Chapter One

Supermassive Black holes

Why did the collapsed matter in the Universe end up making galaxies and not black holes? One would have naively expected a spherical collapse to end with the formation of a point mass at its center. But, as it turns out, tidal torques from neighboring objects torque the infalling material and induce non-sphericity and some spin into the final collapse. The induced angular momentum prevents the gas from reaching the center on a direct plunging orbit. After the gas cools and loses its pressure support against gravity, it instead assembles into a disk in which the centrifugal force balances gravity. The finite size of the luminous region of galaxies is then dictated by the characteristic spin acquired by galaxy halos, which typically corresponds to a rotational velocity that is $\sim 5\%$ of the virial circular velocity, with a negligible dependence on halo mass. This does not imply that no gas accumulates at the center. In fact, galactic spheroids are observed to generically harbor a central black hole, whose formation is most likely linked to a small mass fraction the galactic gas ($< 0.1\%$) which has an unusually low amount of angular momentum. The small mass fraction of the central black holes implies that their gravitational effect is restricted to the innermost cusp of their host galaxy. Nevertheless, these central black holes are known to have a strong influence on the evolution of their host galaxies. This state of affairs can be easily understood from the fact that the binding energy per unit mass in a typical galaxy correspond to velocities v of hundreds of $\mathrm{km\ s^{-1}}$ or a fraction $\sim (v/c)^2 \sim 10^{-6}$ of the binding energy per unit mass near a black hole. Hence a small amount of gas that releases its binding energy near a black hole can have a large effect on the rest of the gas in the galaxy.

We start this chapter with a short introduction to the properties of black holes in general relativity.

1.1 BLACK HOLES

Birkhoff's theorem states that the only vacuum, spherically symmetric gravitational field is the static *Schwarzschild metric*,

$$ds^2 = -\left(1 - \frac{r_{\rm Sch}}{r}\right) c^2 dt^2 + \left(1 - \frac{r_{\rm Sch}}{r}\right)^{-1} dr^2 + r^2 d\Omega, \tag{1.1}$$

where $d\Omega = (d\theta^2 + \sin^2\theta d\phi^2)$. The *Schwarzschild radius* is related to the mass M of the central (non-spinning) black hole,

$$r_{\rm Sch} = \frac{2GM}{c^2} = 2.95 \times 10^5\ {\rm cm} \left(\frac{M}{1M_\odot}\right). \tag{1.2}$$

The black hole horizon, $r_{\rm Hor}$ ($= r_{\rm Sch}$ here), is a spherical boundary from where no particle can escape. (The coordinate singularity of the Schwarzschild metric at $r = r_{\rm Sch}$ can be removed through a transformation to the *Kruskal* coordinate system $(r,t) \rightarrow (u,v)$, where $u = (r/r_{\rm Sch} - 1)^{1/2}\, e^{r/2r_{\rm Sch}} \cosh(ct/2r_{\rm Sch})$; $v = u\, \tanh(ct/2r_{\rm Sch})$.) The existence of a region in space into which particles may fall but never come out breakd time reversal symmetry that characterizes the equations of quantum mechanics. Any grander theory that would unify quantum mechanics and gravity must remedy this conceptual inconsistency.

In addition to its mass M, a black hole can only be characterized by its spin J and electric charge Q (similarly to an elementary particle). In astrophysical circumstances, any initial charge of the black hole would be quickly neutralized through the polarization of the background plasma and the preferrential infall of electrons or protons. The residual electric charge would exert an electric force on an electron that is comparable to the gravitational force on a proton, $eQ \sim GMm_p$, implying $(Q^2/GM^2) \sim Gm_p^2/e^2 \sim 10^{-36}$ and a negligible contribution of the charge to the metric. A spin, however, may modify the metric considerably.

The general solution of Einstein's equations for a spinning black hole was derived by Kerr in 1963, and can be written most conveniently in the Boyer-Lindquist coordinates,

$$\begin{aligned} ds^2 &= -\left(1 - \frac{r_{\rm Sch} r}{\Sigma}\right) c^2 dt^2 - \frac{2jr_{\rm Sch} r \sin^2\theta}{\Sigma} c dt d\phi + \frac{\Sigma}{\Delta} dr^2 \\ &+ \Sigma d\theta^2 + \left(r2 + j^2 + \frac{r_{\rm Sch} j^2 r \sin^2\theta}{\Sigma}\right) \sin^2\theta d\phi^2. \end{aligned} \tag{1.3}$$

where the black hole is rotating in the ϕ direction, $j = [J/Mc]$ is the normalized angular momentum per unit mass, $\Delta = r^2 - r r_{\rm Sch} + j^2$, and $\Sigma = r^2 + j^2 \cos^2\theta$. The dimensionless ratio $a = j/(GM/c^2)$ is bounded by unity, and $a = 1$ corresponds to a maximally rotating black hole. The horizon radius $r_{\rm Hor}$ is now located at the larger root of the equation $\Delta = 0$, namely $r_+ = \frac{1}{2} r_{\rm Sch}(1 + (1 - a^2)^{1/2}$. The Kerr metric converges to the Schwarzschild metric for $a = 0$. There is no Birkhoff's theorem for a rotating black hole.

Test particles orbits around black holes can be simply described in terms of an effective potential. For photons around a Schwarzschild black hole, the potential is simply $V_{\rm ph} = (1 - r_{\rm Sch}/r)/r^2$. This leads to circular photon orbits at a radius $r_{\rm ph} = \frac{3}{2} r_{\rm Sch}$. For a spinning black hole,

$$r_{\rm ph} = r_{\rm Sch}\left[1 + \cos\left(\frac{2}{3}\cos^{-1}[\pm a]\right)\right], \tag{1.4}$$

where the upper sign refers to orbits that rotate in the opposite direction to the black hole (retrograde orbits) and the lower sign to corotating (prograde) orbits. For a maximally-rotating black hole ($|a| = 1$), the photon orbit radius is $r_{\rm ph} = \frac{1}{2} r_{\rm Sch}$ for a prograde orbit and $2r_{\rm Sch}$ for a retrograde orbit.

Circular orbits of massive particles exist when the first derivative of their effective potential (including angular momentum) with respect to radius vanishes, and these orbits are stable if the second derivative of the potential is positive. The radius of the *Innermost Circular Stable Orbit (ISCO)* defines the inner edge of any

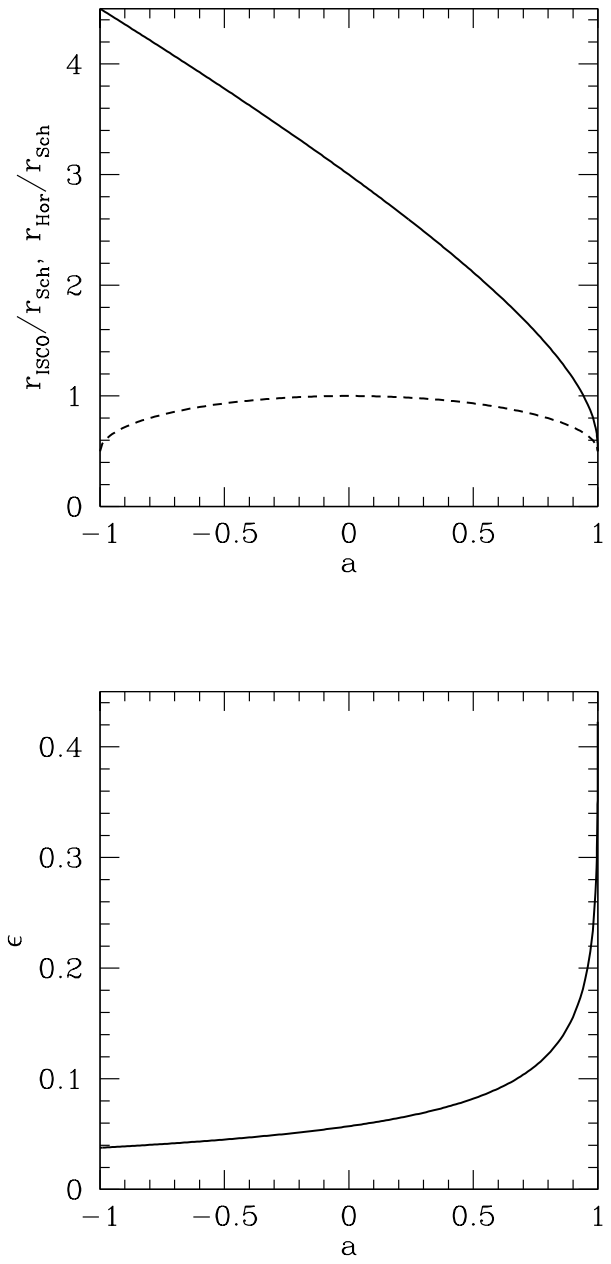

Figure 1.1 The left panel shows the radius of the black hole horizon $r_{\rm Hor}$ (dashed line) and the *Innermost Circular Stable Orbit (ISCO)* around it $r_{\rm ISCO}$ (solid line), in units of the Schwarzschild radius $r_{\rm Sch}$ (see Eq. 1.2), as functions of the black hole spin parameter a. The limiting value of $a = 1$ ($a = -1$) corresponds to a corotating (counter-rotating) orbit around a maximally-spinning black hole. The binding energy of a test particle at the ISCO determines the radiative efficiency ϵ of a thin accretion disk around the black hole, shown on the right panel.

disk of particles in circular motion (such as fluid elements in an accretion disk). At smaller radii, gravitationally bound particles plunge into the black hole on a dynamical time. This radius of the ISCO is given by[1],

$$r_{\rm ISCO} = \frac{1}{2} r_{\rm Sch} \left\{ 3 + Z_2 \pm [(3 - Z_1)(3 + Z_1 + 2Z_2)]^{1/2} \right\}, \qquad (1.5)$$

where $Z_1 = 1 + (1 - a^2)^{1/3}[(1 + a)^{1/3} + (1 - a)^{1/3}]$ and $Z_2 = (3a^2 + Z_1^2)^{1/2}$. Figure 1.1 shows the radius of the ISCO as a function of spin. The binding energy of particles at the ISCO define their maximum radiative efficiency because they spend a short time on their plunging orbit interior to the ISCO. This efficiency is given by,

$$\epsilon = 1 - \frac{r^2 - r_{\rm Sch} r \mp j\sqrt{\frac{1}{2} r_{\rm Sch} r}}{r(r^2 - \frac{3}{2} r_{\rm Sch} r \mp 2j\sqrt{\frac{1}{2} r_{\rm Sch} r})^{1/2}}. \qquad (1.6)$$

The efficiency changes between a value of $\epsilon = (1 - \sqrt{8/9}) = 5.72\%$ for $a = 0$, to $(1 - \sqrt{1/3}) = 42.3\%$ for a prograde (corotating) orbit with $a = 1$ and $(1 - \sqrt{25/27}) = 3.77\%$ for a retrograde orbit.

1.2 ACCRETION OF GAS ONTO BLACK HOLES

1.2.1 Bondi Accretion

Consider a black hole embedded in a hydrogen plasma of uniform density $\rho_0 = m_p n_0$ and temperature T_0. The thermal protons in the gas are moving around at roughly the sound speed $c_s \sim kT/m_p$. The black hole gravity could drive accretion of gas particles that are gravitationally bound to it, namely interior to the radius of influence, $r_{inf} \sim GM/c_s^2$. The steady mass flux of particles entering this radius is $\rho_0 c_s$. Multiplying this flux by the surface area associated with the radius of influence gives the supply rate of fresh gas,

$$\dot{M} \approx 4\pi r_{\rm inf}^2 \rho_0 c_s = 60 \left(\frac{M}{10^8 M_\odot} \right)^2 \left(\frac{n_0}{1\ {\rm cm}^{-3}} \right) \left(\frac{T_0}{10^4\ {\rm K}} \right)^{-3/2} M_\odot\ {\rm yr}^{-1}. \qquad (1.7)$$

In a steady state this supply rate equals the mass accretion rate into the black hole.

The explicit steady state solution to the conservations equations of the gas (mass, momentum, and energy) was first derived by Bondi (1952). The exact solution introduces a correction factor of order unity to equation (1.7). The solution is self-similar. Well inside the sonic radius the velocity is close to free-fall $u \sim (2GM/r)^{1/2}$ and the gas density is $\rho \sim \rho_0 (r/r_{\rm inf})^{-3/2}$. The radiative efficiency is small, because either the gas is tenuous so that its cooling time is longer than its accretion (free-fall) time or the gas is dense and the diffusion time of the radiation outwards is much longer than the free-fall time. If the inflowing gas contains near-equipartition magnetic fields, then cooling through synchrotron emission typically dominates over free-free cooling.

A black hole that is moving with a velocity V relative to a uniform medium accretes at a lower rate than a stationary black hole. At high velocities, the radius

of influence of the black hole would be now $\sim GM/V^2$, suggesting that the sound speed c_s be crudely replaced with $\sim (c_s^2 + V^2)^{1/2}$ in equation (1.7).

1.2.2 Thin Disk Accretion

If the inflow is endowed with rotation, the gas would reach a centrifugal barrier from where it could only accrete farther inwards after its angular momentum has been transported away. This limitation follows from the steeper radial scaling of the centrifugal acceleration ($\propto r^{-3}$) compared to the gravitational acceleration ($\propto r^{-2}$). Near the centrifugal barrier, where the gas is held against gravity by rotation, an accretion disk would form around the black hole, centered on the plane perpendicular to the rotation axis. The accretion time would then be dictated by the rate at which angular momentum is transported through viscous stress, and could be significantly longer than the free-fall time for a non-rotating flow (such as described by the Bondi accretion model). As the gas settles to a disk, the dissipation of it kinetic energy in heat would make the disk thick and hot, with a proton temperature close to the gravitational potential energy per proton $\sim 10^12\ \mathrm{K}(r/r_{\rm Sch})^{-1}$. However, if the cooling time of the gas is shorter than the viscous time, then a thin disk would form. This is realized for the high gas infall rate during the processes (such as galaxy mergers) that feed quasars. We start by exploring the structure of thin disks that characterize the high accretion rate of quasars.

Following Shakura & Sunyaev (1972), we imagine a planar thin disk of cold gas orbiting a central black hole and wish to describe its structure in polar coordinates (r, ϕ). Each gas element orbits at the local Keplerian velocity $v_\phi = r\Omega = (GM/r)^{1/2}$ and spirals slowly inwards with $v_r \ll v_\phi$ as viscous torques transport its angular momentum to the outer part of the disk. The associated viscous stress generates heat, which is radiated away locally from the the disk surface. We assume that the disk is fed steadily and so it manifests a constant mass accretion rate at all radii. Mass conservation implies,

$$\dot{M} = 2\pi r \Sigma v_r = const, \tag{1.8}$$

where $\Sigma(r)$ is the surface mass density of the disk and v_r is the radial (accretion) velocity of the gas.

In the limit of geometrically thin disk with a scale height $h \ll r$, the hydrodynamic equations decouple in the radial and vertical directions. We start with the radial direction. The Keplerian velocity profile introduces shear which dissipates heat as neighboring fluid elements rub against each other. The concept of shear viscosity can be can be easily understood in the one dimensional example of a uniform gas whose velocity along the y-axis varies linearly with the x coordinate, $V = V_0 + (dV_y/dx)x$. A gas particle moving at the typical thermal speed v traverses a mean-free-path λ along the x-axis before it collides with other particles and shares its y-momentum with them. The y-velocity is different across a distance λ by an amount $\Delta V \sim \lambda dV_y/dx$. Since the flux of particles streaming along the x-axis is $\sim nv$, where n is the gas density, the net flux of y-momentum being transported per unit time, $\sim nvm\Delta V$, is linear in the velocity gradient $\eta dV_y/dx$, with a viscosity coefficient $\eta \sim \rho v\lambda$, where $\rho = mn$ is the mass density of the gas.

Within the accretion disk, the flux ϕ-momentum which is transported in the positive r-direction is given by the viscous stress $f_\phi = \frac{3}{2}\eta\Omega$, where η is the viscosity coefficient (in $\mathrm{g\ cm^{-1}\ s^{-1}}$). The viscous stress is expected to be effective down to the ISCO, from where the gas plunges into the black hole on a free fall time. We therefore set the inner boundary of the disk as $r_{\rm ISCO}$, depicted in Fig. 1.1. Angular momentum conservation requires that the net rate of change within a radius r be equal to the viscous torque, namely

$$f_\phi \times (2\pi r \times 2h) \times r = \dot{M}\left[(GMr)^{1/2} - (GMr_{\rm ISCO})^{1/2}\right]. \tag{1.9}$$

The production rate of heat by the viscous stress is given by $\dot{Q} = f_\phi^2/\eta$. Substituting f_ϕ and equation (1.9) gives,

$$2h\dot{Q} = \frac{3\dot{M}}{4\pi r^2}\frac{GM}{r}\left[1 - \left(\frac{r_{\rm ISCO}}{r}\right)^{1/2}\right]. \tag{1.10}$$

This power gives local flux that is radiated vertically from the top and bottom surfaces of the disk,

$$F = \frac{1}{2} \times 2h\dot{Q} = \frac{3\dot{M}}{8\pi r^2}\frac{GM}{r}\left[1 - \left(\frac{r_{\rm ISCO}}{r}\right)^{1/2}\right]. \tag{1.11}$$

The total luminosity of the disk is given by,

$$L = \int_{r_{\rm ISCO}}^{\infty} 2F \times 2\pi r dr = \frac{1}{2}\frac{GM\dot{M}}{r_{\rm ISCO}}, \tag{1.12}$$

where we have ignored genral-relativistic corrections to the dynamics of the gas and the propagation of the radiation it emits.

In the absence of any vertical motion, the momentum balance in the vertical z-direction yields,

$$\frac{1}{\rho}\frac{dP}{dz} = -\frac{GM}{r^2}\frac{z}{r}, \tag{1.13}$$

where $z \ll r$ and P and ρ are the gas pressure and density. This equation gives a disk scale height $h \approx c_s/\Omega$ where $c_s \approx (P/\rho)^{1/2}$ is the sound speed.

Because of the short mean-free-path for particles collisions, the particle-level viscosity is negligible in accretion disks. Instead the magneto-rotational instability (CITE) is likely to develop turbulent eddies in the disk which are much more effective at transporting its angular momentum. In this case λ and v should be replaced by the typical size and velocity of an eddy. The largest value that these variables can obtain are the scale height h and sound speed c_s in the disk. This implies $f_\phi < (\rho c_s h)\Omega \approx \rho c_s^2 \approx P$. We may then parameterize the viscous stress as some fraction α of its maximum value, $f_\phi = \alpha P$.

The total pressure P in the disk is the sum of the gas pressure $P_{\rm gas} = 2(\rho/m_p)k_B T$, and the radiation pressure, $P_{\rm rad} = \frac{1}{3}aT^4$. We define the fractional contribution of the gas as,

$$\beta \equiv \frac{P_{\rm gas}}{P}, \tag{1.14}$$

where $P = P_{\rm rad} + P_{\rm gas}$. In principle, the viscous stress may be limited by the gas pressure only; to reflect this possibility, we write $f_\phi = \alpha P \beta^b$, where b is 0 or 1 if the viscosity scales with the total or just the gas pressure, respectively.

Since the energy of each photon is just its momentum times the speed of light, the radiative energy flux is simply given by the change in the radiation pressure (momentum flux) per photon mean-free-path,

$$F = -c\frac{dP_{\rm rad}}{d\tau}, \tag{1.15}$$

where the optical-depth τ is related to the frequency-averaged (so-called, Rosseland-mean) opacity coefficient of the gas, κ,

$$\tau = \int_0^h \kappa\rho dz \approx \frac{1}{2}\kappa\Sigma, \tag{1.16}$$

where $\Sigma = 2h\rho$. For the characteristic mass density ρ and temperature T encountered at the midplane of accretion disks around supermassive black holes, there are two primary sources of opacity: *electron scattering* with

$$\kappa_{\rm es} = \frac{\sigma_{\rm T}}{m_p} = 0.4\ {\rm cm^2\,g^{-1}}, \tag{1.17}$$

and *free-free* absorption with

$$\kappa_{\rm ff} = 8\times10^{22}{\rm cm^2\,g^{-1}}\left(\frac{\rho}{\rm g\,cm^{-3}}\right)\left(\frac{T}{\rm K}\right)^{-7/2}, \tag{1.18}$$

where we assume a pure hydrogen plasma for simplicity.

It is customary to normalize the accretion rate $\dot{M}$ in the disk relative to the so-called Eddington rate $\dot{M}_E$, which would produce the maximum possible disk luminosity, $L_{\rm Edd}$ (see derivation in equation 1.33 below). When the luminosity approaches the Eddington limit, the disk bloates and h approaches r, violating the thin-disk assumption. We write $\dot{m} = \dot{M}/\dot{M}_E$, with $\dot{M}_{\rm Edd} \equiv L_{\rm Edd}/(\epsilon c^2)$, where ϵ is the radiative efficiency for converting rest-mass to radiation near the ISCO.

Based on the above equations, we are now at a position to derive the scaling laws that govern the structure of the disk far away from the ISCO. For this purpose we use the following dimensionless parameters: $r_1 = (r/10R_{\rm Sch})$, $M_8 = (M/10^8 M_\odot)$, $\dot{m}_{-1} = (\dot{m}/0.1)$, $\alpha_{-1} = (\alpha/0.1)$ and $\epsilon_{-1} = (\epsilon/0.1)$.

In local thermodynamic equilibrium, the emergent flux from the surface of the disk (equation 1.11) can be written in terms of the temperature at disk midplane T as $F \approx caT^4/\kappa\Sigma$. The surface temperature of the disk is the roughly,

$$T_s \approx \left(\frac{4F}{a}\right)^{1/4} = 10^5\ {\rm K}\ M_8^{-1/4}\dot{m}_{-1}^{1/4}r_1^{-3/4}\left[1-\left(\frac{r}{r_{\rm ISCO}}\right)^{1/2}\right]. \tag{1.19}$$

The accretion disk can be divided radially into three distinct regions,

1. *Inner region:* where radiation pressure and electron-scattering opacity dominate.

2. *Middle region:* where gas pressure and electron-scattering opacity dominate.

3. *Outer region:* where gas pressure and free-free opacity dominate.

The boundary between regions 1 and 2 is located at the radius

$$r_1 \approx 54\,\alpha_{-1}^{2/21}(\dot{m}_{-1}/\epsilon_{-1})^{16/21}M_8^{2/21} \quad \text{if b} = 1, \tag{1.20}$$

$$58\,\alpha_{-1}^{2/21}(\dot{m}_{-1}/\epsilon_{-1})^{16/21}M_8^{2/21} \quad \text{if b} = 0, \tag{1.21}$$

and the transition radius between regions 2 and 3 is

$$r_1 \approx 4 \times 10^2\,(\dot{m}_{-1}/\epsilon_{-1})^{2/3}. \tag{1.22}$$

The surface density and scale-height of the disk are given by,
Inner region:

$$\Sigma(r) \approx (3 \times 10^6\,\mathrm{g\,cm^{-2}})\alpha_{-1}^{-4/5}\left(\frac{\dot{m}_{-1}}{\epsilon_{-1}}\right)^{3/5} M_8^{1/5} r_1^{-3/5} \text{ if b} = 1, \tag{1.23}$$

$$(8 \times 10^2\,\mathrm{g\,cm^{-2}})\alpha_{-1}^{-1}\left(\frac{\dot{m}_{-1}}{\epsilon_{-1}}\right)^{-1} r_1^{3/2} \quad \text{if b} = 0, \tag{1.24}$$

$$h(r) \approx R_{\rm Sch}\left(\frac{\dot{m}_1}{\epsilon_{-1}}\right). \tag{1.25}$$

Middle region:

$$\Sigma(r) \approx (3 \times 10^6\,\mathrm{g\,cm^{-2}})\alpha_{-1}^{-4/5}\left(\frac{\dot{m}_{-1}}{\epsilon_{-1}}\right)^{3/5} M_8^{1/5} r_1^{-3/5}, \tag{1.26}$$

$$h(r) \approx 1.2 \times 10^{-2} R_S \alpha_{-1}^{-1/10}\left(\frac{\dot{m}_{-1}}{\epsilon_{-1}}\right)^{1/5} M_8^{-1/10} r_1^{21/20}. \tag{1.27}$$

Outer region:

$$\Sigma(r) \approx (6 \times 10^6\,\mathrm{g\,cm^{-2}})\alpha_{-1}^{-4/5}\left(\frac{\dot{m}_{-1}}{\epsilon_{0.1}}\right)^{7/10} M_8^{1/5} r_1^{-3/4}, \tag{1.28}$$

$$h(r) \approx 10^{-2} R_S \alpha_{-1}^{-1/10}\left(\frac{\dot{m}_{-1}}{\epsilon_{-1}}\right)^{3/20} M_8^{-1/10} r_1^{9/8}. \tag{1.29}$$

The mid-plane temperature is given by,

$$T(r) \approx \left(16\pi^2\right)^{-1/5}\left(\frac{m_p}{\mathrm{k_B}\sigma_T}\right)^{1/5} \alpha^{-1/5}\kappa^{1/5}\dot{M}^{2/5}\Omega^{3/5}\beta^{-(1/5)(b-1)}. \tag{1.30}$$

The above scaling-laws ignore the self-gravity of the disk. This assumption is violated at large radii. The instability of the disk to gravitational fragmentation due to its self-gravity occurs when the so-called Toomre parameter, $Q = (c_s\Omega/\pi G\Sigma)$, drops below unity. For the above scaling laws of the outer disk, this occurs at the outer radius,

$$r_1 \approx 2 \times 10^4 \alpha_{-1}^{28/45}\,(\dot{m}_{-1}/\epsilon_{-1})^{-22/45}\,M_8^{52/45}. \tag{1.31}$$

Outside this radius, the disk gas would fragment into stars, and the stars may migrate inwards as the gas accretes onto the black hole. The energy output from stellar winds and supernovae would supplement the viscous heating of the disk and

might regulate the disk to have $Q \sim 1$ outside the above boundary. We therefore conclude that star formation will inevitably occur on larger scales, before the gas is driven into the accretion disk that feeds the central black hole. Indeed, the broad emission lines of quasars display very high abundance of heavy elements in the spectra out to arbitrarily high redshifts. Since the total amount of mass in the disk interior to this radius makes only a small fraction of the mass of the supermassive black hole, quasars must be fed by gas that crosses this boundary after being volnurable to fragmentation.

1.2.3 Radiatively Inefficient Accretion Flows

When the accretion rate is considerably lower than its Eddington limit ($\dot{M}/\dot{M}_E < 10^{-2}$), the gas inflow switches to a different mode, called a *Radiatively Inefficient Accretion Flow* (RIAF) or an *Advection Dominated Accretion Flow* (ADAF), in which either the cooling time or the photon diffusion time are much longer than the accretion time of the gas and heat is mostly advected with the gas into the black hole. At the low gas densities and high temperatures characterizing this accretion mode, the Coulomb coupling is weak and the electrons do not heat up to the proton temperature even with the aid of plasma instabilities. Viscosity heats primarily the protons since they carry most of the momentum. The other major heat source, compression of the gas, also heats the protons more effectively than the electrons. As the gas infalls and its density ρ rises, the temperature of each species T increases adiabatically as $T \propto \rho^{\gamma-1}$, where γ is the corresponding adiabatic index. At radii $r < 10^2 r_{\rm Sch}$, the electrons are relativistic with $\gamma = 4/3$ and so their temperature rises inwards with increasing density as $T_e \propto \rho^{1/3}$ while the protons are non-relativistic with $\gamma = 5/3$ and so $T_{\rm p} \propto \rho^{2/3}$, yielding a two-temperature plasma with the protons being much hotter than the electrons. Typical models (CITE Narayan & McClintock and refs therein) yield, $T_p \sim 10^{12}$ K$(r/r_{\rm Sch})^{-1}, T_e \sim \min(T_p, 10^{9-11}$ K). Because the typical sound speed is comparable to the Keplerian speed at each radius, the geometry of the flow is thick – making RIAFs the viscous analogs of Bondi accretions.

Analytic models imply a radial velocity that is a factor of $\sim \alpha$ smaller than the free-fall speed and an accretion time that is a factor of $\sim \alpha$ longer than the free-fall time. However, since the sum of the kinetic and thermal energy of a proton is comparable to its gravitational binding energy, RIAFs are expected to be associated with strong outflows.

The radiative efficiency of RIAFs is smaller than the thin-disk value, ϵ. While the thin-disk value applies to high accretion rates above some critical value, $\dot{m} > \dot{m}_{\rm crit}$, the anayltic RIAF models typically admit a radiative efficiency of,

$$\frac{L}{\dot{M}c^2} \approx \epsilon \left(\frac{\dot{m}}{\dot{m}_{\rm crit}} \right), \tag{1.32}$$

for $\dot{m} < \dot{m}_{\rm crit}$, with $\dot{m}_{\rm crit}$ in the range of 0.01–0.1. Here $\dot{m}$ is the accretion rate (in Eddington units) near the ISCO, after taking account of the fact that some of the infalling mass at larger radii is lost to outflows. For example, in the nucleus of the Milky Way, massive stars shed $\sim 10^{-3} M_\odot$ yr^{-1} of mass into the radius of

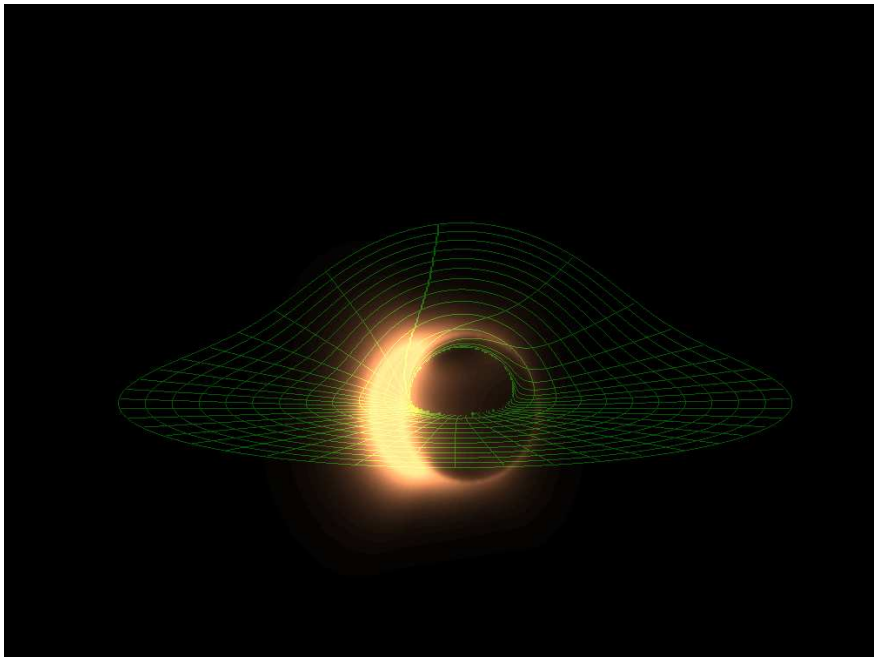

Figure 1.2 Simulated image of an accretion flow around a black hole spinning at half its maximum rate, from a viewing angle of 10° relative to the rotation axis. The coordinate grid in the equatorial plane of the spiraling flow shows how strong lensing around the black hole bends the back of the apparent disk up. The left side of the image is brighter due its rotational motion towards the observer. The bright arcs are generated by gravitational lensing. A dark silhouette appears around the location of the black hole because the light emitted by gas behind it disappears into the horizon and cannot be seen by an observer on the other side. Recently, the technology for observing such an image from the supermassive black holes at the centers of the Milky Way and M87 galaxies has been demonstrated as feasible [Doeleman, S., et al. *Nature* **455**, 78 (2008)]. To obtain the required resolution of tens of micro-arcseconds, signals are being correlated over an array (interferometer) of observatories operating at a millimeter wavelength across the Earth. Figure credit: Broderick, A., & Loeb, A. *Journal of Physics Conf. Ser.* **54**, 448 (2006); *Astrophys. J.* **697** 1164 (2009).

influence of central black hole (SgrA*), but only a tiny fraction $\sim 10^{-5}$ of this mass accretes onto the black hole.

Since at low redshifts mergers are rare and much of the gas in galaxies has already been consumed in making stars, most of the local supermassive black holes are characterized by a very low accretion rate. The resulting low luminosity of these dormant black holes, such as the $4 \times 10^6 M_\odot$ black hole lurking at the center of the Milky Way galaxy, is often described using RIAF/ADAF models.

1.3 THE FIRST BLACK HOLES AND QUASARS

A black hole is the end product from the complete gravitational collapse of a material object, such as a massive star. It is surrounded by a horizon from which even light cannot escape. Black holes have the dual virtues of being extraordinarily simple solutions to Einstein's equations of gravity (as they are characterized only by their mass, charge, and spin), but also the most disparate from their Newtonian analogs. In Einstein's theory, black holes represent the ultimate prisons: you can check in, but you can never check out.

Ironically, black hole environments are the brightest objects in the universe. Of course, it is not the black hole that is shining, but rather the surrounding gas is heated by viscously rubbing against itself and shining as it spirals into the black hole like water going down a drain, never to be seen again. The origin of the radiated energy is the release of gravitational binding energy as the gas falls into the deep gravitational potential well of the black hole. As much as tens of percent of the mass of the accreting material can be converted into heat (more than an order of magnitude beyond the maximum efficiency of nuclear fusion). Astrophysical black holes appear in two flavors: stellar-mass black holes that form when massive stars die, and the monstrous super-massive black holes that sit at the center of galaxies, reaching masses of up to 10 billion Suns. The latter type are observed as quasars and active galactic nuclei (AGN). It is by studying these accreting black holes that all of our observational knowledge of black holes has been obtained.

If this material is organized into a thin accretion disk, where the gas can efficiently radiate its released binding energy, then its theoretical modelling is straightforward. Less well understood are radiatively inefficient accretion flows, in which the inflowing gas obtains a thick geometry. It is generally unclear how gas migrates from large radii to near the horizon and how, precisely, it falls into the black hole. We presently have very poor constraints on how magnetic fields embedded and created by the accretion flow are structured, and how that structure affects the observed properties of astrophysical black holes. While it is beginning to be possible to perform computer simulations of the entire accreting region, we are decades away from true *ab initio* calculations, and thus observational input plays a crucial role in deciding between existing models and motivating new ideas.

More embarrassing is our understanding of black hole jets (see images 1.3). These extraordinary exhibitions of the power of black holes are moving at nearly the speed of light and involve narrowly collimated outflows whose base has a size comparable to the solar system, while their front reaches scales comparable to the

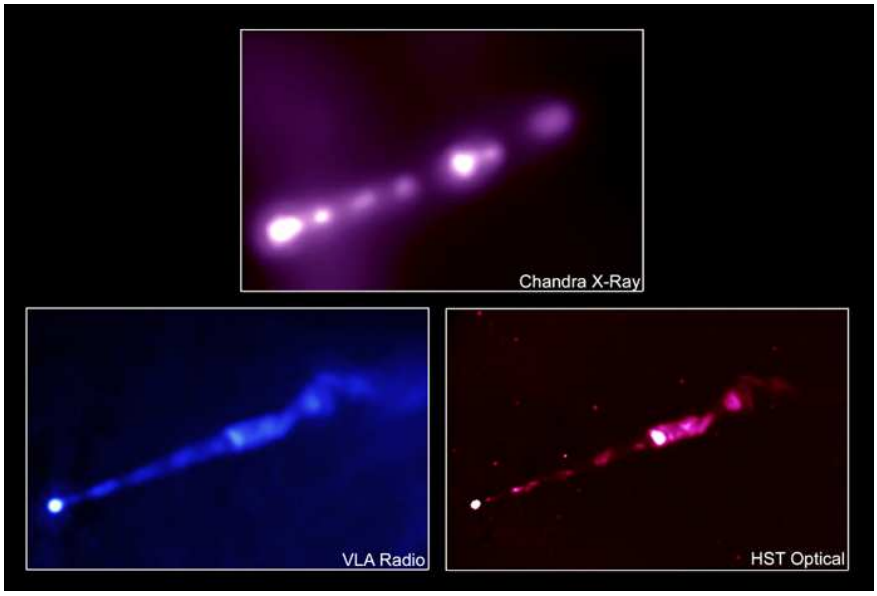

Figure 1.3 Multi-wavelength images of the highly collimated jet emanating from the supermassive black hole at the center of the giant elliptical galaxy M87. The X-ray image (top) was obtained with the Chandra X-ray satellite, the radio image (bottom left) was obtained with the Very Large Array (VLA), and the optical image (bottom right) was obtained with the Hubble Space Telescope (HST).

distance between galaxies.[2] Unresolved issues are as basic as what jets are made of (whether electrons and protons or electrons and positrons, or primarily electromagnetic fields) and how they are accelerated in the first place. Both of these rest critically on the role of the black hole spin in the jet-launching process.

A quasar is a point-like ("quasi-stellar") bright source at the center of a galaxy. There are many lines of evidence indicating that a quasar involves a supermassive black hole, weighting up to ten billion Suns, which is accreting gas from the core of its host galaxy. The supply of large quantities of fresh gas is often triggered by a merger between two galaxies. The infalling gas heats up as it spirals towards the black hole and dissipates its rotational energy through viscosity. The gas is expected to be drifting inwards in an accretion disk whose inner "drain" has the radius of the ISCO, according to Einstein's theory of gravity. Interior to the ISCO, the gas plunges into the black hole in such a short time that it has no opportunity to radiate most of its thermal energy. However, as mentioned in §1.1 the fraction of the rest mass of the gas which gets radiated away just outside the ISCO is high, ranging between 5.7% for a non-spinning black hole to 42.3% for a maximally-spinning black hole (see Fig. 1.1). This "radiative efficiency" is far greater than the mass-energy conversion efficiency provided by nuclear fusion in stars, which is $< 0.7\%$.

Quasar activity is observed in a small fraction of all galaxies at any cosmic epoch.

Mammoth black holes weighing more than a billion solar masses were discovered at redshifts as high as $z \sim 6.5$, less than a billion years after the Big Bang. *If massive black holes grow at early cosmic times, should their remnants be around us today?* Indeed, searches for black holes in local galaxies have found that every galaxy with a stellar spheroid harbors a supermassive black hole at its center. This implies that quasars are rare simply because their activity is short-lived. Moreover, there appears to be a tight correlation between the black hole mass and the gravitational potential-well depth of their host spheroids of stars (as measured by the velocity dispersion of these stars). This suggests that the black holes grow up to the point where the heat they deposit into their environment or the piston effect from their winds prevent additional gas from feeding them further. The situation is similar to a baby who gets more energetic as he eats more at the dinner table, until his hyper-activity is so intense that he pushes the food off the table and cannot eat any more. This *principle of self-regulation* explains why quasars are short lived and why the final black hole mass is dictated by the depth of the potential in which the gas feeding it resides.[3] Most black holes today are dormant or "starved" because the gas around them was mostly used up in making the stars, or because their activity heated or pushed it away a long time ago.

What seeded the formation of supermassive black holes only a billion years after the Big Bang? We know how to make a black hole out of a massive star. When the star ends its life, it stops producing sufficient energy to hold itself against its own gravity, and its core collapses to make a black hole. Long before evidence for black holes was observed, this process leading to their existence was understood theoretically by Robert Oppenheimer and Hartland Snyder in 1937. However, growing a supermassive black hole is more difficult. There is a maximum luminosity at which the environment of a black hole of mass $M_{\rm BH}$ may shine and still accrete gas.[i] This Eddington luminosity, L_E, is obtained from balancing the inward force of gravity on each proton by the outward radiation force on its companion electron at a distance r:

$$\frac{GM_{\rm BH}m_p}{r^2} = \frac{4\pi L_E}{r^2 c}\sigma_T, \tag{1.33}$$

where m_p is the proton mass and $\sigma_T = 0.67 \times 10^{-24}\ {\rm cm}^2$ is the cross-section for scattering a photon by an electron. Interestingly, the limiting luminosity is independent of radius in the Newtonian regime. Since the Eddington luminosity represents an exact balance between gravity and radiation forces, it actually equals to the luminosity of massive stars which are held at rest against gravity by radiation pressure, as described by equation (1.34). This limit is formally valid in a spherical geometry, and exceptions to it were conjectured for other accretion geometries over

[i] Whereas the gravitational force acts mostly on the protons, the radiation force acts primarily on the electrons. These two species are tied together by a global electric field, so that the entire "plasma" (ionized gas) behaves as a single quasi-neutral fluid which is subject to both forces. Under similar circumstances, electrons are confined to the Sun by an electric potential of about a kilo-Volt (corresponding to a total charge of ~ 75 Coulombs). The opposite electric forces per unit volume acting on electrons and ions in the Sun cancel out so that the total pressure force is exactly balanced by gravity, as for a neutral fluid. An electric potential of 1-10 kilo-Volts also binds electrons to clusters of galaxies (where the thermal velocities of these electrons, $\sim 0.1c$, are well in excess of the escape speed from the gravitational potential). For a general discussion, see Loeb, A. *Phys. Rev.* **D37**, 3484 (1988).

the years. But, remarkably, observed quasars for which black hole masses can be measured by independent methods appear to respect this limit. Substituting all constants, the Eddington luminosity is given by,

$$L_E = 1.3 \times 10^{44} \left(\frac{M_{\rm BH}}{10^6 M_\odot} \right) \ {\rm erg\ s^{-1}}, \tag{1.34}$$

Interestingly, the scattering cross section per unit mass for UV radiation on dust is larger by two orders of magnitude than σ_T/m_p (A. Laor & B. Draine ApJ, 402, 441, 1993). Although dust is destroyed within $\sim 10^4 G M_{\rm BH}/c^2$ by the strong illumination from an Eddington-limited quasar (H. Netzer & A. Laor ApJ, 404, L51, 1993), it should survive at larger distances. Hence, the radiation pressure on dust would exceed the gravitational force towards the black hole and drive powerful outflows. Spectral lines could be even more effective than dust in their coupling to radiation. The integral of the absorption cross-section of a spectral line over frequency,

$$\int \sigma(\nu) d\nu = f_{12} \left(\frac{\pi e^2}{m_e c} \right), \tag{1.35}$$

is typically orders of magnitude larger than $\sigma_T \nu_{21}$ where ν_{21} is the transition frequency and f_{12} is the absorption oscillator strength. For example, the Lyα transition of hydrogen, for which $f_{12} = 0.416$, provides an average cross-section which is seven orders of magnitude larger than σ_T when averaged over a frequency band as wide as the resonant frequency itself. Therefore, lines could be even more effective at driving outflows in the outer parts of quasar environments.

The total luminosity from gas accreting onto a black hole, L, can be written as some radiative efficiency ϵ times the mass accretion rate $\dot{M}$,

$$L = \epsilon \dot{M} c^2, \tag{1.36}$$

with the black hole accreting the non-radiated component, $\dot{M}_{\rm BH} = (1-\epsilon)\dot{M}$. The equation that governs the growth of the black hole mass is then

$$\dot{M}_{\rm BH} = \frac{M_{\rm BH}}{t_E}, \tag{1.37}$$

where (after substituting all fundamental constants),

$$t_E = 4 \times 10^7 {\rm years} \left(\frac{\epsilon/(1-\epsilon)}{10\%} \right) \left(\frac{L}{L_E} \right)^{-1}. \tag{1.38}$$

We therefore find that as long as fuel is amply supplied, the black hole mass grows exponentially in time, $M_{\rm BH} \propto \exp\{t/t_E\}$, with an e-folding time t_E. Since the growth time in equation (1.38) is significantly shorter than the $\sim 10^9$ years corresponding to the age of the Universe at a redshift $z \sim 6$ – where black holes with a mass $\sim 10^9 M_\odot$ are found, one might naively conclude that there is plenty of time to grow the observed black hole masses from small seeds. For example, a seed black hole from a Population III star of $100 M_\odot$ can grow in less than a billion years up to $\sim 10^9 M_\odot$ for $\epsilon \sim 10\%$ and $L \sim L_E$. However, the intervention of various processes makes it unlikely that a stellar mass seed will be able to accrete continuously at its Eddington limit with no interruption.

For example, mergers are very common in the early Universe. Every time two gas-rich galaxies come together, their black holes are likely to coalesce. The coalescence is initially triggered by "dynamical friction" on the surrounding gas and stars, and is completed – when the binary gets tight – as a result of the emission of gravitational radiation.[4] The existence of gravitational waves is a generic prediction of Einstein's theory of gravity. They represent ripples in space-time generated by the motion of the two black holes as they move around their common center of mass in a tight binary. The energy carried by the waves is taken away from the kinetic energy of the binary, which therefore gets tighter with time. Computer simulations reveal that when two black holes with unequal masses merge to make a single black hole, the remnant gets a kick due to the non-isotropic emission of gravitational radiation at the final plunge.[ii] This kick was calculated recently using advanced computer codes that solve Einstein's equations (a task that was plagued for decades with numerical instabilities).[5] The typical kick velocity is hundreds of kilometer per second (and up to ten times more for special spin orientations), bigger than the escape speed from the first dwarf galaxies.[6] This implies that continuous accretion was likely punctuated by black hole ejection events,[7] forcing the merged dwarf galaxy to grow a new black hole seed from scratch.[iii]

If continuous feeding is halted, or if the black hole is temporarily removed from the center of its host galaxy, then one is driven to the conclusion that the black hole seeds must have started more massive than $\sim 100 M_{\odot}$. More massive seeds may originate from supermassive stars. *Is it possible to make such stars in early galaxies?* Yes, it is. Numerical simulations indicate that stars weighing up to a million Suns could have formed at the centers of early dwarf galaxies which were barely able to cool their gas through transitions of atomic hydrogen, having $T_{\rm vir} \sim 10^4$K and no H_2 molecules. Such systems have a total mass that is several orders of magnitude higher than the earliest Jeans-mass condensations discussed in §4.1. In both cases, the gas lacks the ability to cool well below $T_{\rm vir}$, and so it fragments into one or two major clumps. The simulation shown in Figure 1.4 results in clumps of several million solar masses, which inevitably end up as massive black holes. The existence of such seeds would have given a jump start to the black hole growth process.

The nuclear black holes in galaxies are believed to be fed with gas in episodic events of gas accretion triggered by mergers of galaxies. The energy released by the accreting gas during these episodes could easily unbind the gas reservoir from the host galaxy and suppress star formation within it. If so, nuclear black holes regulate their own growth by expelling the gas that feeds them. In so doing, they also shape the stellar content of their host galaxy. This may explain the observed

[ii] The gravitational waves from black hole mergers at high redshifts could in principle be detected by a proposed space-based mission called the *Laser Interferometer Space Antenna* (LISA). For more details, see http://lisa.nasa.gov/, and, for example, Wyithe, J. S. B., & Loeb, A. *Astrophys. J.* **590**, 691 (2003).

[iii] These black hole recoils might have left observable signatures in the local Universe. For example, the halo of the Milky Way galaxy may include hundreds of freely-floating ejected black holes with compact star clusters around them, representing relics of the early mergers that assembled the Milky Way out of its original building blocks of dwarf galaxies (O'Leary, R. & Loeb, A. *Mon. Not. R. Astron. Soc.* **395**, 781 (2009)).

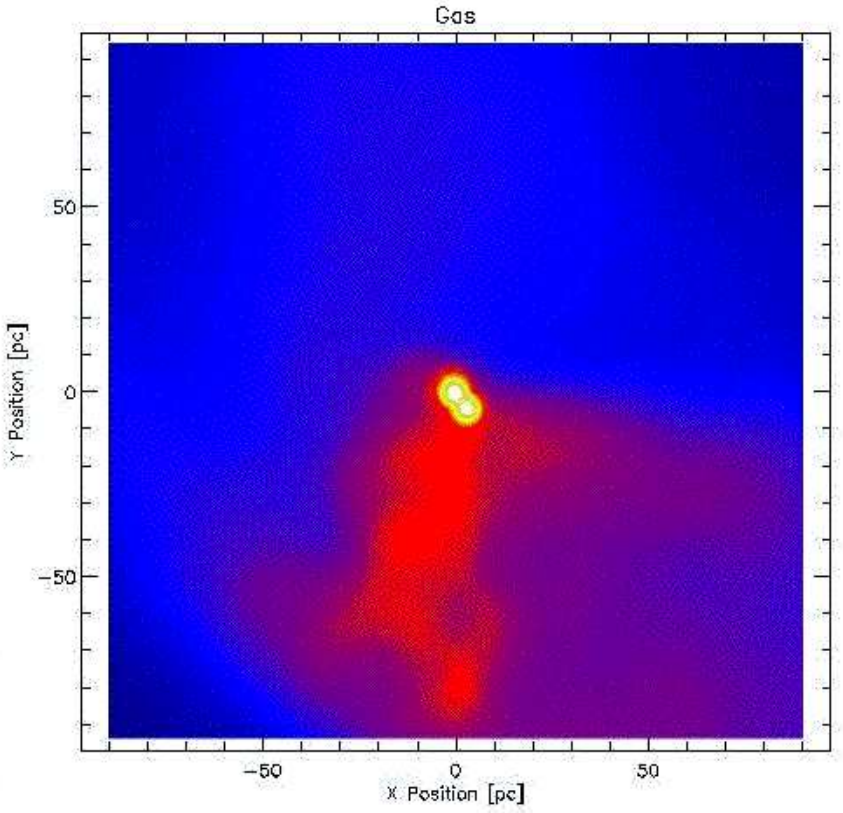

Figure 1.4 Numerical simulation of the collapse of an early dwarf galaxy with a virial temperature just above the cooling threshold of atomic hydrogen and no H_2. The image shows a snapshot of the gas density distribution 500 million years after the Big Bang, indicating the formation of two compact objects near the center of the galaxy with masses of $2.2 \times 10^6 M_\odot$ and $3.1 \times 10^6 M_\odot$, respectively, and radii < 1 pc. Sub-fragmentation into lower mass clumps is inhibited because hydrogen atoms cannot cool the gas significantly below its initial temperature. These circumstances lead to the formation of supermassive stars that inevitably collapse to make massive seeds of supermassive black holes. The simulated box size is 200 pc on a side. Figure credit: Bromm, V. & Loeb, A. *Astrophys. J.* **596**, 34 (2003).

tight correlations between the mass of central black holes in present-day galaxies and the velocity dispersion[8] $\sigma_\star$ or luminosity[9] $L_{\rm sp}$ of their host spheroids of stars (namely, $M_{\rm BH} \propto \sigma_\star^4$ or $M_{\rm BH} \propto L_{\rm sp}$). Since the mass of a galaxy at a given redshift scales with its virial velocity as $M \propto V_c^3$ in equation (**??**), the binding energy of galactic gas is expected to scale as $MV_c^2 \propto V_c^5$ while the momentum required to kick the gas out of its host would scale as $MV_c \propto V_c^4$. Both scalings can be tuned to explain the observed correlations between black hole masses and the properties of their host galaxies.[10] Star formation inevitably precedes black hole fueling, since the outer region of the accretion flows that feed nuclear black holes is typically unstable to fragmentation[11]. This explains the high abundance of heavy elements inferred from the broad emission lines of quasars at all redshifts[12].

The feedback regulated growth explains why quasars may shine much brighter than their host galaxies. A typical star like the Sun emits a luminosity, $L_\odot = 4\times10^{33}\,{\rm erg\,s^{-1}}$ which can also be written as a frcation $\sim 3\times10^{-5}$ of its Eddington luminosity $L_E = 1.4\times10^{38}\,{\rm erg\,s^{-1}}$. Black holes grow up to a fraction $\sim 10^{-3}$ of the stellar mass of their spheroid. When they shine close to their Eddington limit, they may therefore outshine their host galaxy by up to a factor of $\sim (10^{-3}/3\times 10^{-5})$, namely 1–2 orders of magnitude.

The inflow of cold gas towards galaxy centers during the growth phase of their black holes would naturally be accompanied by a burst of star formation. The fraction of gas not consumed by stars or ejected by supernova-driven winds will continue to feed the black hole. It is therefore not surprising that quasar and starburst activities co-exist in ultra-luminous galaxies, and that all quasars show strong spectral lines of heavy elements. Similarly to the above-mentioned prescription for modelling galaxies, it is possible to "dress up" the mass distribution of halos in Figure **??** with quasar luminosities (related to L_E, which is a prescribed function of M based on the observed $M_{\rm BH}$–$\sigma_\star$ relation) and a duty cycle (related to t_E), and find the evolution of the quasar population over redshift. This simple approach can be tuned to give good agreement with existing data on quasar evolution.[13]

The early growth of massive black holes led to the supermassive black holes observed today. In our own Milky Way galaxy, stars are observed to zoom around the Galactic center at speeds of up to ten thousand kilometers per second, owing to the strong gravitational acceleration near the central black hole.[14] But closer-in observations are forthcoming. Existing technology should soon be able to image the silhouette of the supermassive black holes in the Milky Way and M87 galaxies directly (see Figure 1.2).

1.4 BLACK HOLE BINARIES

Nearly all nearby galactic spheroids are observed to host a nuclear black hole. Therefore, the hierarchical buildup of galaxies through mergers must generically produce black hole binaries. Such binaries tighten through dynamical friction on the background gas and stars, and ultimately coalesce through the emission of gravitational radiation.

In making a tight binary from a merger of separate galaxies, the mass ratio of

two black holes cannot be too extreme. A satellite of mass $M_{\rm sat}$ in a circular orbit at the virial radius of a halo of mass $M_{\rm halo}$ would sink to the center on a dynamical friction time of $\sim 0.1 t_H (M_{\rm halo}/M_{\rm sat})$, where t_H is the Hubble time. If the orbit is eccentric with an angular momentum that is a fraction ε of a circular orbit with the same energy, then the sinking time reduces by a factor of $\sim \varepsilon^{0.4}$. Therefore, mostly massive satellites with $M_{\rm sat} > 0.1 M_{\rm halo}$ bring their supermassive black holes to the center of their host halos during the age of the Universe.

As a satellite galaxy sinks, its outer envelope of dark matter and stars is stripped by tidal forces. The stripping is effective down to a radius inside of which the mean mass density of the satellite is comparable to the ambient density of the host galaxy. Eventually, the two black holes are stripped down to the cores of their original galaxies and are surrounded by a circumbinary envelope of stars and gas. As long as the binary is not too tight, the reservoir of stars within the binary orbit can absorb the orbital binding energy of the binary and allow it to shrink. However, when the orbital velocity starts to exceed the local velocity dispersion of stars, a star impinging on the binary would typically be expelled from the galactic nucleus at a high speed. This happens at the so-called the "hardenning radius" of the binary,

$$a_{\rm hard} \approx 0.1 \frac{q}{(1+q)^2} M_6 \left(\frac{\sigma_\star}{100\ {\rm km\ s^{-1}}}\right)^{-2}\ {\rm pc}, \tag{1.39}$$

at which the binding energy per unit mass of the binary exceeds $\frac{3}{2}\sigma^2$, where σ is the velocity dispersion of the stars before the binary tightenned. Here, $M_6 = ([M_1 + M_2]/10^6 M_\odot)$, where M_1 and M_2 are the masses of the two black holes, $q = M_1/M_2$ is their mass ratio, and $\mu = M_1 M_2/(M_1 + M_2)$ is the reduced mass of the binary.

A hard binary will continue to tighten only by expelling stars which cross its orbit and so unless the lost stars are replenished by new stars which are scattered into an orbit that crosses the binary (through dynamical relaxation processes in the surrounding galaxy) the binary would stall. This "final parsec problem" is circumvented if gas streams into the binary from a circumbinary disk. Indeed, the tidal torques generated during a merger extract angular momentum from any associated cold gas and concentrate the gas near the center of the merger remnant, where its accretion often results in a bright quasar.

If the two black holes are in a circular orbit of radius $a < a_{\rm hard}$ around each other, their respective distances from the center of mass are $a_i = (\mu/M_i)a$ ($i = 1, 2$). We define the parameter $\zeta = 4\mu/(M_1+M_2)$, which equals unity if $M_1 = M_2$ and is smaller otherwise. The orbital period is given by,

$$P = 2\pi (GM/a^3)^{-1/2} = 1.72 \times 10^{-2}\ {\rm yr}\ a_{14}^{3/2} M_6^{-1/2}, \tag{1.40}$$

where, $a_{14} \equiv (a/10^{14}\ {\rm cm})$. The angular momentum of the binary can be expressed in terms of the absolute values of the velocities of its members v_1 and v_2 as $J = \Sigma_{i=1,2} M_i v_i a_i = \mu v a$, where the relative orbital speed is

$$v = v_1 + v_2 = (2\pi a/P) = 1.15 \times 10^4\ {\rm km\ s^{-1}} M_6^{1/2} a_{14}^{-1/2}\ . \tag{1.41}$$

In gas-rich mergers, the rate of inspiral slows down as soon as the gas mass interior to the binary orbit is smaller than μ and the enclosed gas mass is no longer

sufficient for carrying away the entire orbital angular momentum of the binary, J. Subsequently, momentum conservation requires that fresh gas will steadily flow towards the binary orbit in order for it to shrink. The binary tightens by expelling gas out of a region twice as large as its orbit (similarly to a "blender" opening a hollow gap) and by torquing the surrounding disk through spiral arms. Fresh gas re-enters the region of the binary as a result of turbulent transport of angular momentum in the surrounding disk. Since the expelled gas carries a specific angular momentum of $\sim va$, the coalescence time of the binary is inversely proportional to the supply rate of fresh gas into the binary region. In a steady state, the mass supply rate of gas that extracts angular momentum from the binary, $\dot{M}$, is proportional to the accretion rate of the surrounding gas disk. Given that a fraction of the mass that enters the central gap accretes onto the BHs and fuels quasar activity, it is appropriate to express $\dot{M}$ in Eddington units $\dot{\mathcal{M}} \equiv \dot{M}/\dot{M}_E$, corresponding to the accretion rate required to power the limiting Eddington luminosity with a radiative efficiency of 10%, $\dot{M}_E = 0.023 M_\odot \ \mathrm{yr}^{-1} M_6$. We then find,

$$t_{\rm gas} \approx (J/\dot{M}va) = \mu/\dot{M} = 1.1 \times 10^7 \ \mathrm{yr} \ \zeta \dot{\mathcal{M}}^{-1}. \tag{1.42}$$

For a steady $\dot{\mathcal{M}}$, the binary spends equal amounts of time per log a until GWs start to dominate its loss of angular momentum.

The coalescence timescale due to GW emission is given by,

$$t_{\rm GW} = \frac{5}{256} \frac{c^5 a^4}{G^3 M^2 \mu} = 2.53 \times 10^3 \ \mathrm{yr} \ \frac{a_{14}^4}{\zeta M_6^3}. \tag{1.43}$$

By setting $t_{\rm GW} = t_{\rm gas}$ we can solve for the orbital speed, period, and separation at which GWs take over,

$$v_{\rm GW} = 4.05 \times 10^3 \ \mathrm{km \ s^{-1}} \ \zeta^{-1/4} (\dot{\mathcal{M}} M_6)^{1/8} \ ; \tag{1.44}$$

$$P_{\rm GW} = 0.4 \ \mathrm{yr} \ \zeta^{3/4} M_6^{5/8} \dot{\mathcal{M}}^{-3/8} \ ; \tag{1.45}$$

$$a_{\rm GW} = 2.6 \times 10^{-4} \ \mathrm{pc} \ \zeta^{1/2} M_6^{3/4} \dot{\mathcal{M}}^{-1/4}. \tag{1.46}$$

For a binary redshift z, the observed period is $(1+z)P_{\rm GW}$. The orbital speed at which GWs take over is very weakly dependent on the supply rate of gas, $v_{\rm GW} \propto \dot{M}^{1/8}$. It generically corresponds to an orbital separation of order $\sim 10^3$ Schwarzschild radii ($2GM/c^2$). The probability of finding binaries deeper in the GW-dominated regime, $\mathcal{P} \propto t_{\rm GW}$, diminishes rapidly at increasing orbital speeds, with $\mathcal{P} = \mathcal{P}_{\rm GW} (v/v_{\rm GW})^{-8}$.

Black hole binaries can be identified visually or spectroscopically. At large separations the cores of the merging galaxies can be easily identified as separate entities. If both black holes are active simultaneously, then the angular separation between the brightness centroids can in principle be resolved at X-ray, optical, infrared, or radio wavelengths. The UV illumination by a quasar usually produces narrow lines from gas clouds at kpc distances within its host galaxy or broad lines from denser gas clouds at sub-pc distances from it. Therefore the existence of a binary can be inferred from various spectroscopic offsets: *(i)* between two sets of narrow lines if the galaxies are separated by more than a few kpc and both show quasar activity at the same time; *(ii)* between the narrow emission lines of the gas and the

absorption lines of the stars due to the tidal interaction between the galaxies at a multi-kpc separation; *(iii)* between narrow lines and broad lines if the black hole binary separation is between the kpc and pc scales. The last offset signature can also be produced by a single quasar which gets kicked out of the center of its host galaxy while carrying the broad-line region with it. Such a kick could be produced either by the anisotropic emission of gravitational waves during the coalescence of a binary (producing up to $\sim 200\ \mathrm{km\ s^{-1}}$ in a merger of non-spinning black holes, and up to $\sim 4,000\ \mathrm{km\ s^{-1}}$ for special spin orientation), or from triple black hole systems that are made when a third black hole is added to a galaxy before the binary coalesces. Aside from providing a test of general relativity in the strong field limit, these kicks have an important effect in suppressing the growth of black holes in small galaxies at high redshifts.