Lecture 13: Black hole growth and formation

Dr. James Mullaney

May 28, 2019

1 Introduction

Last week we started looking at Active Galactic Nuclei (AGN) – sites of rapid supermassive black hole growth at the centres of galaxies. In those lectures, we just assumed that supermassive black holes are a given, whereas, in fact, how these huge black holes were formed remains a major unsolved question in extragalactic research. In this lecture, we will consider when supermassive black holes were formed, and consider current theories of how they were formed.

2 The earliest known supermassive black holes

There is no known way we can measure the "age" a black hole in the local Universe, so the only way we can hope to estimate when the first supermassive black holes were formed is by observing them at high redshifts. Since one of the key ingredients needed to produce an AGN is a supermassive black hole, then finding a AGNs at high redshifts is a clear indication that supermassive black holes existed at early times. To date, the most distant AGN known is ULAS J112001.48+064124.3 (hereafter, ULAS): a quasar at redshift $z = 7.088$. This redshift corresponds to a time only 744 Myr after the Big Bang, meaning the supermassive black hole must have formed soon after the beginning of the Universe.

Although simply identifying an AGN at such high redshifts provides key insights into the formation of supermassive black holes, it would be even better if we could determine the mass of the black hole. With different formation scenarios likely producing different initial masses of black holes, measuring the masses of high redshift black holes helps us to constrain these models. Thankfully, measuring the mass of the black hole at the centre of an AGN is *relatively* straightforward, even for an AGN at $z \sim 7$.

Firstly, we can get a crude estimation of a lower limit of the mass of the black hole via Eddington luminosity arguments. As we saw in the last lecture, the Eddington luminosity of a black hole is given by:

$$L_{\mathrm{Edd}} = 1.3 \times 10^{38} \left(\frac{M}{\mathrm{M}_{\odot}} \right) \ \mathrm{erg\ s^{-1}} \tag{1}$$

which roughly corresponds to the maximum luminosity an AGN can have (it's only "roughly" because the Eddington luminosity is defined for spherical accretion, whereas an AGN accretes matter in the form of a disk). This gives:

$$L_{\mathrm{AGN}} \lesssim 1.3 \times 10^{38} \left(\frac{M_{\mathrm{BH}}}{\mathrm{M}_{\odot}} \right) \ \mathrm{erg\ s^{-1}} \tag{2}$$

where $L_{\rm AGN}$ is the bolometric luminosity of the AGN, and $M_{\rm BH}$ is the mass of the black hole. After rearranging, this gives:

$$M_{\rm BH} \gtrsim \frac{L_{\rm AGN}}{1.3 \times 10^{38}\ {\rm ergs\ s^{-1}}}\ {\rm M_\odot} \tag{3}$$

Since we know the redshift to the quasar, it is easy to calculate its luminosity, which is measured as 2×10^{47} erg s^{-1}. Plugging this into the above formula gives a black hole mass of: $M_{\rm BH} \gtrsim 1.5 \times 10^9\ {\rm M_\odot}$. So, even by this crude approximation, it is clear that we're observing a very massive supermassive black hole just a few hundred million years after the Big Bang.

2.1 A more precise mass estimate

While the Eddington luminosity method provides us with crude lower limit for the mass of the black hole, it would be preferable to have an actual mass measurement. Thankfully, this is possible via the "virial technique" which, like reverberation mapping (see Lecture 12), uses the broad line region (BLR) to provide a mass estimate.

Assuming that the motions of gas within the BLR is dominated by the gravity of the black hole, we can use Newtonian dynamics to measure the black hole's mass:

$$M_{\rm BH} = \frac{v^2 r}{G} \tag{4}$$

where v is the velocity of the gas in the BLR, r is the radius of the BLR, and G is the gravitational constant. As we saw in reverberation mapping, measuring v is straightforward, we simply take the quasar's spectrum and measure the velocity widths of the broad emission lines (in this case, the permitted line C IV, since Hα and Hβ are shifted out of the optical range). Like in reverberation mapping, however, measuring r is more complicated. It is observationally expensive to repeatedly take the spectrum of a high redshift quasar, so as yet we have not obtained the multi-epoch observations required to perform reverberation mapping for ULAS.

There is, however, another way to calculate r that uses the results of reverberation mapping of AGNs in the local Universe. As more and more nearby AGNs had their black hole masses measured by reverberation mapping, it was soon realised that the radius of the BLR is tightly correlated with the luminosity of the AGN. It is thought that this is because the increased light from high luminosity AGNs "pushes out" the regions in which the broad lines are produced. What this means is that we can obtain an estimate of r from the luminosity of the AGN, i.e.,:

$$\log(r/\text{light days}) = -21.3 + 0.519 \log(L_{\rm AGN}/\text{erg s}^{-1}) \tag{5}$$

From this, we get $r = 3.2$ light days, or 8.3×10^{13} m. This gives a value for the mass of the black hole at the centre of ULAS as $2 \times 10^9\ {\rm M_\odot}$. This is in the ball-park of the Eddington luminosity estimate, indicating that ULAS is accreting at close to its Eddington limit.

3 The implications of a $10^9\ {\rm M_\odot}$ BH at $z = 7$.

Now that we have confirmed the presence of a *billion* solar mass black hole just a few hundred years after the Big Bang, what does this imply for our understanding of how supermassive black holes form? Does this imply the presence of "primordial" supermassive black holes formed by the Big Bang, or could these black holes have formed from "stellar mass" black holes accreting rapidly for the previous 744 million years?

To answer this question, we have to consider whether a black hole could accrete mass quickly enough to grow by a billion solar masses in the space of a few hundred million years. We've already seen that there's an approximate upper limit to how quickly a black hole (or, for than matter, any object) can accrete material: the rate corresponding to the Eddington Luminosity (known as the Eddington rate). So, if we assume that ULAS's black hole accreted at its Eddington limit for the previous 744 million years, what is the maximum mass it could have accreted in that time? To calculate that, we need to integrate the Eddington rate with respect to time, since it increases with time as the black hole gains mass.

We'll start with the Eddington luminosity:

$$L_{\mathrm{Edd}} = \frac{4\pi c G m_p}{\sigma_T} M_{\mathrm{BH}} = K c M_{\mathrm{BH}} \tag{6}$$

where I've grouped all the constants (except c: you'll see why later) into K (see Lecture 12 notes for the meaning of all the terms in this equation). We also know that the luminosity of an AGN is related to its accretion rate via:

$$L_{\mathrm{AGN}} = \eta \dot{M}_{\mathrm{BH}} c^2 \tag{7}$$

and that the maximum accretion rate occurs when $L_{\mathrm{AGN}} \approx L_{\mathrm{Edd}}$, so subbing Eqn. 7 into Eqn. 6 gives:

$$\dot{M}_{\mathrm{BH}} = \frac{Kc}{\eta c^2} M_{\mathrm{BH}} = \frac{K}{\eta c} M_{\mathrm{BH}} \tag{8}$$

Since $\dot{M}_{\mathrm{BH}} = dM_{\mathrm{BH}}/dt$ we can separate the differential equation into:

$$\int_{M_{\mathrm{S}}}^{M_{\mathrm{F}}} \frac{dM_{\mathrm{BH}}}{M_{\mathrm{BH}}} = \frac{K}{\eta c} \int_{t_{\mathrm{S}}}^{t_{\mathrm{F}}} dt \tag{9}$$

where t_{S} is the start time and t_{F} is the finish time, and M_{S} and M_{F} are the black hole masses at those two times, respectively. The above equation integrates to:

$$\ln\left(\frac{M_{\mathrm{F}}}{M_{\mathrm{S}}}\right) = \frac{K}{\eta c}(t_{\mathrm{F}} - t_{\mathrm{S}}) \tag{10}$$

or,

$$\frac{M_{\mathrm{F}}}{M_{\mathrm{S}}} = \exp\left(\Delta t \frac{K}{\eta c}\right) \tag{11}$$

where $\Delta t = t_{\mathrm{F}} - t_{\mathrm{S}}$. In SI units, $K = 2.1 \times 10^{-8}$, so assuming $\eta = 0.1$ gives:

$$\frac{M_{\mathrm{F}}}{M_{\mathrm{S}}} = \exp\left(2.2 \times 10^{-8} \Delta t\right) \tag{12}$$

where here Δt is in years.[1] For $\Delta t = 744$ million years, this gives $M_{\mathrm{F}}/M_{\mathrm{S}} \approx 1.2 \times 10^7$. Since, in the case of ULAS, $M_{\mathrm{F}} = 2 \times 10^9\ M_{\odot}$, this implies that $M_{\mathrm{S}} = 170\ \mathrm{M}_{\odot}$. It is just about possible that the most massive early stars had masses of this scale, but this still presents a challenge for black hole formation models, as it implies that ULAS's black hole must have constantly grown at or above its Eddington limit throughout its entire life to that point. That's quite a remarkable feat!

[1] It's a coincidence that there are about $0.1c$ seconds in a year, so the ηc almost cancels out.

4 Possible BH formation channels

In the previous section, we saw how forming a $\sim 10^9$ $M_\odot$ black hole by $z = 7.088$ challenges our assumptions on how these black holes grow and how they are "seeded" (i.e., what do they grow from?). Of course, one possibility is that massive (i.e., > 1000 $M_\odot$, but perhaps not supermassive) black holes were formed out of the Big Bang. These so-called "primordial" black holes would then go on to accrete matter in the early Universe to create supermassive black holes by $z \sim 7$. However, little is known about how these "primordial" black holes would have formed out of the Big Bang, so there isn't much to report on them. Instead, we'll consider three other possibilities in which black holes form from normal matter sometime after the Big Bang. All three scenarios start-off with a lump of gas contained within an early dark matter halo, but it's what happens to this gas that distinguishes between the three models.

4.1 A single massive star

In this scenario, the lump of gas cools to form the very first stars. This cooling happens very slowly because the gas contains no metals, which are an effective way of radiating energy away from the gas cloud (via their emission lines) under normal circumstance (i.e., in today's Universe). Because of this slow cooling, the cloud fragments differently from today's metal-rich gas clouds, forming far more massive stars compared to star-forming clouds today. If the most massive of these stars ends up around 300 times more massive than the Sun, it will collapse to form a 200 $M_\odot$ black hole which, as we saw earlier, is just about massive enough to form a billion solar mass black hole by $z \sim 7$. However, this would require the black hole to accrete at or above its Eddington limit for all the intervening time.

4.2 A single supermassive star

Rather than forming a single massive star, in this model a single *super*massive star forms. This comes about because rather than fragmenting into smaller stars, the gas cloud monotonically collapses in on itself forming a single, supermassive star with masses upwards of 10,000 $M_\odot$. With such a high mass, the pressure in the centre of the star is so great that it quickly collapses to form a black hole embedded in the envelope of the rest of the star. This central black hole then quickly consumes the envelope, rapidly growing in size to form a million solar mass black hole. This black hole can then accrete at (average) rates much lower than the Eddington limit to form a billion solar mass black hole by $z \sim 7$.

On writing this, I can see many drawbacks with this model. Even if it is possible for a gas cloud to monotonically collapse to form a single star, it doesn't explain how the resulting black-hole-embedded-in-star doesn't blast away its outer layers due to super-Eddington accretion.

4.3 A dense cluster of merging stars

Rather than forming a single star, most gas clouds will collapse to form a population of stars. This scenario exploits this feature of gas cloud collapse to produce a massive (i.e., $\sim 10^3$ $M_\odot$) solar mass star. First, the pristine gas cloud collapses to form a population of stars, but for whatever reason (again, possibly due to the low metallicity of the cloud) this forms a far more dense cluster of stars than we see in the local Universe. As a result of the extremely high density of the resulting stellar cluster, the stars soon merge to form one or more massive stars with mass of the order a thousand stellar masses. These massive stars then rapidly age (losing comparatively little mass via winds due to their low metallicities) and die, forming a massive black hole of a few hundred or thousand solar masses. This channel benefits from being less constrained by the Eddington limit, since that only applies to accreting gas. By contrast, merging stars are not affected by the photon pressure that balances gravity in gas accretion.

5 Learning objectives for Lecture 13

With much of this lecture dedicated to the mathematical derivation of how quickly a supermassive black hole can grow, there isn't a huge amount to read. It's important that you understand this derivation, as it highlights the challenges astronomers face in understanding how supermassive black holes were already in place by $z \sim 7$. As you will likely have noted, there remains considerable uncertainties in all our models of how seed black holes formed. This is, in part, because of the extreme difficulties in observing these redshifts and the corresponding lack of empirical evidence for how early, pristine gas clouds collapse to form stars and, ultimately, black holes. It's a fascinating area of research.

So, after this lecture you should:

- have an understanding of the Eddington luminosity/rate/limit (they all refer to the same physical process) and why it arises (i.e., photon pressure on gas balancing gravity);
- be familiar with what the results from observations of the most distant quasar imply for our understanding of the formation of supermassive black holes;
- understand the virial technique of measuring black hole masses;
- be familiar with the current most popular formation mechanisms for seed black holes in the very early Universe.