WHAT QUASARS TELL US ABOUT GALACTIC EVOLUTION

MARTIN J REES
Institute of Astronomy
Madingley Road, Cambridge, CB3 OHA

Abstract. Quasars offer three types of clue to galactic evolution. (i) The formation of massive black holes in the centres of young galaxies; (ii) Clues to subgalactic structure from quasar absorption spectra; (iii) Implications for redshifts higher than 5. These are briefly summarised in this text.

1. AGNs and the Stellar Cores of Galaxies

The most remote quasar so far detected has z=4.89 (Schneider, Schmidt and Gunn 1991). The population genuinely seems to thin out at redshifts above 2.5-3: the comoving density of quasars falls by at least 3 for each unit increase in z (see Shaver, 1995 for a review). Recent progress in the study of active galactic nuclei (AGNs) bring into sharper focus the question of how and when supermassive black holes formed, and how this process relates to galaxy formation.

When the stellar core of a galaxy forms, part of the gas may collapse into a black hole. The formation and growth of the hole then manifests itself as a quasar; the peak in the quasar population (i.e. redshifts in the range 2 - 3) signifies the era when large galactic bulges were forming. This process involves complex gas dynamics and feedback from stars; we are still a long way from being able to make realistic calculations. At the moment, the most compelling argument that a massive black hole forms comes from the implausibility of the alternatives. To stop such an outcome, either:

(i) The formation of stars during the infall must be nearly 100 percent efficient; moreover, the stars would all need to be of low mass (so that no material is expelled out again).

or (ii) Gas must remain centrifugally supported in a self-gravitating disc for hundreds of orbital periods, without the onset of any instability that

R. Bender and R. L. Davies (eds.), New Light on Galaxy Evolution, 303–311.

redistributes angular momentum and allows the inner fraction to collapse.

Neither of these options seems at all likely – the first would require an IMF quite different from what is actually observed in the inner parts of galaxies; the second is contrary to well-established arguments that self-gravitating discs are dynamically unstable.

We first need to know at what stage in its progressive concentration towards the centre the gas stops being able to form stars (because of radiation pressure, magnetic fields, or whatever) and evolves instead into a supermassive object. A crude argument is the following.

When a self-gravitating object releases its binding energy in a dynamical (or free fall) timescale, the luminosity is

$$L_{diss} = \left(v_{virial}^5/G\right) \simeq 10^{59}\,(v_{virial}/c)^5\ \mathrm{erg\,s^{-1}}.$$

When v_{virial} gets high enough, this becomes comparable with the Eddington limit. Fragmentation is then impossible: instead, the gas is 'puffed up' by radiation pressure into a supermassive object. This would happen whenever $L_{diss} \gtrsim L_{Ed}$, i.e. when

$$v_{virial} > 300\left(M/10^6 M_\odot\right)^{\frac{1}{5}}\ \mathrm{km\,s^{-1}}.$$

Note that the above condition, while sufficient, is by no means necessary: fragmentation may stop at a substantially earlier stage. For instance, photoionization cross-sections can greatly exceed the Thomson cross-section , allowing pressure support for a lower L_{diss} (Haehnelt 1995). Dust opacity could also be important.

Magnetic stresses could also prevent fragmentation. Full equipartition requires a field of

$$\sim 0.1\left(M/10^8 M_\odot\right)^{-1}\left(v_{virial}/300\,\mathrm{km\,s^{-1}}\right)^4 G;$$

but even a field 1000 times weaker than this would exert enough pressure to inhibit the condensation of stellar-mass subunits; the free-electron fraction maintained by the intense UV radiation would be too high to permit ambipolar diffusion (which permits stars still to form, despite dynamically-important magnetic fields, within dense interstellar clouds in present-day galaxies). Quite apart from these constraints, fragmentation would be inhibited by tidal effects when the gas developed a sufficient central concentration.

Once a large mass of gas became too condensed to fragment into stars, it would continue to contract and deflate like a spinning supermassive star; viscosity would provide internal dissipation, and redistribute angular momentum. Some material would be shed during the contraction to carry away the angular momentum, but a substantial fraction could continue contracting until it underwent complete gravitational collapse – for example, if 10 percent of the mass had to be shed in order to allow contraction by a factor of 2, about 20 per cent would remain to form a black hole.

2. Quasar Remnants

Considerations of AGN 'demography' , by now well known, suggest that a typical galaxy could have passed through a quasar phase, and that by z = 2 (2-3 billion years) most had developed central holes of $10^6 - 10^9 M_\odot$. Van der Marel has reviewed the evidence for central dark masses – dead quasars – based on studying the spatial distribution and velocities of stars in the central core.

Much the most compelling evidence for a central black hole, however, recently came from a quite different technique: probing gas motions by measuring the 1.3 cm line of H_2O in the nearby peculiar spiral galaxy NGC 4258 (Watson and Wallin 1994; Miyoshi *et al.* 1995). The spectral resolution in this microwave line is high enough to pin down the velocities with accuracy of 1 km/sec. Such observations, combined with VLBA mapping with a resolution of 0.5 milliarc seconds (100 times sharper angular resolution than the HST) have revealed, right in the galaxy's core, a disc with rotational speeds following an exact Keplerian law around a compact dark mass. The circumstantial evidence for black holes has been gradually growing for 30 years, but this remarkable discovery clinches the case completely.

Should we be surprised that the putative holes in nearby galaxies are generally quiescent? Their environment could be almost free of gas, so that very little gets accreted; moreover, when the accretion rate is low, he cooling is so slow (because of the low densities) that only a small fraction of the binding energy gets radiated. The radio sources in the centres of some ellipticals, and perhaps also their X-ray emission, are consistent with accretion in a low-efficiency mode (Fabian and Rees, 1995 and references cited therein).

The case for a black hole in our own Galactic Centre was until recently ambiguous, but is now strengthened by the discovery of stars within the central 0.1 pc (Krabbe *et al.* 1995). The case would be clinched if some of these stars turned out to be moving as fast as (say) 1000 km/sec. If there is indeed a hole of a few million solar masses in our Galactic Center, the most recent tidal disruption of a star (even if it happened tens of thousands of years ago) may have left traces that could still be be detectable. Up to 10^{53} ergs of ionizing radiation could be released by accretion of the captured debris – more photons than would be emitted by steadier UV sources in the entire $\sim 10^5$ years interval between successive disruptions. Moreover the half of the star that is ejected may have left traces in some of the strange patterns of the gas within the central 2 parsecs.

3. Effects of Binary Black Hole Coalescence

Are the putative black holes actually described by a Kerr metric, as relativity predicts? Some important progress on this question has come recently from the ASCA X-ray satellite, which offers sufficient spectral resolution to reveal lines. Whereas optical emission lines come, for straightforward thermodynamic reasons, from a volume much larger than the hole itself, thermal X-ray emission can come from the innermost parts of an accretion flow. The lines should therefore display substantial gravitational redshifts, as well as large doppler shifts. There is already one convincing case (Tanaka *et al.* 1995) of a broad asymmetric emission line indicative of a relativistic disc viewed almost face-on, and others should soon follow

The most impressive test of general relativity would be detecting gravitational radiation from mergers of supermassive binaries: this involves no physics other than that of spacetime itself. Some events of this kind are expected. Quasar statistics suggest that most galaxies should harbour black holes that formed at $z > 2$. Moreover, many galaxies have experienced a merger since that time. When that happens, the holes in the two merging galaxies would spiral together, emitting, in their final coalescence, up to 10 per cent of their rest mass as a burst of gravitational radiation.

This burst would be in a frequency range around a millihertz – too low to be accessible to ground-based detectors. Space-based detectors are needed. One such being proposed, by the European Space Agency, is the Laser Interferometric Spacecraft (LISA) – six spacecraft on solar orbit, configured as two triangles, with a baseline of 5 million km whose length is monitored by laser interferometry.

LISA could detect the mergers of supermassive holes, even whose occurring at high redshifts, with high signal-to-noise. The bad news is that the event rate is low. Even out to $z = 5$, there could be less than one event per decade involving holes above $10^6 M_\odot$, even if there are enough black holes altogether to account for all the quasars (Haehnelt 1994). This is of course a lower limit. There could be lower-mass holes in small galaxies that are more common and underwent more mergers. The sensitivity of LISA is such that it could even detect waves from a stellar-mass object orbiting a supermassive hole.

LISA is at the moment just a proposal – even if it is funded, it is unlikely to fly before 2015. Is there any way of learning, before that date, something about gravitational radiation? The dynamics (and gravitational radiation) when two holes merge has so far been computed only for cases of special symmetry. The more general problem – coalescence of two Kerr holes with general orientations of their spin axes relative to the orbital angular momentum – is one of the US 'grand challenge' computational projects. When this challenge has been met (and it will almost certainly not take all the time until 2015) it will tell us not only the characteristic wave form of the radiation, but also the recoil that arises because there is a net emission of linear momentum.

This recoil could displace the hole from the centre of the merged galaxy – it might therefore be relevant to the low-z quasars that seem to be asymmetrically located in their hosts (and which may have been activated by a recent merger). The recoil might even be so violent that the merged hole breaks loose from its galaxy and goes hurtling through intergalactic space.

Density profiles in the centres of ellipticals would be altered by such an event. The inward spiralling of a binary black hole via dynamical friction would substantially change the orbits of stars whose total mass is at least of the order of the binary's own mass. This would 'puff up' the stellar distribution (as well as expelling roughly its own mass of stars completely from the galaxy when it becomes a 'hard' binary). The binding energy of the stellar core would be further reduced if the binary were completely expelled from the system by the gravitational-wave recoil during final coalescence.

4. Quasars and the end of the 'Dark Age'

The Universe took about half a million years to cool down to 3000 K (corresponding to $z = 1000$). Thereafter, further expansion shifted the primordial radiation into the infrared; a 'dark age' began, which persisted until the first nonlinear perturbations developed into bound systems and released enough nuclear or gravitational energy to light up the universe again. The high-z quasars tell us that, after about one billion years, the dark age had certainly ended. The lack of complete 'Gunn Peterson' absorption in the spectra implies that there had by then been enough energy input to re-ionize the primordial material; the existence of the quasars implies that black holes of as much as $10^8 M_\odot$ had accumulated.

What happened during the timespan from a million to around a billion years? The answer depends on the relation between quasars and galaxies, and on when galaxy formation began. One need only recall three much-studied options for structure formation (none of which can yet be ruled out) to indicate the current level of uncertainty:

(i) According to the simplest CDM model, the first structures – loosely bound systems of subgalactic scale – start to form at redshifts as high as 20 or 30. The galaxies themselves only form more recently, but early enough to have provided enough 'hosts' for the high-z quasars. More detailed studies

https://doi.org/10.1017/S0074180900232531 Published online by Cambridge University Press

(e.g. Haehnelt and Rees 1993) suggest that the z-distribution of quasars can be fitted if the mass of the hole that forms within each dark halo depends in a plausible way on the depth and profile of its potential well. If the CDM model were correct, the quasar density should fall off steeply beyond $z = 5$; the intergalactic medium would have been originally ionized, perhaps as early as $z = 20$, by stars in shallow potential wells of subgalactic scale.

(ii) If at least one species of neutrino has a mass of a few e.v, then the CDM model is modified by the presence of 'hot' dark matter. In the hybrid 'mixed dark matter (MDM) model, fluctuations on small scales have lower amplitude, relative to fluctuations on large scales, than for 'pure' CDM; there is therefore less structure at early times. The existence of quasars at $z = 5$ is a severe constraint on the fraction of 'hot' dark matter. Indeed the inferred ionization of the IGM at high redshifts is then itself a problem (even if it isn't due to quasars), since the fluctuation spectrum in MDM has less power on small scales, so not even subgalactic structures form early.

(iii) The so-called primordial isocurvature baryon (PIB) model could lead to non-linear structures soon after recombination. Bound systems that condense out early, and lose their angular momentum by Compton drag, could evolve directly into black holes (Umemura, Loeb and Turner, 1993), even before the virialisation of galactic-scale potential wells which seem a prerequisite for black hole formation in CDM and MDM.

Discovery of quasars at much higher redshifts than 5 would push back our estimates of when galaxies formed, unless the fluctuations are non-gaussian, or we adopt the radical view (which could be maintained if the PIB model were right) that these quasars are not closely connected with galaxies.

5. Subgalactic Structure

In the CDM model, weakly bound clouds of very low mass would be the first objects to condense out, and virialise. If their virial temperature were below about 500K, they would be unable to cool. The first non-linear structures that could inject energy into the Universe, at $z \gtrsim 10$, would have baryonic masses $\sim 10^5 M_\odot$. These virialise at temperatures exceeding 500K – hot enough for H_2 cooling to allow continuing collapse and fragmentation. We don't know the IMF of these first stars. If they were predominantly of low mass, they would be inefficient at generating energy, but could create a pre-galactic 'macho' component (which would thereafter behave like CDM, and therefore constitute a fraction of the dark matter in galactic halos) – this is an interesting possibility. But any O or B stars (or very massive objects) would produce UV, and perhaps soft X-rays, that could photoionize the medium, raising the Jeans mass to $\gtrsim 10^8 M_\odot$. (Complete photoionisation of the IGM would, as a byproduct, produce by $z = 5$ a universal heavy element abundance of 1 per cent of solar.)

Photoionisation must have been almost completed before $z = 5$. Just how quickly it would happen, even in a specific CDM model, is not clear. The amount of matter that needs to have condensed to provide the UV depends on the IMF in these subgalactic systems. However it has recently been realised that the first UV could have interesting feedback effects on H_2 – depending on its spectrum it could either enhance the H_2 formation, and therefore the cooling, or else destroy the molecules (Heiman, Rees and Loeb 1995) before photoionization has been achieved. These effects determine what adiabat the gas is on at $z = 5$, something which is important for interpreting quasar absorption lines.

6. The Lyman Forest

Quasars offer a wealth of information on galaxy formation and pregalactic structures via their rich absorption spectra. Two decades of effort by many theorists to understand the Lyman forest have bequeathed us a great deal of historical baggage, which confuses the subject even more than necessary. These absorption lines now seem a natural consequence of current ideas on hierarchical structure formation – if they had only just been discovered, this is surely the interpretation that would suggest itself to most of us. But when they were discovered, far less was known about the high-redshift universe, and there were fewer constraints on, for instance, a hot intergalactic medium.

The photoionized medium would be influenced by any inhomogeneities in the dark matter distribution ('minihalos') whose virial temperature exceeded its own temperature of few times 10^4K. The medium would then no longer produce a smooth Gunn-Peterson trough: nonuniformities in its density (and in its velocity field as well) would imprint weak Lyman forest lines on any UV continuum passing through it.

Most of the lines described by Tytler at this meeting correspond to HI column densities of only $\sim 10^{12}\,\mathrm{cm\,s^{-1}}$ and optical depths no more than unity even in the line centre. If a homogeneous IGM has a Gunn-Peterson optical depth of (say) 0.1, then (since the neutral fraction increases linearly with density) material need only be compressed one-dimensionally by a factor 10 to produce such a line.

The regions responsible for the weak lines are therefore definitely in a dynamical state: the relevant gas is being pulled by gravity, but has not yet evolved into a quasi-static virialised cloud (and maybe never will). Any cloud that has achieved virial equilibrium must have at least several hundred times the mean density, and a line of sight through it would intercept an HI column density of more than 10^{15} cm s^{-1}. (Miralda-Escudé *et al.* 1995 and references cited therein).

The observed lines are so close that it is unrealistic to think in terms of widely- separated discrete clouds with a uniform medium between them – just as it would be unrealistic to model an ocean surface as completely flat except for a few isolated high waves. Between the overdense regions, the density falls well below that of a uniform IGM, because matter tends to drain into the potential wells.

We expect deviations from Voigt profiles, because the lines will be Doppler-broadened partly by bulk motions. It is therefore important to set the lowest possible upper limit to the *thermal* broadening. This tells us the temperature (and therefore the form of the UV background). More important, it tells us something about the thermal history of the gas at even higher redshifts than those directly observed. This is because the recombination times at intergalactic densities are so long that adiabatic cooling in the expanding IGM (and subsequent adiabatic heating when a proto-cloud has started to collapse) affects the temperatures; so also, at high redshifts, does Compton cooling on the microwave background (see Miralda-Escudé and Rees, 1994 for some illustrations of such effects). The temperature therefore depends on when and how the gas was first ionized.

7. Realistic Limits of Computations

The purely gravitational clustering of non-dissipative dark matter is being computed with higher and higher resolution. It is now also possible to model the gas dynamics, including shocks and cooling, on all scales from the Lyman forest clouds up to clusters of galaxies. If the primordial gas were heated only by shocks when it fell into a bound system or formed a pancake, then the problem would be well-posed, and calculations would be feasible throughout all the non-linear stages.

In principle, the formation of the first stars and quasars is determined by the initial conditions. But there is of course no realistic chance of modelling the internal processes of star formation (not the central 'active nuclei') within bound systems. The physics and feedback effects from star formation, etc – which are important at the pregalactic era as well as within forming galaxies – involve parameters that cannot be computed a priori. Indeed the importance of the simulations lies in the hope that we will learn about these complex processes by modelling a wide parameter space and testing which assumptions yield the best agreement with the data on galaxies and their evolution – data whose volume and quality are advancing at least as fast as computational techniques.

References

Fabian A.C. and Rees, M.J. 1995 MNRAS 277, L55.
Haehnelt, M 1994 MNRAS 269, 199
Haehnelt, M 1995 MNRAS 273, 249
Haehnelt, M and Rees M.J. 1993 MNRAS 269, 168
Krabbe, K *et al.* 1995 Astr. Astrophys (in press)
Miyoshi et al 1995 Nature 373, 127.
Miralda-Escudé, J, Cen, R., Ostriker, J.P. and Rauch, M, 1995 Ap J.(in press)
Miralda-Escudé, J. & Rees, M.J. 1994 MNRAS 266, 343.
Schneider, D., Schmidt, M and Gunn, J.E. 1991 A.J. 102, 837
Shaver, P., 1995 Ann. N.Y. Acad. Sci (in press)
Tanaka, Y. *et al.* 1995 Nature 375, 659
Umemura, M., Loeb, A. and Turner, E.L. 1993 Ap. J. 419, 459.
Watson, W.D. and Wallin, B.K. 1994 Ap.J. 432. L35

C.S. Frenk: You mentioned that it is unlikely that many low mass galaxies would grow a $10^8 M_\odot$ black hole. A hint that these galaxies may be different from bigger ones is that their inferred dark matter halos have large core radii. Could you comment on this possible connection and on the physical processes that would distinguish small galaxies from big ones?

Yes, that might be part of the story – dissipative effects would be more efficient in a compact high-density core. However, I think the most important difference between high-mass and low-mass galaxies is that the former have deeper potential wells.

R. van der Marel: In studying the stellar dynamical effects of black holes in galactic nuclei it is often assumed (mainly for mathematical convenience) that the black hole grows adiabatically in an already existing stellar system. Do you consider this a plausible and viable assumption?

My guess would be that this is a rather poor assumption (though a profile with slope close to 3/2 may still be a good approximation for other reasons, and is an obvious 'template' to use). Stars might still have been forming at the same time as the hole was growing. The cusp profile could also have been modified if the central black hole had merged with another.

https://doi.org/10.1017/S0074180900232531 Published online by Cambridge University Press