SMBH accretion and mergers: removing the symmetries

Andrew King\(^1\) and Chris Nixon\(^{1,2,3}\)

\(^1\) Department of Physics and Astronomy, University of Leicester, Leicester, LE1 7RH, UK
\(^2\) JILA, University of Colorado and NIST, Boulder, CO 80309-0440, USA

E-mail: ark@leicester.ac.uk

Received 10 July 2013
Published 29 November 2013
Online at stacks.iop.org/CQG/30/244006

Abstract

We review recent progress in studying accretion flows on to supermassive black holes (SMBH). Much of this removes earlier assumptions of symmetry and regularity, such as aligned and prograde disc rotation. This allows a much richer variety of effects, often because cancellation of angular momentum allows rapid infall. Potential applications include lower SMBH spins allowing faster mass growth and suppressing gravitational-wave reaction recoil in mergers, gas-assisted SMBH mergers, and near-dynamical accretion in galaxy centres.

PACS number: 98.62.Js

(Some figures may appear in colour only in the online journal)

1. Introduction

It is now widely accepted that almost all reasonably large galaxies have supermassive black holes (SMBH) in their centres, and further that these holes grow predominantly through luminous accretion of gas (Soltan 1982, Yu and Tremaine 2002). Any realistic model of gas flows around black holes must take into account the angular momentum of the gas. Radiation processes can relieve the gas of orbital energy, but there is no equivalent process which can intrinsically remove angular momentum. As a result, gas flows tend to form rotationally supported discs with characteristic radii given by the angular momentum of the flow (Pringle 1981). Inward gas flow—accretion—would be impossible without some process (usually called ‘viscosity’) transporting angular momentum outwards in the disc, allowing mass to flow inwards (Lynden-Bell and Pringle 1974) and release its gravitational binding energy. Near the black hole this becomes very large, approaching a significant fraction of the rest-mass energy. For this reason, accretion discs probably power the most luminous objects in the universe, and understanding the nature of their angular momentum transport has been a
major goal of modern astrophysics. The most promising candidate is the magnetorotational instability (MRI) (Balbus and Hawley 1991) which injects turbulence into the gas by stretching magnetic field lines.

The viscosity in accretion discs is often parameterized as (Shakura and Sunyaev 1973)

\[ \nu = \alpha c_s H, \]  
\[ (1) \]

where \( c_s \) is the sound speed, \( H \) is the disc angular semi-thickness and \( \alpha < 1 \) is a dimensionless parameter. In principle \( \alpha \) is a function of position within the disc gas. This formalism characterises the maximum viscosity feasible in an accretion disc through an efficiency parameter: any turbulent gas velocities above the sound speed would quickly shock and dissipate, and any turbulent length scales longer than \( H \) would lead to a disc thicker than \( H \).

The maximum viscosity is then \( \approx c_s H \) (Pringle 1981). In practice much of the physics of accretion discs depends only on low powers of \( \alpha \), allowing simple insights by taking \( \alpha \) as a global constant.

From (1) we can write down the accretion timescale \( t_{\text{visc}} = R^2 / \nu \) for a disc as

\[ t_{\text{visc}} \approx 10^{10} \text{ yrs} \left( \frac{\alpha}{0.1} \right)^{-1} \left( \frac{H/R}{10^{-3}} \right)^{-2} \left( \frac{M}{M_8} \right)^{-1/2} \left( \frac{R}{1 \text{ pc}} \right)^{3/2}. \]  
\[ (2) \]

This is evaluated here for typical active galactic nuclei (AGN) disc parameters: \( \alpha \sim 0.1 \) from King et al (2007), and \( H/R \sim 10^{-3} \) from e.g. Collin-Souffrin and Dumont (1990). This form already allows an important conclusion about AGN accretion. From our remarks above, this is how SMBH grow, and so must occur on timescales significantly shorter than the age of the universe. Requiring \( t_{\text{visc}} \) to be shorter than a Hubble time, we see that the disc scale radius must be \( R \ll 1 \text{ pc} \).

We reach a similar conclusion by considering the effects of self-gravity on an AGN disc. At a radius of only \( \sim 0.01 \text{ pc} \) King et al (2008) (see also Levin 2007, Goodman 2003) find that discs become gravitationally unstable. For SMBH this instability is catastrophic and results in a complete fragmentation. Most of the disc gas forms stars, starving the inner disc (Goodman 2003, Levin 2007, King et al 2008). This suggests that discs feeding SMBH must either be fed at mass rates many orders of magnitude below Eddington, or form as small scale ‘shots’. Observations of Eddington accretion suggest the latter.

These simple arguments already point to a new picture for SMBH accretion which we discuss here.

2. SMBH accretion

Astrophysical black holes have two important parameters: mass and angular momentum (spin). The spin controls the specific binding energy of matter accreting on to the hole, and so the accretion luminosity for a given black hole mass. Accretion grows the SMBH mass, but also affects its spin.

The huge disparity in scales between the black hole (\( R_g = 5 \times 10^{-6} M_8 \text{ pc} \)) and the galaxy (\( \sim 10–100 \text{ kpc} \)) strongly suggests that the angular momenta of the black hole spin and of the gas trying to accrete on to it cannot be correlated in direction, at least initially. In particular the orbital plane of most of the disc mass is unlikely to be aligned with the SMBH spin plane. (This is the first example of several we shall encounter where strong symmetry assumptions (here alignment), originally made for simplicity, turn out to have a distorting effect in suppressing various important effects.)

The first work on the behaviour of misaligned discs was by Bardeen and Petterson (1975). This was actually in the context of stellar-mass black hole binary discs where for example
an asymmetric supernova kick may have significantly misaligned the black hole spin and the binary orbit (e.g. Roberts 1974). The physics here is set by the Lense–Thirring effect. Lense and Thirring (1918) showed that the dragging of inertial frames makes misaligned test particle orbits precess around the angular momentum of a gravitating body at a rate

$$\Omega_p = \frac{2GJ_h}{c^2R^3},$$

where $J_h$ is the angular momentum (here of the black hole). This precession is strongly differential—much faster for gas close to the black hole. The precession time is

$$t_{LT} = \frac{1}{|\Omega_p|} = \frac{c^2R^3}{2GJ_h} = \frac{1}{2a} \left( \frac{R}{R_g} \right)^2 \frac{R}{c}$$

where $J_h = |J_h| = aGM^2/c$ (Kumar and Pringle 1985). From (4) it is easy to see that for gas orbiting close to the black hole horizon ($R \approx R_g$) the precession time can be as short as the dynamical time (at which point the orbit is no longer near-circular). However, the precession is strongly dependent on radius, so for gas far from the black hole the precession is entirely negligible:

$$t_{LT} = 6.5 \times 10^{10} a^{-1} \left( \frac{M}{10^8 M_\odot} \right)^{-2} \left( \frac{R}{1 \text{ pc}} \right)^3 \text{ yr.}$$

Thus on scales $\gtrsim 1$ pc, precession induced by the SMBH can be ignored.

Bardeen and Petterson (1975) considered a disc of gas subject to a strongly differential precession of this type (modelled in Newtonian gravity, as everything we shall consider). These authors suggested that dissipation in the disc between the differentially precessing rings causes the disc to align to the black hole spin, more quickly in the centre than in the outer parts. So after some time the inner disc is aligned, the middle disc is warped and the outer disc is still misaligned. For an isolated disc-hole system, the warp propagates outwards until the entire disc is aligned. Later, Papaloizou and Pringle (1983) showed that these early investigations into warped discs (e.g. Petterson 1977, 1978, Hatchett et al 1981) did not properly handle the internal fluid dynamics in a warped disc. In particular their equations did not fully conserve angular momentum, although the conclusions of Bardeen and Petterson (1975) still hold qualitatively. Papaloizou and Pringle (1983) investigated the Navier–Stokes equations for a warped disc in the linear (small warp) regime. They discovered that the warp can propagate in two distinct ways. In the first of these, viscosity dominates ($\alpha > H/R$), and things behave diffusively, as in Bardeen and Petterson (1975). The second mode of propagation is wave-like, and pressure dominates ($H/R > \alpha$). For black hole discs, the diffusive mode is expected to dominate as the discs are generally thin ($H/R \sim 10^{-3}$) and viscous ($\alpha \approx 0.1$). We therefore focus on this case. For wave-like discs see e.g. Papaloizou and Lin (1995), Lubow and Ogilvie (2000) and Lubow et al (2002).

Global solutions in full 3D hydrodynamics were hard to achieve, so to make progress Pringle (1992) used conservation equations to derive an equation governing the evolution of a twisted disc composed of circular rings interacting viscously. Using the equations of Pringle (1992) and Scheuer and Feiler (1996) calculated the secular evolution of warped discs. A subtle error in their calculation led them to conclude that discs always coaligned with the black hole spin—even if they began close to counteralignment. This led to the conclusion that all black hole discs align on timescales $\sim a^2/t_{visc}$. As $\alpha < 1$, with observations suggesting $\alpha \approx 0.1$ (King et al 2007), this disc–BH coalignment occurred long before significant accretion could take place. This largely removed the motivation for studying misaligned or warped discs.

The alignment error had serious consequences: since alignment would always occur rapidly whatever the initial orientation of disc and hole spin, SMBH would always gain almost
all their mass from prograde accretion discs, thus spinning up to near-maximal values after doubling their masses. This in turn made the specific binding energy release maximal (about 40% of rest-mass energy). Since radiation pressure inhibits accretion (via the Eddington limit), this implied that only rather low mass accretion rates were possible. The discovery of SMBH with masses $\gtrsim 10^8 M_\odot$ at redshifts $z \gtrsim 6$ (allowing accretion only for a timescale of order $10^9$ yr) then appeared to require that these holes must have started from initial ‘seeds’ with significant masses $\gtrsim 10^6 M_\odot$.

Almost a decade passed (with the belief in rapid coalignment well entrenched in SMBH models) before King et al (2005) discovered the error in Scheuer and Feiler (1996)—the implicit assumption of infinite disc angular momentum. With this restriction lifted King et al (2005) showed that counteralignment simply requires

$$\cos \theta < -\frac{J_d}{2J_h}$$

where $\theta$ is the angle between the disc and hole angular momentum vectors with magnitude $J_d$ and $J_h$ respectively. This condition ensures that the total angular momentum vector (sum of disc and hole) is shorter than the hole angular momentum. Since the latter is subject only to precessions, its length cannot change during the alignment process, meaning that hole and disc must end up opposed to account for the shorter total angular momentum.

With retrograde discs now plausible—indeed likely, King and Pringle (2006, 2007) began to explore the effect of misaligned discs on the evolution of SMBH. The larger lever-arm of retrograde accretion flows near the black hole horizon offers an obvious way of keeping their spins low, and so making SMBH growth efficient, particularly if accretion disc events have no preferred direction with respect to the host galaxy. This scenario (often called ‘chaotic accretion’; King and Pringle 2006, 2007) is consistent with many observational effects, such as the observed lack of correlation of the directions of AGN jets (thought to be orthogonal to the plane of the accretion disc close to the hole) with large-scale properties of the host (King et al 2008). There was renewed interest in warp propagation through accretion discs. Lodato and Pringle (2006) used the numerical method of Pringle (1992) to explore the evolution predicted by King et al (2005), confirming that the criterion (6) correctly determines the alignment history (co, counter, or more complex) of accreting black holes.

In the chaotic accretion picture, a natural question is what happens when a misaligned accretion event occurs on to a pre-existing disc. Nixon et al (2012a) explored this with numerical simulations. As partially opposed gas flows interact through viscous spreading, significant amounts of angular momentum can be cancelled, leading to gas infall and strong accretion.

Perhaps surprisingly, this cancellation can occur even in a single disc event, as sufficiently inclined discs may change their planes almost discontinuously, rather than in the smooth warps envisaged by BP (Nixon and King 2012). There had been sporadic evidence that discs could ‘break’ in this way (Larwood et al 1996, Fragner and Nelson 2010, Lodato and Price 2010). The theoretical possibility of this idea was further studied by Ogilvie (1999); for a locally isotropic viscosity coefficient, he showed that the stresses in a strongly warped disc would evolve in such a way that the forces trying to bring the disc back into a single plane would actually weaken as the warp grew.

The first systematic investigation of this possibility was by Nixon and King (2012), who used the Pringle (1992) method to explore warp propagation with the full effective viscosity coefficients derived by Ogilvie (1999). This revealed modified Bardeen–Petterson (BP) behaviour where a sharp break in the disc occurs between the aligned inner disc and the misaligned outer disc. However, the method of Pringle (1992) forces the disc to respond viscously (following a diffusion equation) to the Lense–Thirring precession and excludes

$$\text{Eq. (6)}$$
other possible hydrodynamical effects. The sharp disc break found in Nixon and King (2012) suggested that this assumption was too restrictive, and that a full 3D hydro numerical approach was needed.

Accordingly Nixon et al (2012b) made smoothed particle hydrodynamics (SPH) simulations of misaligned discs around a spinning black hole. SPH was already known to reproduce the behaviour derived by Ogilvie in modelling the communication of a warp in a fluid disc (Lodato and Price 2010), and therefore suitable for modelling such discs. Nixon et al (2012b) confirmed the expected BP behaviour of low-inclination discs, but showed that misaligned discs can break. Further, a break can promote cancellation of disc angular momentum. Separated disc annuli precess independently of each other, and so inevitably become partially opposed. In effect each annulus borrows angular momentum from the central black hole to achieve the cancellation, so that in some way the black hole is complicit in feeding itself more rapidly than simple viscous evolution would allow.

This tearing behaviour differs radically from smooth BP evolution, and we are only beginning to understand its implications. Figure 1 shows the 3D disc structure for a small and large inclination disc around a spinning black hole (Nixon et al 2012b).
Importantly, it is clear that although tearing was first noted for accretion on to spinning black holes, the crucial element is that disc orbits precess differentially. This always happens if the effective potential for the disc flow has a quadrupole component. We can therefore expect tearing to occur quite generically for disc-like flows on all scales, not merely around spinning black holes, but also around SMBH binaries (see below) and on the scale of an entire galaxy.

The potential objection to the ideas of breaking and tearing is that current treatments rely on a scalar viscosity (stress proportional to strain), whereas a more general tensor relation might conspire to hold the vertical disc structure together. Intuitively this seems unlikely for the currently-favoured MRI picture: where the ‘horizontal’ viscosity driving angular momentum transport and accretion is constantly pumped by azimuthal winding up of magnetic fields, the vertical relative motion of the two sides of a warp is periodic and bounded, suggesting that if anything an MRI viscosity might be still weaker in the vertical direction. Current MRI simulations are very far from being able to answer such questions, not least because of the unphysical diffusive effects of having a disc plane inclined with respect to the grid symmetries.

3. SMBH binaries

Most galaxies have SMBH, and galaxy mergers are common. Dynamical friction drives the SMBHs of a merging galaxy pair close together in the centre of the merged galaxy: binary angular momentum is lost to surrounding stars. This forms an SMBH binary with a typical separation of order 1 pc, but cannot coalesce the holes, since all the suitable stars have been driven away by the dynamical friction process itself. Since very few SMBH binaries are actually observed, some as yet unknown process must drive the binary to coalescence. This is the last parsec problem (Begelman et al 1980, Milosavljević and Merritt 2001).

There have been various suggested solutions, generally of two types. One type invokes collisionless matter to arrange that stellar orbits passing close to the SMBH binary are constantly refilled, and so available to remove it angular momentum. Recent attempts along these lines have met with some success, e.g. with triaxial dark matter haloes refilling the stellar orbits (Berczik et al 2006), or using the Kozai mechanism in triple SMBH systems (Blaes et al 2002), or non-axisymmetric potentials (Iwasawa et al 2011). However it seems clear that at some level gas must play a role, and attempts to invoke it constitute the second type of proposed solution of the last parsec problem. Shocks in the galaxy gas flows can rob the gas of the angular momentum supporting its orbit so that it falls towards the binary with a random inclination. Quite separately, to observe SMBH binaries in action we need to understand how they interact with gas.

All early studies of gas interacting with an SMBH binary were, as in section 2, limited to discs which were both coplanar and prograde w.r.t. the binary. But then two physical effects severely limit the effect on binary evolution. First, the disc mass is limited by self-gravity to

$$M_d \lesssim \frac{H}{R} M_\text{vir},$$

where $H/R \ll 1$. So any disc with enough mass to affect the binary quickly fragments into stars. Second, an infinite family of prograde disc orbits are resonant with the binary rotation, holding the gas far out. In principle this shrinks the binary by transferring angular momentum to the gas, but much too slowly to be helpful. Rather as in the case of dynamical friction, the effect which removes angular momentum from the binary tends to weaken itself by gradually pushing away the agency (here the prograde disc gas) which effects this removal in the first place.

But there is no compelling reason to suppose that the circumbinary gas disc is prograde. And the reasoning of the last paragraph strongly suggests that things would work much better
with retrograde discs, as happened with SMBH growth in section 2. Retrograde discs do not suffer resonances (Papaloizou and Pringle 1977), so disc gas can instead accrete freely on to the binary with negative angular momentum. Nixon et al (2011a) show that once a retrograde moving mass $\sim M_2$ has interacted with the binary, its eccentricity approaches unity (here ‘interacted’ means gravitationally—the gas need not accrete for example). Even before this happens, gravitational wave losses complete the SMBH coalescence. So the timescale for this phase of the merger is given by $\sim M_2/\dot{M}$, where $\dot{M}$ is the accretion rate through the retrograde disc.

However, (7) tells us that any individual accretion event cannot have a mass $\gtrsim M_2$ unless the mass ratio is very small ($\lesssim H/R$). Therefore we must consider multiple, randomly-oriented events. As prefigured at the end of the last section, Nixon et al (2011b) showed that much of the disc alignment and breaking phenomena derived for a disc around a single spinning black hole hold for a circumbinary disc. Here the quadrupole part of the binary potential induces a similar (but stronger) precession to the Lense–Thirring effect. In particular Nixon et al (2011b) conclude that counter-alignment of the circumbinary disc occurs if and only if

$$\cos \theta < -\frac{J_d}{2J_b}$$

exactly as (6) for the black hole case. Nixon (2012) showed that counteralignment of a circumbinary disc is stable (contrary to previous reports), and that for any reasonable set of parameters the binary dominates the angular momentum of the system. So approximately half of all randomly oriented accretion events are retrograde.

All this means that a sequence of $\sim 2qR/H$ randomly oriented accretion events, each limited by self-gravity, can drive the binary eccentricity close to unity and allow gravitational wave losses to complete the black hole merger. But there is a subtlety here. If we assume that collisionless processes can drive the binary in to 0.1 pc, the viscous time (appropriate for retrograde circumbinary discs) is

$$t_{\text{visc}} \approx 4.5 \times 10^8 \text{ yr} \left( \frac{\alpha}{0.1} \right)^{-1} \left( \frac{H/R}{10^{-2}} \right)^{-2} \left( \frac{M}{10^8 M_\odot} \right)^{-1/2} \left( \frac{R}{0.1 \text{ pc}} \right)^{3/2}. \quad (9)$$

This time is so long that at first glance it appears unlikely that gas accretion can help on these scales. The merger time is given by

$$t_{\text{merge}} \sim 2qR/H t_{\text{visc}} \quad (10)$$

where taking $t_{\text{visc}}$ from (9) gives an upper limit. For typical numbers, $qR/H \approx 10 - 100$, and so (10) would be longer than a Hubble time.

However, Nixon et al (2013) recently showed that misaligned circumbinary discs can tear and cancel angular momentum (see figure 2), and therefore the accretion rates may be increased by factors up to $10^4$ for sustained periods (see figure 3). So depending on the mass supply from the galaxy, it is possible that the gas-driven merger timescale can be as short as $\sim 10^6 - 10^7$ yr from 0.1 pc, and considerably shorter still if collisionless dynamics puts the starting point even further in.

It appears then that collisionless (e.g. Berczik et al 2006, Khan et al 2011) and collisional processes (e.g. those considered here) conspire together to solve the last parsec problem with collisionless processes driving the binary to the scales ($\lesssim 0.1$ pc) where gas can have a strong effect. Then, if collisionless processes can go no further, gas completes the merger provided the galaxy can supply enough mass ($\sim M_2$) on the right (chaotic) orbits.
Figure 2. The 3D structure of the $\theta = 60^\circ$ circumbinary disc from Nixon et al (2013).

Figure 3. Accretion rates for the circumbinary disc simulations in Nixon et al (2013). For tearing discs the accretion rate can be boosted by factors up to $10^4$ compared to the equivalent prograde, aligned disc.

4. Conclusion

The common thread in most of the work reported here is the gradual removal of assumptions of regularity and symmetry of gas flows near supermassive black holes (SMBH).
These assumptions included those of disc flows aligned with the spin plane of a single black hole, or the orbital plane of an SMBH binary, as well as prograde rotation in most cases. Although probably initially made on grounds of simplicity, these assumptions have little basis in reality. We have seen that they arbitrarily rule out a rich variety of phenomena and actually create artificial difficulties in many cases. In particular by keeping the sense of rotation of everything parallel, they make angular momentum barriers to infall and mergers formidable. The central regions of galaxies are not in general strongly coherently rotating, so cancellations between opposed gas flows must be a common phenomenon. So we have seen that in general SMBH do not inevitably have high spins—indeed if accretion events have no preferred direction there is a slow but persistent statistical trend towards lower spins (King et al 2008), something that was already known to be the outcome of repeated SMBH coalescences (Hughes and Blandford 2003). Several problems are greatly eased by this—we have already noted that the Eddington limit of a slowly spinning hole is much less of a barrier to black hole mass growth, and in addition note that low spins make strongly anisotropic gravitational-wave reaction unlikely, allowing galaxies to retain the merged black holes. For completeness we should add that the same physics makes the often-invoked idea of jet precessions seem unlikely (Nixon and King 2013). This paper also points out that SMBH spins barely move under the effect of individual accretion events, and instead perform very slow random walks in direction.

Perhaps the most spectacular consequence of removing symmetry assumptions is that disc flows may generically be subject to breaking and tearing (Nixon and King 2012, Nixon et al 2012b, Nixon et al 2013). By temporarily borrowing angular momentum from the source of the potential, gas rings arrange to cancel it with neighbours to allow accretion to occur on near-dynamical timescales. This may potentially add considerable new insight to our current picture of many gas-dynamical effects in astrophysics.

Acknowledgments

Research in theoretical astrophysics at Leicester is supported by an STFC Consolidated Grant. CN acknowledges support for this work, provided by NASA through the Einstein Fellowship Program, grant PF2-130098.

References

Balbus S A and Hawley J F 1991 Astrophys. J. 376 214–33
Bardeen J M and Petterson J A 1975 Astrophys. J. Lett. 195 L65
Begelman M C, Blandford R D and Rees M J 1980 Nature 287 307–9
Berczik P, Merritt D, Spurzem R and Bischof H P 2006 Astrophys. J. Lett. 642 L21–24
Blaes O, Lee M H and Socrates A 2002 Astrophys. J. 578 775–86
Collin-Souffrin S and Dumont A M 1990 Astron. Astrophys. 229 292–328
Fragner M M and Nelson R P 2010 Astron. Astrophys. 511 A77
Goodman J 2003 Mon. Not. R. Astron. Soc. 339 937–48
Hatchett S P, Begelman M C and Sarazin C L 1981 Astrophys. J. 247 677–85
Hughes S A and Blandford R D 2003 Astrophys. J. Lett. 585 L101–4
Iwasawa M, An S, Matsubayashi T, Funato Y and Makino J 2011 Astrophys. J. Lett. 731 L9
Khan F M, Just A and Merritt D 2011 Astrophys. J. 732 89
King A R, Lubow S H, Ogilvie G I and Pringle J E 2005 Mon. Not. R. Astron. Soc. 363 49–56
King A R and Pringle J E 2006 Mon. Not. R. Astron. Soc. 373 L90–92
King A R and Pringle J E 2007 Mon. Not. R. Astron. Soc. 377 L25–28
King A R, Pringle J E and Hofmann J A 2008 Mon. Not. R. Astron. Soc. 385 1621–7
King A R, Pringle J E and Livio M 2007 Mon. Not. R. Astron. Soc. 376 1740–6
Kumar S and Pringle J E 1985 Mon. Not. R. Astron. Soc. 213 435–42
Larwood J D, Nelson R P, Papaloizou J C B and Terquem C 1996 Mon. Not. R. Astron. Soc. 282 597–613
