QUASAR EVOLUTION AND THE GROWTH OF BLACK HOLES IN THE NUCLEI OF ACTIVE GALAXIES

R. D. BLANDFORD
Theoretical Astrophysics 130-33
California Institute of Technology
Pasadena, California 91125

ABSTRACT. The observed evolutionary behavior of active galactic nuclei is compatible with a model in which black holes form in the nuclei of new-born galaxies and then grow at a rate limited by both radiation pressure and the supply of gas. Individual sources become more luminous with time as long as they are being fueled. However, the rapid decrease in the mean rate of supply of gas causes a strong decline in the space density of active objects. Nearby galaxies should harbor modest size ($\sim 10^6 - 10^8 M_\odot$) black holes. It is suggested that the gas that fuels high redshift quasars is mostly derived from the host galaxy.

1. Introduction

Considerable effort has been devoted to measuring and interpreting the luminosity and redshift distributions of quasars and radio galaxies. (*e.g.* Weedman 1986). It is generally agreed that the density (per unit co-moving volume) of bright quasars and radio sources was larger in the past and that this contrast in space density was less marked for lower power sources. Most interpretations of these evolutionary properties have been empirical in character and have involved the fitting of *ad hoc* functional forms to source counts.

The physical properties of active galactic nuclei (AGN) have also been studied in considerable detail. It is widely believed that nuclear activity is powered by the accretion of gas onto a massive black hole (*e.g.* Rees 1984, Blandford 1987, Begelman, this volume, and references therein). The total power emerging from the neighborhood of a black hole is generally assumed to be limited by the Eddington luminosity, $L_E = Mc^2/t_E$, where M is the mass of the hole and $t_E = \sigma_T c/4\pi G m_P = 5 \times 10^8$yr is the Eddington time. The black hole masses and Eddington limits have increased as the quasars have evolved. Simultaneously, the rate of supply of gas has presumably decreased. These two effects ought to exert a controlling influence on the observed evolution of AGN.

In this contribution I describe a simple physical model of AGN evolution and use it to interpret quasar counts. A preliminary version of this model was outlined in Blandford (1986). Related ideas have been developed independently by Cavaliere and collaborators and have been summarized most recently in Cavaliere *et al.* (1988).

2. Motivation

2.1. RELICT BLACK HOLE MASSES

A deceptively simple calculation (Soltan, 1982, Phinney 1983) furnishes a lower bound on the total black hole mass required to account for the integrated light from quasars. If the observed bolometric flux is F_i from quasar i at redshift z_i, and $N_g \sim 10^{-2}h^{-3}\mathrm{Mpc}^{-3}$ is the local density of bright $(L \gtrsim 0.5L_*)$ galaxies (Efstathiou, Ellis and Peterson 1988), where we adopt an Einstein-De Sitter cosmology of age $t_0 = 6.6h^{-1}\mathrm{Gyr}$, then the minimum relict black hole mass per bright galaxy is

$$\langle M \rangle = \frac{\sum_i F_i(1 + z_i)}{\epsilon_B N_g c^3} \qquad (1)$$

where ϵ_B (assumed constant) is the efficiency of the accretion process for producing blue light (which is unlikely to exceed 0.03). The sum is dominated by $\sim 20^m - 21^m$ quasars at redshifts $z \sim 2$. Making conservative assumption about the bolometric correction, we obtain $< M > \sim 3 \times 10^6 (\epsilon_B/0.03)^{-1} h^{-3} \mathrm{M_\odot}$. (This answer is, of course, independent of beaming). Note that the mass bound increases eightfold if h is decreased from 1 to 0.5. Furthermore, if quasar black hole masses are estimated using the ionization parameter and the broad line widths to measure the size and velocity dispersion respectively, (e.g. Wandel and Petrosian, this volume), then the luminosities of the quasars that dominate the sum in equation (1) are below the Eddington limit and the problem is again exacerbated.

The actual partition of relict black hole mass between galaxies depends upon the duty cycle of activity, $\delta(M,t)$. For example, if a fraction ~ 0.01 of galaxies are active for a whole evolution time scale ~ 1 Gyr at $z \sim 2$, then the relict masses are likely to be ten times larger than if ten times as many nuclei are active with a duty cycle of ten per cent. As explained by Dressler (this volume) high resolution observations of stellar velocity dispersions and rotation velocities are starting to constrain high mass models.

2.2. EDDINGTON-LIMITED ACCRETION

The majority of quasars are radio-quiet and X-ray soft. In addition, as several contributors to this volume have argued, the infra-red emission is probably mostly reprocessed dust emission. Therefore most of the power from these sources is emitted quasi-thermally at the effective temperature of an inner accretion disk. Now, despite all their uncertainties, models of accretion onto black holes generally predict a sensitivity to the parameter $\lambda_B = L_B/L_E$ which measures the ratio of the B luminosity to the Eddington limit. In particular, efficient thermalization of the emergent, radiation may require the accretion rate to be near the Eddington value $\lambda_B \sim 0.1$). (In "Grand Unified AGN models", e.g. Blandford 1984, the non-thermal radio quasars and galaxies are hypothesized to be accreting below the Eddington rate. These objects are identified with higher mass black holes found preferentially in elliptical galaxies. Conversely lower mass holes associated with Eddington-limited accretion are found in spirals.)

Allowing for a bolometric correction, $\lambda_B \sim 0.1$ corresponds to the Eddington limit.(e.g. Begelman this volume). All of this suggests that λ_B has a reasonably

standard value in radio-quiet quasars, which from §2.1 cannot be much less than 0.1.

2.3 FUEL SUPPLY

Although the manner in which gas is supplied to a galactic nucleus in order to fuel activity therein is still a matter of controversy, many of the more serious alternatives involve the intermittent deposition of large parcels of gas, triggered by interactions with neighboring galaxies(*e.g.* Fricke, this volume). The gas is therefore likely to be deposited in the nucleus at rates in excess of the Eddington value. As we know from §2.1 that the accretion rate cannot greatly exceed the Eddington value, this suggests that the flow is limited by the black hole, presumably by radiation pressure at or near the Eddington limit. This "feast or famine" hypothesis is only likely to be relevant at high accretion rates; steady stellar mass loss appears to be adequate to fuel lower luminosity AGN.

2.4 QUASAR LUMINOSITY FUNCTION

As Smith discusses in this volume, Boyle *et al.* (1987), report a prominent break in the quasar luminosity function (with a luminosity which declines $\propto t^{-2.3}$ at constant value of the luminosity function) for $0.5 \lesssim z \lesssim 2.5$. Below the break, the luminosity function is quite flat. The behavior of the luminosity function at high power can only be guessed at this time, but the preponderance of bright quasars among the high redshift discoveries suggests that the luminosity function is also quite flat at these epochs. The current quasar-Seyfert luminosity function is reported not to exhibit such a prominent break (but see Marshall 1987).

The two simplest empirical descriptions of the luminosity function are density evolution (in which the probability that a source is extinguished is luminosity-independent) and luminosity evolution (in which a population of sources fades in such a way that the rate of change of the source magnitude is independent of the magnitude). Note that density evolution does not require that sources have only one epoch of activity. If it did, then rare bright local quasars, like 3C273 would contain extremely massive black holes ($M \gtrsim 3 \times 10^{11} M_\odot$). Much lower masses are predicted if all galactic nuclei have a small probability of being reactivated.

Pure density evolution is clearly incorrect as it over-predicts the local density of weaker sources. However, it could describe the variation of powerful sources. Pure luminosity evolution appears to describe the data at intermediate redshifts but is apparently inconsistent with low z and high z data. However, a combination of luminosity evolution at low power and density evolution at high power might be acceptable.

3. Evolution of the Luminosity Function

3.1 ASSUMPTIONS OF THE MODEL

The foregoing considerations motivate a model derived from two fairly restrictive assumptions, namely that most AGN radiate near the Eddington-limit with constant λ_B and ϵ_B and that the duty cycle of activity is determined by the mean rate

of gas supply through the relation

$$\delta(M,t) = \min\left(1, \frac{\lambda_B < \dot{M} > t_E}{\epsilon_B M}\right) \tag{2}$$

where $\langle \dot{M} \rangle (M,t)$ is the mean accretion rate for holes of mass M.

3.2 NORMALIZED LUMINOSITY FUNCTION

Let $N(M)$ be the density of black holes per unit comoving volume of mass M in interval dM. This density will satisfy the conservation equation

$$\frac{\partial N}{\partial t} + \frac{\partial}{\partial M}\{N\langle \dot{M} \rangle\} = S(t) \tag{3}$$

where S is the source function which can be set proportional to $\delta(M - M_i)$, at a small injection mass M_i. (An alternative equation using luminosity rather than mass as the independent variable has been used by Cavaliere et al. 1985, but this is inappropriate if the sources are abruptly extinguished.) Note that equation (3) is first order and therefore requires holes of a given mass to evolve along a single track in the (M,t) plane. A model incorporating a range of accretion rates at a given mass and time would naturally lead to a second order equation. This is unnecessary at present.

It is convenient to change the independent variable from mass to luminosity $L = \lambda M c^2/t_E$. and the dependent variable to the normalized luminosity function $\Phi(L,t) = \delta M N/N_g$ per bright galaxy per unit $\ln L$, (roughly magnitude).

Equation (3) can be integrated along its characteristics. Our procedure is to specify the luminosity function where it is best measured near the break and to integrate forward and backward in time so as to infer the functional forms of the duty cycle and the birth rate.

3.3 GENERAL FEATURES OF THE SOLUTION

We naturally interpret the break as marking the mass below which holes accrete continuously and above which they are progressively more starved of fuel and consequently less likely to be active. The rate of fuel supply diminishes so rapidly, that the hole masses (and consequently active luminosities) are essentially constant well above the break.

Detailed models adequately reproduce the flat, low power luminosity function. This is because the mass of small holes grows exponentially with time with time constant much shorter than t. The flux in mass space is therefore effectively constant and so the number per logarithmic interval of mass is then also constant. At intermediate redshifts, well above the break, $\Phi \propto L^{-2.6}t^{-6}$. As the break luminosity is roughly constant, we deduce that $\delta \propto< \dot{M} > /M \propto \Phi \propto M^{-2.6}t^{-6}$. The model therefore requires that $< \dot{M} > \propto M^{-1.6}t^{-6}$. Flattening of the luminosity function at $z \gtrsim 3$ can also be reproduced as these sources contain the first holes to form and their total growth times are comparable with the age of the universe

then. The largest masses required are $M \lesssim 3 \times 10^9 M_\odot$. and they stop growing when the fuel supply becomes inadequate. They are only rarely re-activated at low and intermediate redshifts. The low power luminosity function essentially reflects the recent black hole birth rate.

4. Implications

4.1 FUELING RATE

The inferred fueling rate exhibits a dramatic overall decline with time and requires that high mass holes deplete their gas supply before low mass holes. This is not what might be expected if the gas supply is related to the frequency of interactions. Instead it might suggest that the gas is mostly derived from the host galaxy and that large scale mass transfer between galaxies is comparatively rare. However, episodes of galactic accretion could still be *triggered* by interactions.

4.2 RELICT BLACK HOLE MASSES

In this model, black hole mass is used with nearly optimal efficiency to make quasar light and so the predicted relict black hole masses are generally lower than those obtained using some alternative schemes. The frequency of the relict black holes is roughly $(dN/d\ln M)/N_g \sim 10^{-3}$ for holes less massive than those responsible for the break in the luminosity function at $z \sim 2 - 2.5$, *i.e.* with $M \lesssim 10^8 (\lambda_B/0.1)^{-1} h^{-2} M_\odot$. Heavier masses will be much rarer; their frequency will be calculable when the high z luminosity function is measured. The best search sites for large black holes are the nuclei of giant elliptical galaxies associated with energetic extended radio sources.

4.3 GALAXY FORMATION

Most black holes must have been formed comparatively recently, but the existence of highly luminous quasars with $z \gtrsim 4$ requires the formation of massive galaxies at even earlier times, a conclusion which constrains cold dark matter scenarios for galaxy formation (Efstathiou and Rees, 1988). The present model minimizes this constraint as the holes grow rapidly and only require Eddington-limit masses.

4.4 SEYFERT GALAXIES

Neither of the hypotheses need apply to low power Seyfert galaxies. The rates of mass accretion needed could be supplied by stellar collisions in a nuclear star cluster or normal stellar mass loss in the inner parts of a galaxy. Also, there are too few galaxies for low power AGN to be Eddington- limited and little observational evidence that this is occurring. Observations (Dressler, this volume) suggest that the majority of bright galaxies contain holes with masses $\gtrsim 10^6 M_\odot$ and consequently most of them have had periods of activity in the past.

4.5 RADIO SOURCES

We can extend this model to apply to radio source evolution if we further hypothe- size that the luminous extended sources are powered by black holes that have been

spun up in an earlier epoch of mass accretion but are currently starved of gas in the case of radio galaxies and limited in the case of radio quasars (*e.g.* Rees *et al.* 1982). The epoch of peak radio source activity should slightly follow the corresponding quasar epoch. Compact radio sources (including the BLLac objects) are believed to be relativistically beamed, intrinsically weak sources (e.g. Zensus and Pearson 1987) and should therefore evolve less dramatically than the extended sources.

Acknowledgements

References

Blandford, R. D., 1984, *Ann. N. Y. Acad. Sci.,* **422**, 403.

Blandford, R. D., 1986, in *Quasars*, ed V. Kapahi and G. Swarup, (Dordrecht: E. Reidel) p. 359.

Blandford, R. D., 1987, in *300 Years of Gravitation*, ed. S. W. Hawking and W. Isreal, (Cambridge: Cambridge University Press), p. 277.

Boyle, B. J., Fong, R., Shanks, T. and Peterson, B. A., 1987, *Mon. Not. R. astr. Soc.,* **227**, 717.

Cavaliere, A., Giallongo, E., Padovani, P. and Vagnetti, F., 1988, in *Optical Surveys for Quasars* (in press).

Cavaliere, A., Giallongo, E. and Vagnetti, F., 1985, *Astrophys. J.,* **296**, 402.

Efstathiou, G., Ellis, R.S., and Peterson, B.A., 1988, *Mon. Not. R. astr. Soc.,* **232**, 431.

Efstathiou, G. and Rees, M.J., 1988, *Mon. Not. R. astr. Soc.,* **230**, 5P.

Marshall, H.L, 1987, *Astr. J.,* **94**, 638.

Phinney, E. S., 1983, Unpublished PhD thesis, University of Cambridge.

Rees, M.J., 1984, *Ann. Rev. Astr. Astrophys.,* **22**, 471.

Rees, M. J., Begelman, M. C., Blandford, R. D., and Phinney, E. S., 1982, *Nature,* **195**, 17.

Sołtan, A., 1982, *Mon. Not. R. astr. Soc.,* **200**, 115.

Weedman, D. W., 1986, *Quasar Astronomy* (Cambridge: Cambridge University Press).

Zensus, A., and Pearson, T.J., editors, 1987, *Superluminal Radio Sources*, (Cambridge: Cambridge University Press).

DISCUSSION

   MARSHALL (A comment) I have shown that the best power law luminosity functions that fit quasar data (Koo and Kron, $Ap.$ $J.$, **325**, 92, 1988; Boyle et $al.$, $M.N.R.A.S.$, **227**, 717, 1987) can also fit the Seyfert galaxy data (Marshall, $A.J.$, **94**, 628, 1987). In particular, the data cannot be fitted by a simple power-law luminosity function.

   WANDEL Commenting on your suggestion that high redshift objects should be accreting (and radiating) close to the Eddington value, while low-luminosity ones should have a low Eddington ratio – we find this behavior in the masses and accretion rates we derive from fitting the UV spectrum by an accretion-disk model for a large sample of objects (Wandel and Petrosian 1988).

   GASKELL (A comment) I think there is a lot of observational evidence for the "feast or famine" model as you call it. This is as follows (Gaskell, $Ap.$ $J.$, submitted): 1. The Dibai mass-luminosity relationship seems to be confirmed – we don't see high mass objects at low luminosities. 2. The "blue-bump" has a shape independent of luminosity – in accretion models this does not allow different accretion rates at the same mass. 3. There is a (weak) correlation of quasar luminosity with host galaxy luminosity. Galaxies certainly don't lose many orders of magnitude of mass with time!