Disc Accretion in Active Galactic Nuclei

Andrew King

Theoretical Astrophysics Group, University of Leicester, Leicester LE1 7RH, UK

Abstract

I review disc accretion in AGN. I consider the conditions for forming discs small enough to accrete within a Hubble time, black hole feedback and the M–sigma and SMBH–bulge mass relations, and the spin of supermassive black holes.

Key words:

1 Introduction

Accretion of matter on to a black hole is the most effective way of extracting energy from normal matter. This process must therefore power the brightest objects in the Universe, including AGN, and shows that the black hole mass in these objects is growing. The centre of almost every galaxy is now known to host a supermassive black hole (SMBH). The need to grow these holes to their current huge masses must mean that almost every galaxy is active from time to time.

As we shall see, accretion on to the supermassive black holes in AGN requires the infalling gas to lose almost all of its angular momentum. Some form of disc accretion is therefore inevitable. This allows us to profit from much of the knowledge gained by studying accretion discs in other contexts, particularly accreting binary systems. For example mass loss through winds is a very common feature of disc accretion, and is particularly important when the hole is fed mass at rates above the Eddington value. Basic ideas about discs show that such systems have a strong effect on their surroundings, either the interstellar medium for binaries, or the entire galaxy bulge for SMBH. In the latter case this feedback leaves an imprint of the SMBH on the whole structure of the bulge.

A clear difference between accretion in close binaries and in AGN is that in binaries the angular momentum of the accreting gas is often constrained to be always in the same sense. By contrast in AGN each new accretion episode may have angular momentum completely uncorrelated with earlier or later episodes. Thus while accreting black holes in binaries generally spin up, this is much less obvious in AGN. Since the black hole mass in AGN increases by large factors this opens the possibility of changing the accretion efficiency, which becomes very large for rapid black–hole spin.

I review these and other problems briefly below.

2 Accretion disc theory

Accretion disc theory is the subject of many books and reviews (see e.g. Frank et al., 2002 and Pringle, 1981). Accordingly this section simply summarizes the main results without giving detailed derivations.

2.1 Disc formation

Matter accreting on to a mass M forms a disc if its specific angular momentum J is too large for it to impact the object directly. I define the circularization radius

$$R_{\rm circ} = \frac{J^2}{GM}, \tag{1}$$

which is where the matter would orbit if it lost energy but no angular momentum. The condition for disc formation is that $R_{\rm circ}$ should exceed the effective size of the accretor (a parabolic orbit with specific angular momentum J would reach a minimum separation $0.5R_{\rm circ}$). In AGN the accretor is a black hole, and the the effective size is the radius of the innermost stable circular orbit $R_{\rm ISCO}$.

Any conceivable source of accreting matter in AGN has specific angular momentum J large enough to ensure $R_{\rm circ} >> R_{\rm ISCO}$; indeed we shall see later that the scale of this inequality is itself a problem for AGN feeding. Under these conditions disc formation follows if, as usual, energy is lost through dissipation faster than angular momentum is redistributed. Since the orbit of lowest energy for a given angular momentum is a circle, matter follows a sequence of circular orbits about the compact accretor.

The agency for both energy dissipation and angular momentum transport is usually called viscosity (although this cannot be the standard small–scale viscosity of kinetic theory). Even after several decades of work it is still not properly understood. The best candidate mechanism invokes the magnetorotational instability (MRI: Balbus & Hawley, 1991). Here a comparatively weak magnetic field threading the disc is wound up by the shear, and transports angular momentum outwards. Reconnection limits the field growth and produces dissipation. Numerical simulations show that this is a promising mechanism, but it is not yet clear that the effect is large enough (cf King et al, 2007) to satisfy observational constraints. There are even worries (Fromang & Papaloizou, 2007) that current simulations give results which depend on numerical resolution.

This lack of knowledge of viscosity means that there is as yet no deterministic theory of accretion discs. We cannot make definite predictions as to what will happen in a completely general case. This leaves open alternative possibilities, only some of which will survive as realistic when a full understanding of viscosity emerges. A typical example of such alternatives is the question of whether gas always tries to cool and spiral slowly inwards, or if there exist conditions under which it can rid itself very rapidly of angular momentum and advect rapidly inwards without radiating significantly (a so–called ADAF). Similarly the question of how jets form and remove matter from accretion discs remains unsolved because of our lack of understanding of viscosity.

This fundamental indeterminacy has allowed theorists free rein in imagining various interesting possibilities. Fortunately there are certain situations in which we can sidestep our lack of understanding of viscosity and draw fairly clear conclusions. The resulting good agreement with observation does suggest that the broad outlines of the incomplete theory are reasonable.

2.2 Thin discs

While viscosity transports angular momentum and thus spreads the initial ring at $R_{\rm circ}$ into a disc, the nature of this accretion disc is determined by the efficiency with which the disc can cool. Very often this efficiency is high. Intuitively this suggests that the matter cannot have dynamically significant pressure, so that the circular orbits of disc gas are actually Keplerian. Simultaneously it seems likely that the disc is *thin*: that is, its scaleheight H obeys

$$H \simeq \frac{c_{\rm s}}{v_{\rm K}} R << R \qquad (2)$$

at disc radius R, where $c_{\rm s}$ is the local sound speed, and

$$v_{\rm K} = \left(\frac{GM}{R}\right)^{1/2} \tag{3}$$

is the Kepler velocity, with M the accretor mass. In this state the azimuthal velocity is close to $v_{\rm K}$ and the radial and vertical velocities are much smaller. The properties of being thin, Keplerian and efficiently cooled are all equivalent, and if any one of them breaks down so do the other two.

If the thin disc approximation holds, the vertical structure is almost hydrostatic and decouples from the horizontal structure, which can be described in terms of its surface density Σ. Mass and angular momentum conservation imply that this obeys a nonlinear diffusion equation

$$\frac{\partial \Sigma}{\partial t} = \frac{3}{R}\frac{\partial}{\partial R}\left(R^{1/2}\frac{\partial}{\partial R}[\nu\Sigma R^{1/2}]\right). \tag{4}$$

Here ν is the kinematic viscosity, which is usually parametrized as

$$\nu = \alpha c_{\rm s} H. \tag{5}$$

where α is a dimensionless number. In a steady state this gives

$$\nu\Sigma = \frac{\dot M}{3\pi}\left[1 - \beta\left(\frac{R_{\rm in}}{R}\right)^{1/2}\right], \tag{6}$$

where $\dot M$ is the accretion rate and the dimensionless quantity β is specified by the boundary condition at the inner edge $R_{\rm in}$ of the disc. For example, a disc ending at the radius R_* of a non–rotating star has $R_{\rm in} = R_*$. There is some debate as to the correct value of β for a black hole accretor, as this depends on whether the accretion flow within the ISCO has significant magnetic content to connect it to the disc flow further out (cf Krolik, 1999).

In a steady thin disc dissipation $D(R)$ per unit surface area is also proportional to $\nu\Sigma$, i.e.

$$D(R) = \frac{9}{8}\nu\Sigma\frac{GM}{R^3}\left[1 - \beta\left(\frac{R_{\rm in}}{R}\right)^{1/2}\right], \tag{7}$$

so that the surface temperature T is independent of the viscosity ν despite being entirely generated by it:

$$T = T_{\rm visc} = \left\{\frac{3GM\dot M}{8\pi R^3\sigma}\left[1 - \beta\left(\frac{R_*}{R}\right)^{1/2}\right]\right\}^{1/4}. \tag{8}$$

The total accretion luminosity is given by integrating $D(R)$ over both sides of the disc. For a black hole we have to use general relativity at radii close to the Schwarzschild radius. The result is

$$L_{\rm acc} = \epsilon \dot{M} c^2 \tag{9}$$

where $\epsilon \sim 0.05 - 0.42$ is the efficiency of rest–mass conversion, which depends on the inner boundary condition, and in particular the black hole spin.

2.3 Disc timescales

Equation (4) shows that Σ changes on a timescale

$$t_{\rm visc} \sim \frac{l^2}{\nu} \tag{10}$$

if its spatial gradient is over a lengthscale l. Hence we would expect a disc to make significant changes in its surface density and thus its luminosity on a timescale $\sim R^2/\nu$, where R is its outer radius. We can use this fact to get an idea of the magnitude of the viscosity in observed discs in close binaries. In dwarf novae, which are short–period white–dwarf binaries, the disc size is $R \sim 1{-}3\times 10^{10}$ cm, and surface density changes take a few days. This suggests that $\alpha \sim 0.1$.

There are two other obvious timescales in a disc. The first is the dynamical timescale

$$t_{\rm dyn} \sim \frac{R}{v_{\rm K}} = \left(\frac{R^3}{GM}\right)^{1/2}, \tag{11}$$

characterizing states in which dynamical equilibrium is disturbed; note that vertical hydrostatic balance is resored on a timescale

$$t_z \sim \frac{H}{c_{\rm s}} = \frac{R}{v_{\rm K}} = t_{\rm dyn} \tag{12}$$

where we have used eqn (2). The second is the thermal timescale

$$t_{\rm th} = \frac{\Sigma c_{\rm s}^2}{D(R)} \sim \frac{R^3 c_{\rm s}^2}{GM\nu} = \frac{c_{\rm s}^2}{v_{\rm K}^2}\frac{R^2}{\nu} = \left(\frac{H}{R}\right)^2 t_{\rm visc} \tag{13}$$

where we have used eqn (7). The alpha–disc parametrization (4) can be used to show that

$$t_{\rm visc} \sim \frac{1}{\alpha}\left(\frac{H}{R}\right)^{-2} t_{\rm dyn} \tag{14}$$

so we finally have the ordering

$$t_{\rm dyn} \sim t_z \sim \alpha t_{\rm th} \sim \alpha (H/R)^2 t_{\rm visc}, \tag{15}$$

i.e. dynamical < thermal < viscous.

3 AGN Discs

To get some idea of typical AGN disc conditions we consider a case with $M = 10^8 {\rm M_\odot}$, $\dot M = 1 {\rm M_\odot\,yr^{-1}}$. The gravitational energy release is dominated by the central regions of the disc, where $R \sim$ few $\times\, GM/c^2 \sim$ few $\times\, 10^{13}$ cm. We can easily check that the condition for a thin disc (i.e. efficient cooling) is satisfied here. The temperature (8) is of order $\sim$ few $\times\, 10^5$ K Thus we expect most of the luminosity from AGN to be emitted in the UV and soft X–rays (in the rest–frame). The dynamical and thermal timescales in the central regions are $\sim 10^3 - 10^4$ s respectively. These are the shortest possible timescales for significant variability.

Although the centre of the disc dominates the emitted luminosity, most of the mass is stored in the outer regions, and must move inwards under viscosity to power the AGN. To estimate the timescale

$$t_{\rm visc} \sim \frac{R^2}{\alpha c_s H} \tag{16}$$

we have to solve the steady-state disc equations (see e.g. Frank et al., 2002). These show that for radii $R = 10^{18} R_{18}$ cm we have $H/R \sim 10^{-3}$ and $c_s \sim 10^5 R_{18}^{-1/2}$ ${\rm cm\,s^{-1}}$, so that

$$t_{\rm visc} \sim 10^{10} \left(\frac{\alpha}{0.03}\right) R_{18}^{3/2}\ {\rm yr}. \tag{17}$$

In other words, *the timescale on which mass moves inwards to power the AGN approaches the Hubble time for disc radii of order 0.3 pc.* This is an extremely powerful constraint. It shows that gas feeding an AGN must have low angular momentum *before* it forms a disc, otherwise the value of $R_{\rm circ}$ will be so high that there is no hope of the gas ever reaching the black hole. The gas must fall towards the hole with an impact parameter of no more than a few tenths of a parsec, which is tiny on the scale of a galaxy. Such a precise aim is very unlikely unless the feeding process somehow involves a much wider distribution of matter, most of which never accretes on to the SMBH. This accords at least qualitatively with the idea that the basic mechanism driving black hole growth is the same that builds up the bulge of a galaxy, namely mergers of smaller galaxies. As we will see, this typically gives black hole masses M which are $\sim 10^{-3}$ of the bulge mass, pointing to a process of SMBH growth which is inherently wasteful in mass terms, just as we deduced above. A further qualitative agreement is that the randomness of the accretion process means that there is no correlation between its instantaneous axis, as revealed by the observed directions of radio jets, and the large–scale structure of the host galaxy.

The conditions discussed above typify bright AGN, i.e. those whose black holes are growing rapidly. Of course the thin disc condition itself must fail if the accreting matter does not cool efficiently. This can for example happen in low–luminosity AGN (LLAGN).

4 SMBH feedback

The disc theory discussed above assumes that the accretion luminosity has no effect on the accretion flow itself. However this assumption fails at luminosities $L_{\rm acc} \geq L_{\rm Edd}$, where

$$L_{\rm Edd} = \frac{4\pi GMc}{\kappa} \tag{18}$$

is the Eddington value, with $\kappa \sim 0.3\,{\rm cm^2\,g^{-1}}$ the electron scattering opacity. For $L_{\rm acc} \geq L_{\rm Edd}$ the disc drives off the excess accretion at each radius R so as to keep its local accretion luminosity $\sim GM\dot{M}/R$ just below the radiation pressure limit. Thus $\dot{M}(R)$ decreases as R, and the hole gains mass at a rate which is just $\dot{M}_{\rm Edd} = L_{\rm Edd}/\epsilon c^2$, where ϵ is the radiation efficiency specified by the ISCO (and thus dependent on the Kerr spin parameter a). The result (Shakura & Sunyaev, 1973) is a luminosity only logarithmically above $L_{\rm Edd}$, and an outflowing wind carrying away the super–Eddington mass rate $\dot{M}_{\rm out} = \dot{M} - \dot{M}_{\rm Edd}$ at a speed $v \sim (\dot{M}_{\rm Edd}/\dot{M})c$. This carries total momentum

$$\dot{M}_{\rm out} v \sim \frac{L_{\rm Edd}}{c}, \tag{19}$$

and total energy $\sim \dot{M}_{\rm out}v^2/2 \sim L_{\rm Edd}v/c$.

There is direct evidence of such outflows with $v \sim 0.1c$ in some AGN (e.g. Pounds et al., 2003a, b), and good reason to assume that they occur during the most rapid growth phases of SMBH, as even growth at the rate $\dot{M}_{\rm Edd}$ is barely enough to account for observed SMBH masses at high redshift. It is also obvious that they can have a major effect on the host galaxy. The Eddington outflow must impact the gas of the host bulge and sweep it up in a shell. The speed of the shell depends on whether the shocked outflowing gas cools or not. If it does, the host gas feels simply the momentum rate (19) (a momentum–driven outflow). If the gas cannot cool within the flow timescale, it also communicates its thermal pressure to the host gas, driving this outwards at higher speed (an energy–driven outflow). King (2003, 2005) shows that in a typical bulge, Compton cooling establishes momentum–driven conditions at small radii. The outflow sweeps up a shell, which stalls fairly close to the SMBH, until this grows its mass to the critical value

$$M_\sigma = \frac{f_g \kappa}{\pi G^2}\sigma^4 = 2\times 10^8 {\rm M_\odot}\sigma_{200}^4 \tag{20}$$

Here $f_g = 0.16$ is the cosmic gas fraction $\Omega_{\rm baryon}/\Omega_{\rm matter}$ and $\sigma = 200\sigma_{200}$ km s^{-1} is the velocity dispersion of the host bulge. At this point the shell expands rapidly, reaching radii where Compton cooling is no longer effective. It then accelerates, cutting off the mass supply to the SMBH, and indeed the gas in the bulge, at a value

$$M_{\rm bulge} \sim \left(\frac{m_p}{m_e}\right)^2 \frac{\sigma}{c} M \sim 10^3 M \tag{21}$$

where m_p, m_e are the proton and electron masses, and at the last step I have assumed a typical velocity dispersion $\sigma \sim 200$ km s^{-1}.

Despite having no free parameter, (20) is in excellent agreement with observations of the $M-\sigma$ relation (Ferrarese & Merritt, 2000; Gebhardt et al., 2000). The SMBH–bulge mass relation is similarly close to observation. Note that it is actually of the form $M_{\rm bulge} \propto M^{5/4}$, which agrees well with the Faber–Jackson relation (McLaughlin et al, 2006).

The agreements here suggest that the $M-\sigma$ and $M_{\rm bulge}-M$ relations are consequences of momentum–driven feedback from an Eddington outflow at the black hole. It is easy to show that an energy–driven outflow would be too efficient in driving mass away, and produce too small a value for M_σ and $M_{\rm bulge}$. Cosmological simulations of these effects adopt for numerical reasons a form of distributed energy deposition, rather than solving the interaction of the outflow with the bulge. These produce acceptable answers for M_σ and $M_{\rm bulge}$ if one assumes that the distributed energy is only a small fraction (actually $\sim \sigma/c \sim 10^{-3}$) of that radiated by the black hole. The need to put this fraction in by hand is a clear sign that a good deal of the physics producing these relations is missing from this approach.

5 SMBH spin

As remarked in the Introduction, accretion on to SMBH in AGN differs from stellar–mass black hole accretion in close binaries in its randomness. In particular the initial sense of the accretion flow's angular momentum must be retrograde with respect to the hole spin about one–half of the time. One might expect that this would automatically lead to slowly–spinning SMBH, as retrograde accretion would cancel prograde. Indeed the retrograde case has a larger lever arm, strengthening the argument. However until recently the opposite view, that SMBH are all rapidly spinning, was the accepted one (cf Volonteri et al, 2005).

The reason for this is the Lense–Thirring (LT) effect, i.e. dragging of inertial frames. In the context of black–hole accretion this means that a test–particle orbit inclined wrt the black hole spin must precess, at a rate which goes as R^{-3}. However the matter in an accretion disc is not test particles, but gas which has viscosity. This means that the differential precession caused by the LT effect produces a viscous torque between the hole spin and the disc. By Hawking's theorem this must tend to produce an axisymmetric situation. The first calculations of the effect (Scheuer & Feiler, 1996) suggested that the end effect was a disc co–aligned with the hole spin.

Since this co–alignment occurs on a viscous timescale, which is much shorter than the mass–doubling timescale on which the hole accretes angular momentum, this result would imply that all the mass–doubling takes place with the disc accreting in a prograde fashion on to the hole. Since the hole increases its mass enormously over time, this would mean that all SMBH should be spinning at an almost maximal rate (Kerr a parameter ~ 1). Although this makes them bright, as it increases the accretion efficiency to a value $\epsilon \sim 0.42$, the result creates a major difficulty. For since $L_{\rm Edd}$ is uniquely fixed by the mass, the maximum rate $\dot{M}_{\rm Edd} = L_{\rm Edd}/\epsilon c^2$ at which the hole can accrete is severely reduced. This increases the e–folding time for the growth of the SMBH mass. With $a \sim 1$ the most massive SMBHs observed at redshift $z \sim 6$ must have had 'seed' masses which were themselves already $\sim 10^6 {\rm M_\odot}$ or more, before accretion started. By contrast, with more modest values $a \sim 0.5$ growth from even stellar masses is possible (cf King et al. 2008 and references therein).

There have been several attempts to explain how such large seed masses could arise. However they may not be necessary, since Scheuer & Feiler's (1996) result that the LT effect causes co–alignment makes an implicit assumption, namely that the total angular momentum of the disc J_d is much larger than that of the hole J_h. If this assumption is removed, King et al., 2005 showed that *counter*–alignment of disc and hole occurs provided that the two angular momentum vectors are misaligned by an angle θ with $\cos\theta < -J_d/2J_h$. In this case retrograde accretion would be rapidly established, and reduce the hole spin.

There remains the question of whether the condition $J_d < 2J_h$ is ever satisfied. In a recent paper King et al (2008) suggest that the disc size R_d, and thus its total angular momentum $J_d \sim M_d(GMR_d)^{1/2}$, are limited by the fact that the disc becomes self–gravitating outside a radius such that the disc mass M_d exceeds $(H/R)M \sim 10^{-3}M$. From this they draw a number of conclusions. (a) AGN black holes should on average spin moderately; (b) coalescences of AGN black holes in general produce modest recoil velocities, so that there is little likelihood of their being ejected from the host galaxy; (c) black holes can grow even from stellar masses to $\sim 5\times10^9\,\mathrm{M}_\odot$ at high redshift $z \sim 6$; jets produced in successive accretion episodes can have similar directions, but after several episodes the jet direction deviates significantly. They argue that rare examples of massive holes with significant spin may result from coalescences with SMBH of similar mass, and are most likely to be found in giant ellipticals. There currently seems to be no flagrant disagreement with observation for any of these conclusions. Indeed statistical arguments using the inferred background light provided by quasars (Soltan, 1982, and subsequent papers) suggest an average accretion efficiency $\epsilon \sim 0.1$, favouring moderate black hole spin.

References

[18] Balbus, S.A., Hawley, J.F., 1991, ApJ 376, 214

[18] Frank, J., King, A.R., Raine, D.J., 2002, *Accretion Power in Astrophysics* 3rd Ed., Cambridge University Press, Cambridge.

[18] Ferrarese, L., Merritt, D., 2000, ApJ 539, L9

[18] Fromang S., Papaloizou J., 2007, A&A, 476, 1113

[18] Gebhardt, K., Bender, R., Bower, G., Dressler, A., Faber, S.M., Filippenko, A..V., Green, R., Grillmair, C., Ho, L.C., Kormendy, J., Lauer, T.R., Magorrian, J., Pinkney, J., Richstone, D. and Tremaine, S., 2000, ApJ 539, L13

[18] King, A.R., 2003, ApJ, 596, L27

[18] King A., 2005, ApJ, 635, L121

[18] King, A.R., Lubow, S.H., Ogilvie, G.I., Pringle, J.E., 2005, 363, 49

[18] King A. R., Pringle J. E., Livio M., 2007, MNRAS, 376, 1740

[18] Krolik J. H., 1999, ApJ, 515, L73

[18] McLaughlin D.E., King A.R., Nayakshin S., 2006, ApJ, 650, L37

[18] Pounds, K.A., Reeves, J.N., King, A.R., Page, K.L., O'Brien, P.T., Turner, M.J.L., 2003a, MNRAS, 345, 705
[18] Pounds, K.A., King, A.R., Page, K.L., O'Brien, P.T., 2003b, MNRAS 346, 1025
[18] Pringle, J.E., 1981, ARAA 19, 137
[18] Scheuer, P.A.G., Feiler, R., 1996, MNRAS, 282, 29
[18] Shakura, N.I., Sunyaev, R.A., 1973, A&A, 24, 337
[18] Soltan, A., 1982, MNRAS, 200, 115
[18] Volonteri, M., Madau, P., Quataert, E., Rees, M.J., 2005, ApJ, 620, 69