_\odot\), corresponding to a bulge luminosity of current hosts (from Fig. 2) of \(10^{49} L_\odot\). Correcting for the total-to-bulge luminosity ratio of 3 (ref. 52), and using the luminosity function of bright galaxies\(^8\), we identify the bright quasars of \(z \geq 2\) with the half of modern galaxies with \(M_8 < -20.7\) that have bulges\(^3\), and find a co-moving density of such spheroids today of \(\rho_8 = 10^{-3} \text{ Mpc}^{-3}\).

Thus, luminous quasar MBHs at \(z \approx 3\) are only \(\sim 10^{-3}\) as numerous as their galaxy-host descendants today. One way to resolve this discrepancy is to assume that quasars have very short duty cycles in their bright phase, of order \(\eta \approx 10^{-3}\). Because the quasar epoch runs about \(t_q \sim 10^{9} \text{ yr}\) from \(z = 1\) to \(z = 3\) (the full-width at half-maximum of the quasar plot in Fig. 2), the lifetime in the bright phase would then be only \(t_q = \eta t_q = 10^{9} \text{ yr}\). The fractional mass change in this phase is only \(8M/M = t_q/t_e < 0.02\), where the ‘Salpeter time’ \(t_e\) (for an accreting black hole to e-fold in mass) is:

\[
t_e = M/M = 4 \times 10^8 \text{yr}
\]

where we have parametrized the radiative efficiency in terms of \(\epsilon_0 \approx 0.1\) because popular geometrically thin optically thick accretion disk models rarely exceed efficiencies of \(\epsilon_0 \approx 0.1\).

This leads to the disturbing conclusion that quasars accrete only a tiny fraction of their mass while in their bright phase. This result, which depends critically on comparing the upper end of the AGN luminosity function with the upper end of the present-day MBH mass spectrum, is in sharp contrast to the near equality of the integral quantities (see Figure 1 and for the total mass density in present-day MBHs, underpredicting the latter by a modest factor of 5 (consistent with significant accretion's occurring as an advection-dominated flow\(^6\)). Because this integral constraint is dominated by bright objects, this conflict probably reflects a misidentification of the current fossil masses of the bright quasars.

A plausible explanation is that MBHs might not have stayed constant in mass from the quasar era until now, but rather grew in mass by an average factor \(F\). This growth might have occurred because hierarchical clustering merged these MBHs with their protogalaxies. These mergers need not emit light. In that case, luminous quasars should be identified with MBHs today that are larger by the factor \(F\) and their spheroids would be brighter by nearly the same factor. The exponential cut-off in the bright end of the luminosity function makes such spheroids much rarer and closer in abundance to the space density of quasars.

Specifically, for \(F = 5\) the limiting MBH mass today becomes \(2 \times 10^8 M_\odot\), the limiting quasar mass is \(4 \times 10^{10} L_\odot\) and the new benchmark galaxy in the luminosity function has \(M_8 = -22.2\), which is 300 times rarer. The difference in space density between quasars and spheroids is reduced to a factor of 3, the duty cycle comes up to 1/3, and the lifetime of the bright phase is \(\sim 3 \times 10^9 \text{ yr}\). In this model, bright quasars now spend a few Salpeter lifetimes in the bright phase, which is more plausible.

Is growth by a factor \(F = 5\) reasonable? Data on the recently discovered Lyman-break galaxies (LBGs) indicate slightly more growth. Several authors have suggested that LBGs are the early formation phase of spheroidal components\(^5\). Small radii (1–2 kpc (refs 56, 58)) and small velocity dispersions (measured for only a handful of the brightest objects\(^6\)) indicate modest masses of \(2\)–\(10 \times 10^8 M_\odot\). If these merge to form typical spheroids of today with \(L > 10^{10} L_\odot\) and masses \(M > 10^9 M_\odot\), growth by more like a factor of 10 would be required. Models of hierarchically clustering protogalaxies\(^6\) also suggest growth by \(F = 10\).

A second item that favours more growth is the lack of detection of AGNs in LBGs. If MBH are forming everywhere in protogalaxies together with stars at the universal ratio \(M_\bullet /M_* = 0.005\) (Fig. 1)

---

**Figure 2** A comparison of the density of very bright quasars in the Universe with the density of star formation, as a function of redshift and cosmic time. The solid line and the filled circles are the co-moving number density of quasars \(n_Q\); see ref. 12) in units of \(\text{Mpc}^{-3}\), and the open circles represent estimates of the co-moving star formation rate (SFR). The SFR can be read off the left axis, and \(n_Q\) should be read off the right axis. The arrows on the two highest redshift SFR points indicate an estimate of plausible correction for extinction. The peak of quasar activity in the Universe appears to pre-date most of the star formation. The times are derived from the redshift assuming that \(\Omega = 1\) and \(H_0 = 80 \text{ km s}^{-1} \text{ Mpc}^{-1}\).
that faint, compact galaxies (for example, M32) have the highest density satellites (by tidally shredding them), which might otherwise last for a few months to a year.6,8 Plausible models predict that some close to the MBH. Calculations show matter distributions with dense centres66, and dissipation is likely to be a common feature of our understanding of many phenomena, ranging from galaxy rotation curves to the formation of structure in the Universe. In hierarchical models and often

The tests of the picture. The detection of supermassive black holes and the discovery of dark matter share a common feature. In both cases there was scepticism of dynamical mass measurements, and acceptance was preceded by decades of debate. Dark matter has since come to be an essential feature of our understanding of many phenomena, ranging from galaxy rotation curves to the formation of structure in the Universe. We now seem engaged in a similar transition in the prevailing view of the centres of galaxies.

An inevitable source of fuel for dead quasar engines is the debris from tidally disrupted stars. MBHs with masses $M_\bullet \approx 10^6 M_\odot$ disrupt main-sequence stars rather than swallowing them whole. Some of the debris from the star is ejected, but a portion remains bound to the MBH, forming an accretion disk that undergoes a "flare" lasting for a few months to a year.6,8 Plausible models predict a V-band luminosity of $\sim 10^3 L_\odot$. The event rate is controlled by how quickly stars can drift into the "loss cone" of low-angular-momentum orbits that some close to the MBH. Calculations show that faint, compact galaxies (for example, M32) have the highest disruption rates, $\sim 10^{-4}$ yr$^{-1}$. Larger, more diffuse galaxies (for example M87) have much lower rates of $\sim 10^{-6}$ yr$^{-1}$, and often have sufficiently massive black holes to consume main-sequence stars whole. It is possible that a stellar disruption by an MBH has already been witnessed spectroscopically. The nucleus of the spiral galaxy NGC 1097 exhibited an ephemeral, broad, double-peaked H$\alpha$ emission line, whose profile matched that expected from an accretion disk.

Finally, it might be possible to detect the gravitational-wave signature of merging MBHs, and thereby constrain the merger history of galaxies. In hierarchical models a typical bright galaxy has merged a few times since the quasar era. The timescales for the decay of binary black holes in different regimes indicate that, for MBHs with $M_{\bullet} \gg 10^3 M_\odot$, the binary holes will merge on a timescale short compared with the next merger time.2,2 The merger rate for galaxies above 0.01 L$^*$ might exceed 1 per year in the visible Universe. For an equal-mass binary black hole, the final orbit produces a luminosity of order $L_{\text{GW}} \sim c^2 G M^2 \approx 10^{46}$ erg s$^{-1}$ in gravitational radiation, independent of $M_{\bullet}$. These mergers are the most powerful events in the Universe, but ironically they might not produce electromagnetic radiation. The energy is emitted over a time $t \sim GM_{\bullet}^2/c^3$. The distinctive signature of a supermassive MBH merger as opposed to two stellar-mass black holes is the lower frequency and longer duration. Two $10^6 M_\odot$ black holes radiate much of their energy at a frequency $\sim 10^{-3}$ Hz. The events are too slow for the Laser Interferometry Gravitational-wave Observatory, but are easy for LISA (the Laser Interferometric Space Array proposed as a cornerstone mission for the European Space Agency). A key test of the ideas in this paper is the observation of gravitational radiation from merging black holes at the centres of merged galaxies since $z \approx 3$.

D. Richstone is at the Institute for Advanced Study, Princeton, and the University of Michigan, Department of Astronomy, Ann Arbor, Michigan 48109-1090, USA; E. A. Ajhar, G. Bouwens, R. Green and T. R. Lauer are at Kitt Peak National Observatory, 950 N. Cherry Avenue, Tucson, Arizona 85726-6732, USA; R. Bender is at Universität-Stuttgart, Ludwig-Maximilians-Universität, Schei nerstrasse 1, D-81679 München, Germany; A. Dressler and L. C. Ho are at the Carnegie Observatories, 813 Santa Barbara Street, Pasadena, California 91101-1292, USA; S. M. Faber and Karl Gebhardt are at the University of California, Lick Observatory, Santa Cruz, California 95064, USA; A. V. Filippenko is at the University of California, Department of Astronomy, Berkeley, California 94720-3411, USA; J. Kormendy is at the University of Hawai`i, Institute of Astronomy, 26800 Woodlawn Drive, Honolulu, Hawai`i 96822, USA; John Magorrian is at the University of Toronto, Canadian Institute for Theoretical Astrophysics, Toronto, Ontario, MSS 3H8, Canada; Scott Tremaine is at Princeton University, Department of Astrophysical Science, Princeton, NJ 08544-1001, USA.
