The 100,000th image taken with the Hubble Space Telescope shows a bleak view, dominated by a pair of very bright star-like objects surrounded by a few faint galaxies. The object to the right is a star, but the central object just looks like a star; it appears to have the same brightness yet billions of light-years separate them. In fact, it's the most distant thing in this image, lying further away than all the faint galaxies. To a rough order of magnitude, the object at the centre is a million times further away than the star. This means (given the way light spreads out by the inverse square of the distance) it has to be a million million times brighter than a single star - ie it is brighter than the light from an entire galaxy. Welcome to the extraordinary world of the quasar, where some of the most incredibly bright objects in the Universe are powered by matter falling onto a super-massive black hole lying in the heart of a galaxy.

It is particularly apt to revisit the topic of quasars now (2013) as it is 50 years since their original identification opened up new realms of the Universe to astronomers. Their discovery was serendipitous, but they have become one of the most intriguing objects to observe. Their extreme luminosity means that they can be seen right the way across the visible Universe; and because they are so distant, the light travel-time means we are seeing these far-off objects when the Universe was very young. Quasars are thus readily observable tracers that help us map out the very early, very distant Universe - and of the space lying between us and them (but this will have to await another lecture!). They also reveal one of the important influences in the growth and development early in the history of many of the seemingly staid galaxies that surround us in the present day. But let us first backtrack to pick up the story from more than fifty years ago...

Active Galaxies

There are billions of 'ordinary' galaxies in the Universe. Each comprises a vast collection of stars, gas and dust gathered together under gravity into spiral or elliptical shapes. The light given off by all these stars, gas and dust is predominantly thermal 'blackbody' radiation, related to the temperature of the emitting body: the hotter the object, the bluer the radiation. A small percentage of all galaxies are active galaxies. These have all the light of a normal galaxy - and much more besides. An active galaxy is not only much more luminous at all wavelengths, but the nature of the energy it emits differs substantially from a normal galaxy.

There were early indications that some galaxies were different, even long before such 'nebulae' were recognized as separate galaxies lying outside the Milky Way. Photographs of the giant elliptical galaxy M87 taken by the American astronomer Heber Curtis in 1918 showed it to have a long 'ray' that stretched from the centre to well outside its envelope of stars. This feature remained an exception, and a puzzle.

Several decades later (in 1943) the American astronomer Carl Seyfert studied a selection of spiral galaxies that had long been known to show unusually bright and blue point-like cores, or 'nuclei'. The light from the nucleus alone was excessively bright compared to the rest of its galaxy, and it suggested that there could be something spectacular and unusual occurring at the centre of some galaxies. This impression was confirmed when Seyfert studied the visible spectra of these galaxies, which were very different from that given off by an ordinary galaxy.

The underlying light given off at all wavelengths (the 'continuum') is too bright, too blue, and lacking the absorption features that one would expect if it were simply the summed light of millions of ordinary stars at a range of temperatures. In addition, the spectrum shows strong peaks of emission at discrete wavelengths. These colours are produced by excited atoms, where the exact wavelength of light given off is uniquely dependent on the atomic structure of the chemical element. The presence of such emission lines in a Seyfert spectrum betrays not only the presence of clouds of excited atoms of gas - mainly hydrogen, but also heavier elements such as neon, nitrogen, oxygen, magnesium - but also a plentiful supply of the very energetic (ultraviolet) radiation required to excite them. Different emission lines sample different temperature, density and ionization regimes, and so can be used to measure the physical conditions around the active nuclei. One of the most interesting facts is that many of the lines given off by the gas clouds in these nuclei appear much broader than comparable lines observed from gases in the lab on Earth.

The light given off by a gas cloud is the sum of all the different photons emitted by individual atoms - all the same type of atoms in the same physical state should radiate light at exactly the same wavelength. Rather than this light appearing as an incredibly thin and bright spike at exactly one colour, however, the observer sees photons arriving with a slight spread of wavelength because the atoms are not still. Some of the photons are slightly blue- or red-shifted, due to the Doppler effect; any motions of the atoms to or away from us smear out the emission line. The amount the line is widened thus reveals the range of velocities of the atoms within the cloud. There will be simple random jostling movements because the atoms always have an average energy related to their temperature; this produces a predictable thermal broadening, where the width of the line depends only on the temperature and the mass of the atom. Doppler broadening can also result from any bulk flows of material within a gas cloud (such as rotation, or in/outflow), augmenting the width of the line still further. All Seyfert galaxies show a set of comparatively narrow lines where the width indicates a motion of around 400 km/s; these tend to be 'forbidden' lines, only emitted from regions of incredibly low density that we can't replicate in the lab.
on Earth. Some Seyferts, however, show a much broader set of 'permitted' lines (such as the well-studied Balmer series of Hydrogen lines) with widths revealing speeds some 10 to 1000 times faster than the ordinary rotation of material round the centre of spiral galaxy. The gas within the clouds responsible for these emission lines must be extremely turbulent. The exact interpretation of these observations was unclear at the time, and for many years Seyfert galaxies remained odd, pathological exceptions to galaxy behaviour.

A different type of active galaxy was revealed as the new science of radio astronomy developed in the late 1940's/early 1950's. Early work included a systematic mapping of the sky to catalogue the brighter radio sources, although the first observations were not yet very refined: the telescope antennae could detect radio emission from a source and measure its brightness at particular frequencies. The exact position of the source on the sky, however, could not be determined with much accuracy; the astronomers could only know an approximate direction in which the source was located on the sky. Originally many of the bright radio sources were identified as stars within the Milky Way, but by 1949 it was clear that sometimes an obvious large elliptical galaxy lay in the correct part of the sky to be the source. It was not a straightforward identification, however. If these galaxies were truly the source of the radio emission, their large distance implied a radio luminosity that would be far greater than expected from a normal galaxy. This would mean that there was a new kind of active galaxy, later known as a radio galaxy.

The first radio galaxy to be identified was Cygnus A, in 1954, which happens to be the most powerful example locally.

More modern observations show that the radio source is split into three components - two diffuse regions of emission known as radio 'lobes' surround a bright compact core that is aligned with the core of the galaxy. The lobes are symmetrically located to either side, but far out of, the galaxy; they are separated by about a million light years and have no visible identification. Long thin jets link one or both of the lobes back to the compact core. This radio emission is not thermal radiation stemming from the temperature of the source; instead it is produced by energetic electrons, moving around magnetic field lines at nearly the speed of light in a process known as synchrotron radiation. When the energy required to power the vast size and brightness of the Cygnus A radio structure was first estimated, it was a surprise to deduce that an otherwise 'normal' elliptical galaxy could be responsible for such a phenomenal release of energy at its core.

One of the most famous surveys was the 3C (Third Cambridge) catalogue, published in 1959, which lists several hundred of the brightest radio sources in the northern-hemisphere sky. Many of these are classical radio galaxies, identified with either the dominant galaxy of a cluster, or with fainter, smaller (and thus presumably more distant) elliptical galaxies.

Quasi-stellar Objects

Some of the brightest 3C sources evaded identification. Even though their radio emission seemed similar to that of the radio galaxies, there was no easy association of the source with an obvious elliptical galaxy. An early (1960) tentative identification of one of these radio sources, known as 3C48, was with a faint star. If so, it would have to be a very strange star – one that not only showed an unusually blue colour, but also varied enormously in brightness. The optical spectrum of the star showed lots of emission lines similar to those emitted by the Seyfert galaxies, but not lines that could be recognised, as they were not at the colours for known chemical elements. In addition, it was a struggle to perceive how a star could produce such prodigious radio emission. Two more of these curious radio 'stars' were identified (3C196 and 3C286); although they were both blue, the emission lines in their optical spectra were again very different from anything else seen – different both from each other and from 3C48.

In 1962 an opportunity arose for a much clearer determination of the position of one of these powerful but as yet unidentified sources, 3C273, when it would be eclipsed by the Moon three times - in May, August and October. The advantage of such an event is that we always know the position of the Moon exactly. If one then compares the precise location of the edge of its disc against the sky at the moments when the radio signal from 3C273 first disappeared and then re-emerged, the intersection of the two arcs pinpoints the source position much more accurately than by the radio emission on its own. The observations were carried out by Cyril Hazard (with MB Mackey and J Shrimmins) at the new Parkes radio dish in Australia. They discovered that the radio emission from 3C273 was distributed into two close components. The improved coordinates clearly indicated that one component originated from a faint, blue, and ordinary-looking star; the second component was associated with a very faint spur of gaseous material that seemed to jet from the star. The position of the star was passed to Maarten Schmidt who was able to use the 200-inch Hale telescope at Mount Palomar to obtain the star's visible spectrum in late December 1962. The resulting spectrum was again different from that of a normal star, and also like that of the other 'radio stars' in that it showed a bright blue continuum with an abundance of emission lines that appeared at different wavelengths/colours from the light given off by known chemical elements. Again, the positions of these lines didn't match what was seen in the other 3C 'radio stars'.

Schmidt pondered the spectrum for a while before he made the breakthrough of realising that the emission lines observed in 3C273 were the right lines, just occurring in the wrong part in the spectrum. The relative pattern was the same for all the usual emission lines due to hydrogen and other elements - and indeed, the same as had been seen in the visible spectrum of Seyfert galaxies - only if all the lines had been systematically shifted to redder wavelengths, indicating a velocity of recession of 48,000 km/s. This high a redshift was itself not new - it
was comparable to that seen in distant faint galaxies at the time – but no-one had considered that the spectrum of an individual star could be explained in this way. The redshift meant that either the star was close, but escaping out our galaxy at a ridiculously high speed; or that like the distant galaxies at such a redshift, it was moving away from us as part of the universal cosmological expansion. The latter explanation seemed (marginally) the less bizarre, but still presented a problem of interpretation. In the expansion of the Universe, there is a direct relationship (established in the late 1920's as Hubble's law) between the speed with which a cosmic object is receding and its distance away from us – the further objects move away faster. 3C273's redshift indicated that this 'star' was two and a half billion light-years distant; galaxies at such a redshift are faint. For any object to appear just like an ordinary star from such a distance it would have to be tremendously bright – some million million times brighter than our Sun, and over a thousand times the luminosity of an entire galaxy. 3C273 became the prototype that allowed astronomers to recognize a new class of strange sources which became known as quasi-stellar objects (QSO's or now contracted to quasars): compact, incredibly luminous and lying at huge distances.

In the light of Schmidt's fresh understanding of 3C273's spectrum, 3C48 and other radio stars were rapidly re-interpreted, and the way was paved for new quasars to be discovered over the next few months. In retrospect, it is not surprising that the interpretation of the optical spectrum of these radio stars had to await the identification of 3C273. Not only is it the brightest quasar optically, but it has a much smaller redshift (compared to the those of the other 'radio stars' of the time) and so recognition of the slight shift of the emission lines did not perhaps require such a large leap of intuition.

Although the original QSO were discovered through their radio emission, the peculiar blue colour and spectrum of the sources meant that soon many more could be discovered by their optical properties alone. It was also quickly realised that all quasars are variable in their brightness. Today we know that only about 10% of quasars have the strong radio emission – most have none. Surveys (using observations taken in a range of wavebands) are continually identifying and collecting more and more quasars to the present day – hundreds of thousands of these objects are now known. They litter the Universe; the most distant yet discovered is so far away that the time it takes its light to reach us means that we are seeing it as it was when the Universe was just 770 million years old.

Exactly what quasars were was subject to enormous controversy all through the 1970's. The similarity of the quasar optical spectrum to those of Seyfert galaxy nuclei strongly suggested that they could also be an analogous but more extreme version of the phenomenon occurring at the cores of active galaxies. The problem with this interpretation was that there was no sign of a surrounding galaxy, and this problem could not be resolved simply by observations. The light from any galaxy at that redshift would be impossible to see as it would be faint, close to, and simply swamped by the enormously bright glare of the quasar… much like trying to determine the shape and colour of a lampshade around an extremely bright lightbulb. Advances came only with the development of digital detectors known as ccd's (charged couple devices) in the 1980's; not only were these more sensitive than photographic plates, but several observations could be digitally combined to improve the signal in the data. The observations also required adaptive optics of ground-based imaging telescopes (or diffraction-limited telescopes such as the Hubble Space Telescope) so that the quasar light was not blurred. Finally the observations were good enough to reveal that the bright star-like emission was surrounded by a faint fuzz, which had a luminosity and size consistent with a surrounding galaxy at the same redshift. We now know that the bright blue core only appears to look like a star as it completely outshines its otherwise 'ordinary' host galaxy. Quasars live in both spiral and elliptical galaxies (though the latter tend to be predominantly the quasars that show the radio luminosity and structure akin to the radio galaxies). Sometimes the host galaxy appears to have a disturbed morphology, suggesting that it has undergone a gravitational interaction with another galaxy.

Observational properties of quasars

Quasars (whether with or without the radio emission) share basic observational properties with each other, as well as with radio galaxies and the Seyfert galaxies; this strongly suggests that they are different manifestations of the same physical process, just operating at a different intensity or in slightly different host galaxies. Grouping such systems as active galactic nuclei can help interpret the underlying process.

A quasar nucleus can emit over a thousand times more light than its host galaxy. This large a luminosity in itself is not a problem; however it can't simply originate from a thousand times more stars for two reasons: the nature of the emission, and the tiny region it is emitted from. A quasar shines at such high luminosities not only in the visible waveband, but some also have powerful radio and/or infra-red emission, and all emit most of their energy in the X-ray and UV part of the spectrum. The nature of the radiation produced by a quasar across all of the electromagnetic spectrum cannot be characterised as simple thermal emission due to the temperature of the source; the bulk of the light is thus not being generated by a simple collection of stars, gas and dust such as might be seen from a normal galaxy. It is produced instead predominantly by non-thermal emission processes.

Both Seyfert and quasar nuclei vary strongly in brightness with time at all wavelengths. These changes are large and swift, particularly in the very energetic X-ray waveband; for example, the Seyfert galaxy MCG-6-30-15 is seen to vary by a factor of 3 in its X-ray luminosity in only about 5 and a half hours. The timescale of the variability and the sharpness of the jumps in brightness limit the size of the region responsible for emitting all this power to be smaller than the distance that light travels in the same amount of time. The signal to 'change in
brightness' moves across the source at the speed of light; it can only result in a rapid change in luminosity if the signal reaches all parts of the object at once so it brightens coherently - otherwise the change in brightness would not appear as a sharp jump, but be smeared out with time. For the X-ray luminosity to change so abruptly within only five and a half hours, the deduced size of the emitting region is about 6 billion km, equivalent to the average separation between Pluto and the Sun. This leads to the staggering conclusion that the luminosity equivalent to several galaxies is emitted from a volume of space smaller than the size of our Solar System - and a region far too compact a region to physically accommodate hundreds of billions of stars.

The quasars with radio emission share the extended radio structure seen in the population of radio galaxies, and this also gives clues to the origin of the nuclear activity. The thinly collimated radio jets extend a few million light years from the compact core, and track where an energetic stream of charged particles has been expelled from the centre of the galaxy at relativistic speeds. Two such jets are seen to emerge from a radio galaxy, and the matter they contain splatters further out in space to form the diffuse radio lobes. A more one-sided structure is observed in radio quasars; although there are two lobes, often only one of the two jets is seen. The other jet is assumed to be directed away from our line of sight and thus not apparent to us due to the way its light is focused into its direction of motion through relativistic beaming. The existence of the jets again suggests that the central process not being simply due to stars. Their regular appearance demands a generic production mechanism which also naturally yields a preferred direction along which this matter is ejected; the straightness of the jets over huge distances also indicates that a memory of this direction and a continual output has been sustained for several million years.

The optical spectrum of quasars always shows an excess of blue/ultraviolet light and the plentiful emission lines. All quasars show the incredibly broad emission lines seen only in a subset of Seyfert galaxy activity, and the widths imply that the emitting gas clouds are moving at speeds in excess of 5000 km/s.

The Paradigm of a Central Supermassive Black Hole

The British astronomer Donald Lynden-Bell first showed in 1969 that material falling under gravity onto a massive black hole at the centre of a galaxy could generate tremendous amounts of energy. Over the years since, this explanation has become the accepted paradigm for the underlying process powering active galactic nuclei. Not only is it the only means we know to generate so much enough energy in such a tiny space, but it has proven consistent with all the observations. However, we have to appeal not to just an ordinary black hole, but it has to be a supermassive black hole – an object some millions to billions greater in mass than our Sun.

The accretion process is highly efficient. Matter is captured by gravity to swirl down and onto the black hole. The large amount of energy released when material falls through a gravitational field provides the source of all the power. The conversion of mass to energy implied by the luminosity of the brightest quasars can be estimated using the very famous equation E=mc^2. The conversion is unlikely to be completely efficient, and a conservative guess is that only about 10% of the energy implied by the rest mass is liberated (which is still much more efficient than the 0.7% efficiency conversion rate of the nuclear fusion processes at the heart of stars). The most luminous quasars only need to be consuming fuel at the rate of less than 1 solar mass a year. The matter to be accreted could come from disrupted stars, or gas in the host galaxy's interstellar medium – or could be driven towards the nucleus from the outside during a merger event with another galaxy. The prodigious power cannot be radiated directly from the black hole itself, as nothing can escape from the event horizon; it must come instead from the matter immediately outside this region.

The accreting material first has to lose angular momentum before it can fall onto the black hole, and does so through the formation of a thin, wide accretion disc (see my lecture on Rotation in Space for more information about such structures) which extends over a scale much larger than our Solar system. The matter in the disc collides and squeezes together, heating up to temperatures of tens of thousands of degrees. The resulting thermal blackbody radiation given off by the hot accretion disc provides the strong blue/ultraviolet continuum prevalent in spectra of quasar and Seyfert nuclei.

Not all the matter swirling down the vortex surrounding the black hole is accreted. Some must be launched outwards to create the twin jets of charged plasma seen in the radio-emitting quasars. The precise process to generate the jets is still far from clear; at the moment it is a problem primarily for theoretical study, as the region where the jets are produced is hidden from observation. The generation mechanism has to be able to finely collimate the jets. Most ideas focus on the role of the magnetic fields that lace through the matter in the rotating accretion disc, and the way they could wind up tightly enough to then propel charged plasma outwards at high velocity. The entire system can be assumed to operate around a single rotational axis, with the jets then be squirted out perpendicular to the plane of the accretion disc. The powerful X-ray emission comes from the region closest to the black hole - not from the inner regions of the accretion disc itself, but from a hot turbulent atmosphere of gas distributed below and above the disc in a 'corona'. This gas may well be associated with the production of the powerful jets. The X-rays also heat up the disc and thus contribute to the ultraviolet emission indirectly.

Hot ionized gas must be present to give rise to the emission lines seen in the Seyfert and quasar spectra. The gas clouds are sufficiently far from the black hole that they are not being accreted, but close enough to still be accelerated by the strong gravitational field. The atoms in the clouds are heated and excited by the energetic X-ray and powerful light emitted by the accretion disc and the corona. The large line-width will then be due to
rotational motions: the closer a gas cloud is to the black hole, the faster its orbital speed. The broad lines are thus emitted by the clouds closer in, moving at speeds up to 10,000 km/s, and the narrow lines are from clouds further out from the black hole, moving more slowly at hundreds of km/s. In the same way that the speed at which the Earth revolves around the Sun depends most strongly on the mass of the Sun, and the distance between the two bodies, we can use the rotational speed of the gas clouds to estimate the mass of the black hole they are in orbit around.

The range of brightness from an active nucleus relative to its host galaxy is most simply explained as due to varying rates of accretion; this suggests a scaling of activity between a quasar and a Seyfert galaxy. If all active galaxies are powered by the same basic central engine, we still have to appeal to variables such as the mass of the black hole, the accretion rate, its galaxy environment and the angle you view it from to reconcile the different manifestations of nuclear activity in galaxies in a 'unified' picture.

The appearance (and hence the classification) of the activity of a galaxy is thought to be strongly affected by its orientation to the viewer's line of sight; this is complicated by the presence of an obscuring torus of dusty opaque gas. Be clear that this is not the same as the accretion disc around the black hole - even though we expect it to be aligned around the same rotational axis, its inner edge is thousands of times wider than the luminous regions of the accretion disc, and barely observable with current telescopes. Quasars and Seyfert II galaxies (those active galaxies that show the broadest emission lines) are seen pole-on, giving the observer a clear line of sight to the black hole and all the processes at play around it. If the same system were viewed edge-on, however, both the light from the inner, broad-line gas clouds and the blue/UV light from the central accretion disc are now obscured; either all the activity is completely hidden, or only partially viewed to appear as a Seyfert I galaxy, or an elliptical radio galaxy with some narrow emission lines in its spectrum. Any dust contained in the obscuring torus can absorb the UV radiation from the central engine, heating up to re-radiate strongly at infra-red wavelengths and contribute to the quasar spectrum.

Firm support for this unified interpretation comes from observations of some Seyfert 1 galaxies (those where the inner regions are not seen directly). Sometimes a faint component of the spectrum can be isolated in polarised light which shows the features of an unobscured Seyfert II nucleus, such as broad lines and blue light. This is reflected emission – seen only where more distant gas clouds have acted as mirrors to redirect hidden light from the central regions towards our line of sight.

One consequence of the surrounding obscuration is that it can make quasars much more difficult to identify. If we don't observe the central activity at a favourable orientation, and if their presence is not revealed through the extended radio emission they can be hard to distinguish. Thus modern surveys counting the distribution of quasars have to employ discovery techniques that look either for the strong infrared emission re-radiated by the dust in the torus, or detect the most energetic X-rays which may be the only direct light that can escape from the centre. As we will see, it is important to be able to estimate the fraction of completely obscured quasars that might be missing from more traditional searches – recent work suggests that there may be twice as many quasars which appear to us as obscured rather than unobscured quasars. This has implications for what we understand about the development of the active galaxy with time – and may be compounded if the level of this dusty obscuration also changes as a galaxy ages.

The Distribution of Quasars

A further complication is that whether or not we see a quasar may not be simply due to our viewing angle; there seems to be an evolution of this activity with time as well. Remember, quasars observed at a high redshift are being observed 'back in time' as well. It's clear that their basic observational properties don't seem to change – quasar spectra across all wavebands show little evidence for change right out to the highest redshifts – so it would appear that the accretion mechanism powering the activity remains the same. However, the number of galaxies going through a quasar phase seems to vary as the Universe ages. Surveys count the relative numbers of quasars at different luminosities; then from seeing how this number changes with redshift enables an estimate of how the quasar number density changes with time (and of course such calculations take into account the stretching of space in the expansion of the Universe). The fact that it took so long for quasars to be discovered immediately reveals that quasars are not common in the nearby Universe – in fact they appear to occur in fewer than 1/100,000 galaxies. But the number density of quasars increases strongly with redshift; they were over a hundred times more common when the Universe was only a quarter of its present age and quasar activity seems to have peaked when the Universe was about 2-3 billion years old. For a period of time they were very common, and then the population has gradually petered out to the present day.

The fact that the majority of quasars are found at high redshift implies that galaxies were far more active in the earlier Universe. So far few quasars have been found in the very high redshift Universe, although the more distant a quasar is, the fainter and more difficult it will be to discover. However there does seem to be a genuine paucity of them in the early Universe. Maybe quasars only became active as their host galaxies developed. At the time of writing (October 2013), the furthest quasar yet known is ULAS J1120+0641, with a redshift of z=7.1. It is so remote that we are seeing it as it appeared only 770 million years after the Big Bang; but despite only having only had this comparatively short time to get its act together, it already contains a supermassive black hole at its core with the mass of 2 billion Suns – altogether not too dissimilar to some of the largest black holes that we see in the present-day Universe.
How to Form a Supermassive Black Hole

Such early quasar activity presents a problem - how do we grow such a massive black hole so quickly? A large progenitor has to be created to act as a 'seed' black hole that will accrete rapidly. Maybe we have to appeal to exceptional conditions occurring only within the very early galaxies to create it. Another complication is the similarity of quasar spectra throughout the Universe, in particular that they show the same relative intensities of emission lines from the same elements. This implies that the heavy elements (oxygen, iron, carbon...) that are produced at the cores of massive stars and supernovae already had to be formed early on; the galaxies that contain quasars must have hosted early and intense bursts of star formation. The creation of the black hole would then be associated with the first intense phase of star formation, suggesting that either one causes the other, or both are byproducts of the surrounding galaxy taking shape. Such star formation needs to have started quickly and efficiently, but as luck would have it, this first epoch of galaxy formation remains out of reach of our observational capacity. As yet, we have to appeal more to ideas that extrapolate from our understanding of similar processes in nearby galaxies.

One theory appeals to a hypothetical first generation of stars which were incredibly massive (100-1000 times the mass of our Sun), and composed only of the primordial hydrogen and helium. Computer simulations suggest that stars of over 300 Solar masses could collapse at the end of their (very short) lives to form a black hole of around 100-200 Solar masses. Even though massive for a stellar black hole, such objects are probably not sufficiently massive to settle to a privileged position in line to accrete a lot of material at the centre of the forming galaxy; and other calculations predict the resulting black hole to be even less massive. At least 100,000 solar masses worth of black hole is required as the seed in order for it to begin accreting substantial amounts of mass from other nearby stars. If the first generation of stars were tightly packed into a large star cluster at the centre of the galaxy, however, they can be expected to interact with each other through gravity. They could experience collisions and mergers - either as the stars or as the end products of stellar collapse, such as neutron stars and black holes. Simulations suggest that the end product could be either a giant black hole, or exceptionally massive star which would promptly collapse into a giant black hole of sufficient size. Alternatively, maybe a very rapid infall and accumulation of gas within the gravitational field at the core of a forming galaxy could create a single unstable supermassive star of up to a million solar masses in size that then promptly collapses.

Once having formed our original black hole, we then have to fuel it at a phenomenal rate in order for it to grow in time to be in place as a supermassive object within a few hundred million years. There is a fundamental limit to the rate at which a black hole can be fuelled, called the Eddington Limit. The faster a black hole accretes matter, the hotter and more luminous its accretion disc will become. All light exerts a physical radiation pressure, which manifests as an outward push on any nearby matter. It is created as from of all the photons hitting the surface of material – the pressure from each individual photon's hit (whether it is absorbed, or reflected) is very small, but the net effect from very many such photos becomes more important. (For example, sunlight exerts a very slight pressure, best seen for the way it forces a comets tail to stream out and away from the Sun... more of this in my Comets lecture next month!). If the accretion rate gets sufficiently high, then the radiation produced becomes so intense that its pressure will start to push the accreting matter away from the BH and deprive it of its fuel supply. All luminous quasars are seen to accrete close to this limiting rate. Even starting from a massive progenitor, the central black hole needs to have grown at a rate close to the Eddington limit for most of its life (or alternatively we have to suppose that black holes could somehow short periods where they are able to accrete at above this rate). This would require an almost continuous supply of matter from its immediate surroundings, or regular resupplies of fresh fuel delivered towards the core in galaxy mergers. To model this correctly, we need to have an idea of how long an individual galaxy spends in its quasar phase.

The rise and decline of the quasar population does not necessarily represent the activity of any one individual object. It likely maps out an overall population where each galaxy undergoes several short-lived bursts of black hole activity which start and die off several times in its history. The longer the quasar phase lasts, the more massive the central black hole grows; many successive generations of shorter epochs of quasar activity predict a smaller final mass. It is most likely that the quasar activity in any individual galaxy will last for only 10s to 100s of millions of years, and that the most galaxies were active when the Universe was about 3 billion years old.

Black hole feedback

Once formed, there is no way to dispose of a supermassive black hole. Even if it is no longer accreting, it will still occupy the centre of a galaxy. So even though there are no very luminous quasars nearby, giant black holes lurk at the centre of most the ordinary massive galaxies in our local Universe. Their presence is revealed, and their mass weighed, by a variety of methods – such as the orbital dynamics of gas clouds, water masers or stars all moving in response to the black hole's gravity. Our nearest giant elliptical galaxy is M87, the dominant central galaxy of the Virgo cluster and the galaxy with a jet photographed back in 1918. Some of the early observations taken with the Hubble Space Telescope revealed a rapidly whirling disc of gas at its core; the emission lines given off by this gas show a blue- and a red-shift from either side of the disc, the magnitude of which indicates that the material is moving around the centre with a rotational speed of about 500km/s. The mass of the central black hole is deduced to be four billion Solar masses. A dormant supermassive black hole is known even closer to home in our own Milky Way galaxy. The orbits of the innermost stars contained in the core of a giant star cluster at the centre of the Galaxy have been tracked for the past 20 years. Their orbital motions and speeds indicate that they are all in rotation around a supermassive black hole some four million times the mass of our Sun. Our
nearest neighbour, the Andromeda Galaxy, also contains a supermassive black hole, with a mass of over 20 million Solar masses.

We find that nearly every ordinary galaxy we search seems to contain a dormant supermassive black hole at its heart and thus must have gone a quasar phase (or several such phases) in its history. Indeed, it would appear that quasar activity is an important part of the development of every luminous galaxy. The presence of a central black hole may well be an unavoidable part of a galaxy's formation and evolution. There are then further questions about the fuelling – the centres of galaxies are environments rich in gas and dust, containing enough fuel to power a very luminous black hole if it were needed, so why would they stop being active? And did the black holes form alongside the galaxies, pre-date, or post-date their formation?

Further clues to the history and development of galaxies and their black holes come from a surprisingly tight correlation between the mass of a central supermassive black hole and the mass of its host galaxy. Central black holes may vary in mass between 10,000 to 10 billion Suns, but their mass is always about 0.3% of the mass of the stars in the spheroid of their host galaxy (this is the bulge at the centre of the spiral galaxies, or the whole galaxy in the case of ellipticals). There is no strong dependence on either the mass or type of the host. The black hole 'knows' about the mass of its surrounding galaxy and vice-versa, suggesting that the black hole and its galaxy have grown in tandem. Furthermore, it implies that the black hole can in some way influence the development of its host galaxy. A central massive black hole is not just merely decorative, but could be vital for regulating the process of galaxy formation, by limiting the growth of the surrounding host.

If energy is released into galaxies while they are still forming, it can curtail the collapse of gas clouds into stars. This energy can't simply be heating and winds from generations of giant stars exploding as supernovae - such a process would have produced abundances of heavy elements far in excess for what is observed. Instead the energy has to come from another, freely-available source, and certainly the activity from supermassive black hole is an obvious suspect. In fact, it's almost an embarrassment of riches as there's too much energy available - only a tiny fraction is needed to influence and halt star formation on the scale of an entire galaxy. The problem then is not one of supply, but of how to tap and distribute just a small percentage of the black hole's energy. The radiation can't be just absorbed directly as heat, because it would then re-emerge as another kind of emission that we would observe. Current ideas involve either pushing out kinetic energy in the form of winds, or just the simple radiation pressure acting on the dust embedded in the gas clouds to disperse and dispel them further – the exact method of how a black hole's energy halts star formation is currently a matter of great debate. As the black hole regulates the size of its surrounding galaxy it will also starve itself of fresh fuel, and its own mass will be regulated in turn. The few black holes still active today may be exceptional cases where a fresh supply of material is supplied from an external source, perhaps when the host is undergoing a disruptive merger with another galaxy.

Even when dormant, a central supermassive black hole will still happily snack on any fuel that comes its way. The one at the centre of our own Milky Way has been under close scrutiny by observers over the Summer, as a cloud of gas and dust known as G2 hurtles towards, following an orbit that has brought it on a close approach. The cloud only has a mass some three times that of the Earth, and we don't know where it's from - possibly it was torn off the outer atmosphere of a dying giant star. The predictions are that it won't be completely devoured, as it never gets close enough to the black hole to disappear over the 'event horizon'. But it's not expected to completely survive either - first it will heat up to incredibly high temperatures, and will then be stretched, stripped and shredded by the intense gravitational forces over a period of several month. So far it has gone past the black hole, and it is already being distorted into a long arc of material – providing us at close hand with a fascinating glimpse of the kind of processes responsible for the most powerful objects in the Universe, the distant quasars.

© Professor Carolin Crawford 2013