Evidence for BH: Active Galaxies

This is the second lecture in which we'll talk about evidence for the existence of black holes in the universe. Here we focus on active galactic nuclei, or AGN. Black holes in these systems are thought to have masses $M \sim 10^{6-9}\,M_\odot$.

The term "active galactic nucleus" encompasses a bewildering variety of sources. It includes quasars, blazars, ULIRGs, LINERs, BL Lac objects, OVV, Seyferts 1 and 2 (and 1.5 and 1.7 and 1.8 and...), radio galaxies, and so on. The only thing that ties them together observationally is that they are typically much more energetic than the average galactic nucleus: total isotropic luminosities might typically be 10^{42-46} erg s^{-1}, although similar sources are seen with much lower luminosities. Note, by the way, that if these sources are accreting at the Eddington rate, this implies (at the high end) masses of more than $10^8\,M_\odot$. The theoretical challenge is to put them all on some approximately equal footing in terms of their central engine. We are *not* going to go through the complicated taxonomy; it isn't worth it unless you plan to focus on AGN in your research. What we will do is try to apply some of the physics we've discussed earlier to understand the basics.

The understanding that there exist such unusual sources can be traced back to the 1950s, when it was discovered that Cygnus A and other galaxies have powerful radio lobes. An important breakthrough came in 1963 when Maarten Schmidt, in puzzling over the highly unusual spectrum of 3C 273, had the brilliant idea that it might be a normal hydrogen spectrum that was redshifted by a then-unprecedented amount (of $z = 0.17$!). Prior to this people had tried to fit the spectrum with things like uranium transitions, but the redshift idea worked much better! Combined with the observed flux, the high redshift of quasars indicated that they had incredibly high luminosities. This was especially impressive because they looked pointlike (thus "quasi-stellar radio source"=quasar). That meant that the energy had to be generated in an unbelievably small region compared to normal galaxies. The problem was sharpened by the observation of rapid variability; these sources often varied faster than a day, meaning that the simplest inference was a source smaller than a light-day (larger sources could not be causally connected [barring a loophole to be discussed soon], and hence could not all light up or all dim in concert). Some of the fastest variation seen is on the order of minutes.

This, however, led to a problem. The spectra of AGN are often "flat" out to MeV energies or beyond. That means that the photon number spectrum is

$$\frac{dN}{dE} \sim E^{-2} \tag{1}$$

out to at least an MeV. Doesn't look flat, does it? What the terminology means is that the amount of energy emitted is equal in equal logarithmic intervals of the photon energy (e.g., there is as much energy between 10 keV and 100 keV as there is between 100 keV and 1 MeV). This means that in a number of sources the gamma-ray energy is $\sim 10^{44}$ erg s^{-1}.

Ask class: suppose, then, that we want to know the optical depth of such photons to photon-photon pair production, for a spherical source of luminosity L and radius R. How do we do it? The optical depth is $\tau = n\sigma d$ in general. Here $n \approx L/(4\pi c R^2 m_e c^2)$, $\sigma \approx \sigma_T/10$, and $d = R$, so

$$\tau \approx 0.1 \frac{L\sigma_T}{4\pi m_e c^3 R} = 0.1\ell_{\rm MeV} \ , \tag{2}$$

where $\ell_i \equiv L_i\sigma_T/(4\pi m_e c^3 R)$ is the *compactness parameter* for photons of energy i. In our case, $L = 10^{44}$ erg s^{-1} and $R = 3\times 10^{14}$ cm (assuming a variation time of an hour), so $\tau \approx 10$.

Ask class: what would this imply? It would mean that practically no MeV photons could escape, so there would be at the very least a pronounced dip in the spectrum, which is not seen. One can make similar arguments based on an analysis of synchrotron radiation. Relativistic electrons will spiral around magnetic field and produce synchrotron radiation. If the source is very compact, then those photons will inverse Compton scatter off of the same relativistic electrons that produced the photons to begin with. A source that has too high a temperature (and hence too high a radiation energy density) will put the energy in gamma rays instead of in radio photons. This allows a limit to be placed on the radio brightness temperature (which is the effective temperature of a blackbody with that much radio emission). The condition that Compton power be less than synchrotron power leads to a limit on the brightness temperature at 1 GHz of

$$T_m < 10^{12}\ {\rm K}\ . \tag{3}$$

There are a number of sources that violate this limit!

Ask class: what is a way out? As first realized by Martin Rees, a way out is that there must be relativistic motion in these sources. If the emission we see comes from a region moving towards us relativistically, then everything is beamed and all of the energy we see (whether in MeV photons or in radio photons) implies a much smaller energy density in the rest frame, so optical depth or inverse Compton scattering limits are not violated. But, as Rees also realized, such relativistic motion has another consequence. He predicted that relativistic bulk motion of, e.g., blobs of radio-emitting ejecta would be seen to move "superluminally".

The reason for this is that if a source is moving with a speed βc in a direction an angle θ from our line of sight, then the measured transverse expansion speed is

$$\beta_{\rm obs} = \frac{\beta\sin\theta}{1-\beta\cos\theta}\ . \tag{4}$$

This is a maximum when $\theta = \cos^{-1}\beta$, in which case $\beta_{\rm obs} = \gamma\beta$. If $\beta > 1/\sqrt{2}$, the maximum transverse velocity exceeds c. This effect has now been seen in many sources, with a maximum $\beta_{\rm obs} \sim 30c$, meaning that the Lorentz factor has to be at least 30 in those sources. Nothing much higher is seen, though; this may be because 30 is a limit, or it could be because the solid angle to see the maximum goes like $\sim \gamma^{-2}$, so observation of $\sim 10^4$ sources would be required to expect to get a $\gamma \sim 10^2$ source.

We can use this to think of the phenomenon of one-sided jets. We'll talk more about jets in a second, but one thing often seen is that a source has only one jet. Assuming that sources always have two identical jets in opposite directions, **Ask class:** how can the observation of only one side be explained? With relativistic motion, the ratio of the frequency-integrated specific intensities (remembering $I \sim \nu^4$) is γ^8, so for $\gamma = 30$ the ratio is 10^{12} and one jet would be invisible even if the other is bright. In a given waveband $\hbar\omega$ one must also correct for the fact that one is looking at rest frame photons of energy $\gamma\hbar\omega$ from the away jet and $\hbar\omega/\gamma$ from the towards jet.

Central Engine

It is by now well-established that most active galaxies are powered by accretion onto black holes. Early on in the game other central engines were suggested, such as active phases of starbursts. However, for many sources the black hole origin is greatly preferred. Reasons include (1) rapid variability implies large energy generation in a small volume, not compatible with a starburst, (2) relativistic motion of jets and the hard spectrum imply a relativistic potential somewhere, also not consistent with starbursts. Some active galaxies, however, such as ultraluminous infrared galaxies, may well be powered by starbursts, as might some of the less luminous active galaxies of other types.

Jets

Now let's think some more about jets. Jets are seen in many places in astrophysics, not just active galactic nuclei. These include protostellar systems, stellar-mass black holes and neutron stars, and AGN.

Jets across astrophysics have certain things in common. The speed of the jet seems correlated with the escape speed at the surface of the star. This is true from protostellar jets up to jets from black hole sources such as AGN, where the projected motion can be faster than the speed of light by factors up to $\sim 20 - 30$, as said above. Anyway, this suggests that jets are formed deep in the potential well, meaning near the star or in the innermost portions of the disk.

The problem of the formation of jets is twofold: acceleration and collimation. The acceleration problem here is different from the acceleration problems we discussed earlier.

There, we talked about how one could get a small population of high-energy particles. Here, we need to get bulk acceleration to high velocity ($\gamma \sim 30$ in some AGN!). There are three general ways that have been proposed to form jets:

(1) Hydrodynamic acceleration. If the jet material starts flowing outwards in a region with sufficient external pressure, its cross-sectional area decreases and its velocity increases. Then, when the speed reaches the speed of sound, Bernoulli's equations show that the cross-sectional area increases (but not faster than the radius, so the collimation angle is small) and the speed also increases. A converging-diverging channel like this is called a de Laval nozzle, and is an essential principle behind rockets. This is therefore an efficient way to produce fast, collimated jets. In some cases (e.g., extragalactic jets) the external gas pressure cannot be high enough to provide the required initial collimation, because if it were then the gas would radiate a higher X-ray flux than observed. It might, however, play some role in protostellar jets.

(2) Radiative acceleration. Another possibility is that radiation from the central object can accelerate matter, which is separately collimated by a surrounding torus of gas. **Ask class:** suppose we had a gas of fully ionized hydrogen; what approximate luminosity would be required for such acceleration? About the Eddington luminosity, or around 10^{38} erg s^{-1} for a solar mass. This can be decreased if higher cross section absorption processes dominate, but not enough in protostellar systems. It is thought that interactions with the radiation may explain why the maximum Lorentz factor is around 20 in AGN: higher than that, and radiation *drag* slows down the jet. In particular, detailed analyses (e.g., Abramowicz, Ellis, & Lanza 1990, ApJ, 361, 470) show that if the source has a maximum luminosity $L = xL_E$, where L_E is the Eddington luminosity for the particles of interest, then $\gamma_{\max} \sim x^{1/3}$. This means that electron-positron jets (with $L_E \approx 10^{-3}L_E$ for ionized hydrogen) might get up to $\gamma = 20 - 30$, given that observed luminosities are no more than about 10 times the ionized hydrogen Eddington luminosity.

(3) Hydromagnetic acceleration. Power comes from, and symmetry is provided by, the rotation of the accretion disk. In AGN, the rotational energy of the black hole itself could be important; this is the basis of the Blandford-Znajek effect.

A particular model is that of Blandford & Payne (1982). The disk is threaded by a poloidal field, and material is accelerated and collimated by the magnetic funnel (toroidal fields). The less-collimated wind may be generated from a wider variety of radii in the disk and may be caused by strong gradients in the magnetic pressure close to the star/disk boundary.

Fueling of AGN

Another question that occurs in the study of AGN is where the central black hole gets the fuel to generate all these effects. This is a question that is of great current interest. One reason is that the number density of quasars and other AGN was much higher in the past, peaking at around $z \sim 2$, and is small now. The star formation rate also was much higher at $z \sim 2$ (it might have been constant from $z \sim 2-5$); is this coincidence, or is it an indication of a relation? Another discovery is that black holes in the centers of galaxies seem to have masses roughly 0.1% of the mass of the central bulge of the galaxy. This suggests that either black holes play an important role in the formation of the central parts of galaxies, or vice versa, or both.

Ask class: what are some ways that black holes might build up their mass in the early universe? Accretion of gas or of stars; merger with other black holes.

This is very much a current topic, so clear answers aren't easy to come by. It seems most likely that black holes are initially formed by or during the first generation of stars, so their initial masses would be at most $10^{2-3}\,M_\odot$, which is much less than their current masses. Then growing to masses of $> 10^9\,M_\odot$ by a redshift of $z = 6.5$ is extremely challenging, and probably requires (1) some way to maintain low rotation and hence low efficiency, so the growth rate can be higher, (2) really high initial mass (on the order of $10^5\,M_\odot$), so fewer e-foldings are necessary, (3) super-Eddington rates of growth, or (4) some non-radiative growth mechanism such as black hole mergers or growth from very low angular momentum dark matter.

Grand Unified Models

Finally, we will very briefly describe one attempt (due to many people) to unify the very diverse phenomenology of AGN. The idea is that there are a relatively small number of driving parameters of the system (some intrinsic, like black hole mass, and some extrinsic, like our viewing angle). For intrinsic parameters, the suggestion is that the behavior of the accretion disk is driven primarily by the BH mass and the accretion rate, and to a lesser degree by the spin of the black hole. One may also have a dusty obscuring torus, which from some viewpoints makes the AGN look different than from others (this is supposed to be the reason for the difference between Seyfert 1s and 2s). However, an important extrinsic parameter is our viewing angle. If we look straight down a relativistic jet (as probably happens in blazars) then we see a hard spectrum extending to very high energies, rapid variability, and we infer a huge isotropic luminosity. Looking at an angle to the jet, variability and inferred luminosity are much milder. There is also a viewing effect expected relative to the disk itself. One effect is that if we can see emission lines from the disk, they will be more broadened if we see the disk edge-on than if we see it face-on.

These unified models have had a lot of success, but as with any large set of complicated objects there are tons of things not especially well explained. However, this seems to have merit in general and probably explains most of the gross variations of phenomenology.

Additional references: In our library we have "Active Galactic Nuclei" by Julian Krolik of Johns Hopkins, which is a nice general reference and goes into many more specifics than we have time for in this class.