LETTERS

Massive black hole binaries in active galactic nuclei

M. C. Begelman*, R. D. Blandford† & M. J. Rees‡

* Department of Astronomy, University of California, Berkeley, California 94720
† Theoretical Astrophysics, California Institute of Technology, Pasadena, California 91125
‡ Institute of Astronomy, Madingley Road, University of Cambridge, Cambridge CB3 0HA, UK

Most theoretical discussions of active galactic nuclei (including quasars) attribute their energy production either to an accreting black hole or to a precursor stage—for instance a dense star cluster or a supermassive star—whose inevitable end point is a massive black hole[1]. We explore here the possibility that some active nuclei may contain two massive black holes in orbit about each other. This hypothesis suggests a new interpretation for the observed bending[2] and apparent precession[3] of radio jets emerging from these objects and may indeed be verified through detection of the direct consequences of orbital motion.

There are straightforward reasons for surmising that supermassive binaries exist[4,5]: mergers between galaxies appear to be frequent; cD galaxies in clusters and small groups were quite probably formed in this manner[6–9]; and there is direct evidence that the nearby active galaxy Centaurus A is a merger product[9]. If most galaxies contain massive black holes in their nuclei—relics of earlier activity—then they will settle, under dynamical friction, in the core of a merged stellar system. Many cD galaxies do indeed contain double or multiple cores that may have been formed in this way[10]. There are other scenarios for binary black hole formation. For instance, a rotating super-massive star may undergo bar-mode instability and fission into two components that both collapse[1].

The dynamical evolution of a pair of massive black holes in the violently relaxed core of a newly merged galaxy proceeds through several stages.

(1) The black holes are probably surrounded by dense stellar clusters; that within the less massive galaxy should be denser and will survive intact until the nuclei merge. Dynamical friction is extremely efficient[6] and the cores merge and undergo violent relaxation in a characteristic galactic dynamical time $t_{gal} \sim 10^8$ yr.

(2) The captured hole sinks toward the centre of the stellar distribution on the dynamical friction time scale:

$$t_{df} \sim \frac{6\times10^6}{\log N}\left(\frac{v_c}{300\ \mathrm{km\ s^{-1}}}\right)\left(\frac{r_c}{100\ \mathrm{pc}}\right)^2 m_8^{-1} \quad \mathrm{yr} \qquad (1)$$

where the core is presumed to have a central velocity dispersion v_c, a core radius r_c and contain N stars, mass m^*. m_8 is the mass of the smaller black hole, in units of $10^8\,M_\odot$, presumed to be associated with the smaller galaxy. M is the mass of the larger black hole, assumed to be less than Nm^* (the mass of the core). t_{df} is generally longer than the core crossing time and so the merged cores should be fairly relaxed.

(3) When the black hole separation r shrinks to a value $r_B \sim (M/Nm_*)^{1/3} r_c$, they will become bound to each other with an orbital velocity that is initially smaller than the velocity dispersion of the stars by a factor $(M/Nm_*)^{1/3}$.

(4) The holes continue to spiral together under dynamical friction. As the binary becomes 'harder', the maximum impact parameter for effective energy transfer diminishes as $(r/r_B)^{3/2} r_c$ because a binary cannot lose energy effectively to a field star during an encounter which lasts longer than one orbital period. For longer encounter times, the orbital energy is adiabatically invariant and the energy transfer diminishes exponentially[11]. This affects t_{df} only through the logarithm as long as the binding energy of the binary is small compared with the total kinetic energy of all the stars whose orbits pass sufficiently close to interact effectively.

(5) When $r \lesssim r_{lc} = (m/M)^{1/4}(r_B/r_c)^{9/4} r_c$, this condition is violated and further stellar dynamical evolution is controlled by the rate at which stars whose orbits do not come close to the binary can diffuse into more radial 'loss cone' orbits[12] that pass within a radius $\sim r$. Two-body 'star–star' relaxation will repopulate the loss cone in a time ~(solid angle of loss cone) × (two-body relaxation time). The loss cone must be repopulated $(r_{lc}/r)^4$ times to halve the binary orbit. The relevant binary evolution time is then:

$$t_{lc} \sim t_{df} \max\left[1, \left(\frac{m}{M}\right)^{7/4} N \left(\frac{r_B}{r_c}\right)^{27/4} \left(\frac{r_{lc}}{r}\right)\right] \qquad (2)$$

Collective effects may, however, lead to a more rapid repopulation of the loss cone[13] and evolution on a time closer to t_{df}.

(6) The binary becomes 'hard' when its orbital speed reaches v_c at a separation $r_h \sim (r_B/r_c)^3 r_c$. The maximum impact parameter for dynamical friction is now $(rr_h)^{1/2}$. Each interacting star on average carries off a fraction $\sim m_* m/M^2$ of the binary's binding energy. The binary evolution time scale is now

$$t_h \sim t_{df} \max\left[\left(\frac{r_h}{r}\right) \ln N, \left(\frac{M}{Nm^*}\right)^2 N\right] \qquad (3)$$

(7) When the binary becomes sufficiently tightly bound, gravitational radiation shrinks the orbit on a time scale $|r/\dot{r}|$ given by

$$t_{GR} \sim 3\times10^5 \left(\frac{m}{M}\right)^{-1} M_8^{-3} r_{16}^4 \quad \mathrm{yr} \qquad (4)$$

(8) The evolution now proceeds quite rapidly to coalescence (which takes a time $\sim M_8$ h). This may be accompanied by a powerful burst of electromagnetic radiation, depending on the local gas density and magnetic field strength, as well as a pulse of gravitational radiation[4,5]. Indeed, the final plunge will lead to the emission of net linear momentum and the recoil may be sufficient to eject the combined hole from the core and hence from the galaxy[5].

In addition to these stellar dynamical effects, infall of gas onto the binary can also lead to some orbital evolution. Gas may be flung out of the system, acquiring energy (and angular momentum) at the expense of the binary; alternatively, gas may accrete onto the larger hole, causing orbital contraction as the product Mr is adiabatically invariant. In either case, the evolution time scale is

$$t_{gas} \sim 10^8 M_8 (\dot{M}/1 M_\odot\ \mathrm{yr}^{-1})^{-1} \quad \mathrm{yr} \qquad (5)$$

During phases of nuclear activity, gas infall rates $\dot{M}$ of 1–$10\,M_\odot\,\mathrm{yr}^{-1}$ are required for times 10^7–10^8 yr. It is, therefore, quite likely that this process controls the binary evolution when r just exceeds r_{GR} and the stellar dynamical effects are weakest.

These time scales, $|r/\dot{r}|$, are given schematically in Fig. 1. It is clear that the longest evolutionary phase, and therefore the one containing most potentially observable binaries, is that with $r_h \gtrsim r \gtrsim r_{GR}$. During this phase a tight binary will persist in a quiescent state with almost constant r, except when it is undergoing fuelling and it becomes a conspicuous source of non-thermal power. The energy dissipated by the material flung out (at the expense of the binary orbital energy) is generally negligible at this time[14]. Slow secular contraction proceeds during active phases and pauses during intervals of quiescence. The details and statistics of observability depend on the time-dependence of $\dot{M}$ (is it intermittent or recurrent?, are the active phases as long as $M/\dot{M}$?). In conditions appropriate to galaxies like M87 ($r_c \sim 100$ pc, $v_c \sim 300$ km s^{-1}, $Nm^* \sim 3\times10^9\,M_\odot$) it is

0028–0836/80/390307—03$01.00

clear that the two-body evolution times given by equations (2) and (3) are much longer than t_{gas}. (This may not be true for much less massive cores and small holes, $M \leq 10^6 M_\odot$.) If the binary gains mass at a constant rate, then the probability of observing a binary with separation greater than r will fall as r^{-1}. If $M \propto M$, then binaries are distributed uniformly in log r. In any case, the separation r_{GR} at which gravitational radiation becomes dominant is insensitive to the values of t_h and t_{gas},

$$r_{GR} = 5\times 10^{16}\left(\frac{m}{M}\right)^{1/4} M_8^{3/4} \times\left[\frac{\min(t_h, t_{gas})}{10^8\ \mathrm{yr}}\right]^{1/4} \mathrm{cm} \quad (6)$$

For $r \leq r_{GR}$ the orbital evolution is very rapid and the probability of observing a binary with separation smaller than r will decline as r^4. Most observed binaries will have $r_{GR} \lesssim r_h$.

The most direct evidence for a massive black hole binary would be some manifestation of the keplerian motion, which has a period

$$P_{orb} \sim 1.6 r_{16}^{3/2} M_8^{-1/2} \quad \mathrm{yr} \quad (7)$$

Any eccentricity in the orbit will almost certainly be erased by the frictional drag and so the orbit is assumed to be circular. We generally expect the spin axes of the holes to be misaligned unless thay have both accreted a substantial amount of gas with a fixed angular momentum[15]. The spin axes will then undergo geodetic precession about the total angular momentum. For the more massive hole, the period is,

$$P_{prec} \sim 600 r_{16}^{5/2} (M/m) M_8^{-3/2} \quad \mathrm{yr} \quad (8)$$

The spin angular momentum will be small compared with that in the binary provided that $(m/M) \geq (GM/rc^2)^{1/2}$ and so the spins precess in a cone with the orbital angular momentum as axis.

The beams of plasma which give rise to compact and extended radio structure in active galaxies may be collimated sufficiently near a massive black hole that the Lense–Thirring effect aligns them with the hole's axis[15], irrespective of the original momentum of the inflowing gas. (This requires the collimation scale to be much smaller than the binary separation r in the phases that concern us here.) The beams will then process with period P_{prec}. It has been suggested that a scaled-down version of this mechanism causes a pair of jets to precess in SS433 (refs 16, 17).

As (r_h/r_{GR}) can be as large as 10^4, P_{orb} and P_{prec} can span many orders of magnitude and be made manifest in different ways. We outline below two examples that could correspond to different phases in the evolution of a binary. The observed activity is assumed to be only associated with the more massive component.

Wide binary ($r_{16} \sim 30$, $M_8 \sim 1$, $m_8 \sim 0.1$). In this case the precession time scale is $\sim 3\times 10^7$ yr—comparable with the inferred lifetime of extended radio components. Precession of the beam axis would then result in an inversion symmetry in the source structure. Such precession has been inferred[3] on phenomenological grounds, but the massive binary seems to be the only way to set this in a dynamical context. (Jet alignment can occur, however, with a lone black hole[15].) Note that the morphology may differ from that associated with a ballistic trajectory, because of gas-dynamical effects, but should nevertheless have the same average symmetry (M.C.B., in preparation).

The orbital period (300 yr) would be too long to be detected directly. However, the broad emission lines may originate in a region around the more massive hole while the narrow lines would come from a much larger system enveloping the binary. The broad lines may then peak at a velocity up to $\sim 10{,}000\ r_{16}^{-1/2} M_8^{-1/2} m_8$ km s^{-1} different from that of the galaxy as determined by the narrow lines. Blue- and redshifts (relative to the narrow lines) should be about equally common.

Close binary ($r \sim r_{GR}$, $t_{gas} \sim 10^8$ yr). In this case $P_{orb} \sim 18(m/M)^{3/8} M_8^{5/8}$ yr, $P_{prec} \sim 3\times 10^4 (M/m)^{3/8} M_8^{3/8}$ yr. Any extended radio structure associated with jets, built up over many

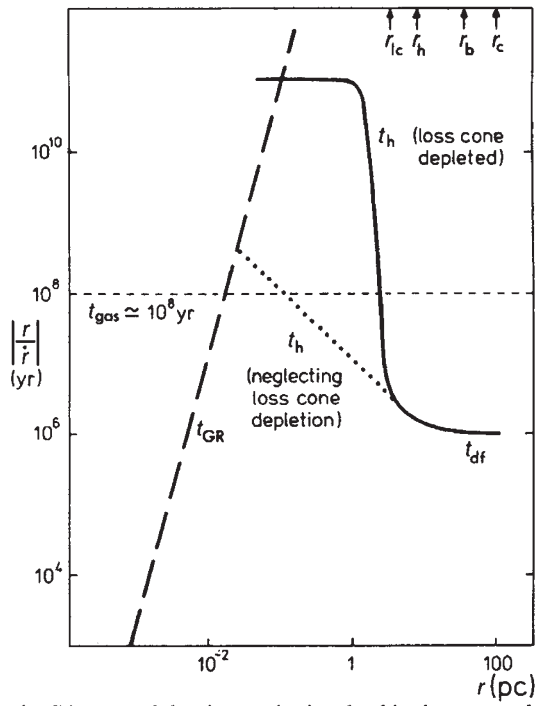

Fig. 1 Diagram of the time scales involved in the approach, and eventual coalescence, of a supermassive binary. The core parameters chosen are those that might be appropriate to a giant elliptical: $r_c \sim 100$ pc, $v_c \sim 300$ km s^{-1}, $N \sim 2\times 10^9$ and $m_* \sim 1$. The members of the binary are taken to have masses $M_8 = 1$ and $m_8 = 0.3$. For this system, $t_{df} \sim 10^6$ yr. Within r_h the evolution time scale would be $(r_h/r)t_{df}$ if loss cone depletion could be ignored; however, unless collective effects permit replenishment of the loss cone on much less than the ordinary stellar relaxation time, the evolution within r_{lc} proceeds on a very much longer time scale $\geq 10^{11}$ yr (compare equation (3)). Influx of gas into the system at a rate $\dot{M} \sim 1\ M_\odot\ \mathrm{yr}^{-1}$ would however yield a time scale $\sim 10^8$ yr. Gravitational radiation would take over as the dominant mechanism within $1.3\times 10^{-2}(t_{gas}/10^8\ \mathrm{yr})^{1/4}$ pc. The values of the various critical radii are indicated. (Note that in this particular example r_{lc} lies inside r_h; for the case of two similar masses, as discussed in the text, this ordering is reversed.)

precession periods, will be aligned with the orbital angular momentum; the precession related curvature will be discernible, however, on the scales probed by VLBI or the VLA.

Suppose that the jets precess around a cone of semi-angle ϕ. If the jets were narrow with a non-relativistic speed $V \ll c$, a 'snapshot' would reveal the beam material tracing out a spiral pattern around the core: successive turns would be separated by a fixed distance vP_{prec}, the pattern becoming more transverse at larger distances from the apex. However, in the 'superluminal' sources the jet velocity is believed to be relativistic with Lorentz factor $\gamma = (1 - v^2/c^2)^{-1/2}$ in the range 5–10 (ref. 2). What is actually observed from a precessing jet is modified by Doppler boosting and light travel time effects that depend on the observer's orientation. We can readily show that this geometry can account for a large bending and consequent misalignment between fast-moving VLBI features and any extended structure.

Let the observer's line of sight make an angle Ψ with the cone axis. The jet will be observed to be brightest when it makes an angle $\leq \gamma^{-1}$ with the line of sight. If $\phi \geq \gamma^{-1}$, this implies that for the strongest sources, the observer direction must lie in the hollow cone, $\phi + \gamma^{-1} \geq \Psi \geq \phi - \gamma^{-1}$. Assuming that the emissivity of the jet decreases away from the apex, the source will again appear brightest when the central region points within an angle γ^{-1} at the observer direction, which it will do for a fraction $\sim(2\pi\gamma\phi)^{-1}$ of each precession period. When observed from this direction, a precessing jet will be brightest and appear to bend through an angle ~ 1 radian. Individual inhomogeneities will, however, appear to move rectilinearly. Superluminal expansion of inhomogeneities convected outwards by the jet with apparent

speed $\sim\gamma c$ will occur. Curvature of this order has been reported in a number of superluminal radio sources[2]. For 3C273 and 3C345, continuous bending through an angle $\sim 30°$ is visible on angular scales θ from ~ 5 to ~ 30 m arc s. If the distance of the source is d, we can infer the necessary precession period using the fact that the projected expansion speed is $\sim\gamma c$ to obtain,

$$P_{\rm prec} \sim 2\pi\phi\theta d/c \qquad (9)$$

For 3C273, we substitute $\phi \sim 60°$, $\theta \sim 30$ m arc s, $d = 800$ *Mpc* to obtain $P_{\rm prec} \sim 500$ yr. We could obtain this short precessional period if $r_{16} \sim 1.7$, $M_8 \sim m_8 \sim 1$, $t_{\rm GR} \sim 3 \times 10^6$ yr. More distant sources require somewhat larger and more probable precessional periods. If, as we have assumed, $\phi \gg \gamma^{-1}$, then at large distance from the nucleus the jet will not be subject to large Doppler shifts.

Orbital motion at the source of a jet is another possible source of jet curvature[18]. However, the maximum transverse displacement in a single wavelength is of order the binary separation r, far too small to account for observed bent superluminal sources, though possibly relevant to small scale bends in sources like NGC6251 (ref. 18) which has a jet 1 pc long, misaligned by 3°. This would be expected from a binary with separation $r \sim 10^{17}$ cm.

We have argued that there are good reasons to expect binary black holes in galactic nuclei and that most observed examples will have separations in the range $10^{16}\,{\rm cm} \leqslant r \leqslant 10^{19}$ cm. Wide binaries may be responsible for inversion-symmetric features in large scale jets. Very close binaries may produce bending and misalignment in VLBI jets through the kinematic consequences of precession. Orbital motion may be reflected in the profiles of permitted emission lines; short orbital periods (~ 10 yr) may even be detectable photometrically using archival plate libraries.

If the binary hypothesis is correct, then we expect that jets that appear to be orbiting or precessing should be associated with cD or irregular galaxies in clusters or groups with large numbers of close companions. Furthermore, estimating the rate of merging and making an evolutionary correction we can compute the frequency of black hole coalescence events from objects at large redshifts to be in the range 10^{-3}–$1\ {\rm yr}^{-1}$ (refs. 4, 5). These events (hypernovae) would be accompanied by detectable bursts of electromagnetic and gravitational radiation of powers that can, in principle, approach $c^5 G^{-1} \sim 10^{59}\,{\rm erg\,s^{-1}}$ for a duration of ~ 1 h. However, as with a supernova, the actual peak electromagnetic power will be moderated by the amount of gas surrounding the explosions, and the strength of the gravitational radiation pulse by a fairly small efficiency factor[5].

Of course the evolutionary properties of binaries could be far richer than we have described. For example, drawing an analogy with stellar evolution, if one of the holes possessed a massive extended disk or envelope in which the inflow time is as long as $|r/\dot{r}|$ then mass transfer between the components could be important and may even account for much of the variability associated with quasars. Also, merging may occur more rapidly than evolution of the binary and three or more black holes may collect in the core. In this case a gravitational sling-shot[19] rather than coalescence will probably occur.

The detection of a stellar cusp around a central mass, as has been claimed for M 87 (ref. 20), implies that this mass is probably not a recently merged binary, because if black holes were brought together by dynamical friction they would have scoured out all the stars passing within $r_{\rm h}$ and there should actually be a deficit of stars in this region. The cusp may, however, be restored if a large mass of gas has subsequently been accreted thus repopulating the cusp.

We thank Juhan Frank for helpful comments. M.C.B. and R.D.B. thank the Institute of Astronomy, Cambridge for hospitality and acknowledge support under NSF grants AST77-23069 and AST78-05484 respectively.

Received 12 June; accepted 30 July 1980.

1. Begelman, M. C. & Rees, M. J. *Mon. Not. R. astr. Soc.* **185,** 847 (1978).
2. Readhead, A. C. S., Cohen, M. H., Pearson, T. J. & Wilkinson, P. N. *Nature* **276,** 768 (1978).
3. Ekers, R. D., Fanti, R., Lari, C. & Parma, P. *Nature* **276,** 588 (1978).
4. Thorne, K. S. & Braginsky, V. B. *Astrophys. J. Lett.* **204,** 1 (1976).
5. Blandford, R. D. *Sources of Gravitational Radiation* (ed. Smarr, L.) (Cambridge University Press, 1979).
6. Ostriker, J. P. & Tremaine, S. D. *Astrophys. J. Lett.* **202,** 113 (1975).
7. Ostriker, J. P. & Hausman, M. A. *Astrophys. J. Lett.* **217,** 125 (1977).
8. White, S. D. M., *Mon. Not. R. astr. Soc.* **191,** 1p (1980).
9. Tubbs, A. preprint (1980).
10. Wyndham, J. D. *Astrophys. J.* **144,** 459 (1966).
11. Heggie, D. C., *Mon. Not. R. astr. Soc.* **173,** 729 (1975).
12. Frank, J. & Rees, M. J. *Mon. Not. R. astr. Soc.* **176,** 633 (1978).
13. Lynden-Bell, D. in *Relativity Theory and Astrophysics* Vol. 2 (ed. Ehlers, J.) 131 (American Mathematical Soc., 1967).
14. Collins, G. W. *Astron. Quart.* **1,** 39 (1977).
15. Rees, M. J. *Nature* **275,** 516 (1978).
16. Martin, P. G. & Rees, M. J. *Mon. Not. R. astr. Soc.* **189,** 19p (1979).
17. Begelman, M. C. *et al. Astrophys. J.* **238,** 722 (1980).
18. Cohen, M. H. & Readhead, A. C. S. *Astrophys. J.* **233,** L101 (1979).
19. Saslaw, W. C., Valtonen, M. J. & Aarseth, S. J. *Astrophys. J.* **190,** 253 (1974).
20. Young, P. J. *et al. Astrophys. J.* **221,** 721 (1978).

Spectral broadening measurements of the ionospheres of Jupiter and Saturn

Richard Woo & J. W. Armstrong

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91103

When Pioneers 10 and 11 flew by Jupiter, their 2.3-GHz monochromatic radio signals underwent substantial spectral broadening as they propagated through the planet's ionosphere[1,2]. Although spectral broadening has been observed during radio occultation by the solar corona[3,4], the Pioneer measurements of Jupiter represent the first time it has been observed during radio occultation by a planetary atmosphere. The Pioneer data are also unique in that r.f. spectral broadening, a scattering phenomenon normally associated with strong intensity scintillations ($\sigma_\chi^2 \sim 1$ where σ_χ^2 is the variance of the log-amplitude scintillations), occurred when the intensity scintillations were weak. This unusual opportunity makes it possible to compare the theory for spectral broadening with the more widely used weak intensity scintillation theory[5,6]. We report here that good agreement is found, and present the inferred characteristics of the electron density irregularities for the ionospheres of both Jupiter and Saturn. The Saturn results are based on measurements made during the 1979 Pioneer 11 encounter of Saturn.

The details of the Pioneer 10 radio occultation experiment have been discussed elsewhere[1,2,7]. Figure 1 shows the measured r.f. spectrogram at 2.3 GHz for the lower ionosphere ($\sim$200–500 km altitude region designated region C by Woo and Yang[7], where altitude is determined by reference[2] to the atmospheric level at which the neutral gas has a refractivity of 10), measured during exit occultation which took place at 26 °N latitude when the solar zenith angle was about 81°. The electron density irregularities in this region have been found to be isotropic[7]. The inherent linewidth over a comparable integration time as determined from measurements outside the ionosphere is less than 0.5 Hz, and can therefore be ignored when interpreting Fig. 1. The three-dimensional spatial wave number spectrum of the refractive index fluctuations corresponding to a power-law spectrum can be written as[6,7]

$$\Phi(x,\kappa) = \frac{\Gamma(p-1)}{4\pi^2} \sin\left[\frac{\pi(p-3)}{2}\right] c_{no}^2 \exp\left[\left(\frac{x-L}{a}\right)^2\right] \kappa^{-p} \qquad (1)$$

where Γ is the gamma function, κ is the spatial wave number, p is the spectral index, c_{no} is the structure constant of refractive index fluctuations, x is the wave propagation direction, L is the spacecraft to planetary limb distance, and a is the extent of the region of irregularities in the x direction.

Figure 2 shows a plot of the structure wave function $D(\tau)$ divided by 2 (refs 4–6). For the power-law density spectrum, a log–log plot of $D(\tau)$ against τ gives an estimate of the spectral index directly from the slope of a linear fit to the data. We have

0028–0836/80/390309—03$01.00