PHY405: Galaxy Formation and Evolution

Dr. James Mullaney
F19 Hicks
j.mullaney@sheffield.ac.uk

January 8, 2020

1 Course description

This module will cover one of the most exciting and fast-moving topics in current astrophysics research, the formation and evolution of galaxies from an observational perspective. Starting with a brief historical introduction, the module will then summarise what we can learn about galaxy evolution from studies of galaxies in the local Universe, before discussing the results obtained from recent deep field observations of the high redshift Universe. The last part of the module will concern the important role that active galactic nuclei play in galaxy evolution. The final lecture will cover the future of galaxy evolution over the next $10^{100}+$ years.

2 Administration

2.1 Lectures

The course will consist of 18 lectures (+2 revision lectures) with two lectures held per week:

- Monday 10-11am; Hicks Building F30
- Tuesday 2-3pm; Hicks Building K14

2.2 Course Notes

This booklet of course notes are to be read in conjunction with attending the lectures. As you'll come to see, the course material is somewhat more descriptive than perhaps you're used to. This is intentional, as most research astronomers don't carry round mathematical proofs in their heads, but generally understand key concepts more descriptively. Because of this, I wanted to make sure you had a set of comprehensive notes to work from. Ideally, you'll read the notes prior to attending that lecture, and use the lectures to gain further clarification on concepts.

2.3 Assessment

- Exam: 70%; Answer three out of five questions.
- Coursework: 20%; Paper summaries – see MOLE for deadlines (10% each).
- Presentation: 10%; 15 minutes (+5 minutes for questions) on a galaxy evolution topic pitched to your peers – we'll arrange suitable dates between us.

2.4 Office Hours

Mondays between 12pm and 2pm.

Lecture 1:
Historical Introduction

Dr. James Mullaney

February 2, 2018

1 The dawn of Extragalactic research

While galaxies external to our own Milky Way have been observed for many centuries – the Magellenic Clouds and the Andromeda Galaxy are all easily observable with the naked eye – it wasn't until the bulding of the 100-inch telescope (i.e., ≈ 2.5 m diameter aperture) on Mt Wilson in 1917 that they were systematically studied. This was the first fully steerable large telescope and was immediately put to use to study galaxies external to our own.

At the time the 100-inch was built, debate still raged over whether galaxies – then referred to as "spiral nebulae" – were internal or extenal to the Milky Way. Edwin Hubble used the 100-inch to make the first distance measurements to these galaxies. He did this by using Cepheid variables, whose peak intrinsic luminosities are known to be tightly related to the period of their variation. By measuring the period of Cepheid variables, Hubble was able to calculate their intrinsic luminosity (L), and by measuring their peak flux (F) was able to determine their distance (r) using:

$$F = \frac{L}{4\pi d^2} \tag{1}$$

In 1924, Hubble published his results which demonstrated that the galaxy NGC 6822 lies at a distance of 214 kpc (7×10^5 light years; it is now known to be even more distant: 500 kpc) – well beyond the most distant of stars in our Milky Way – thus confirming its extragalactic nature.

Hubble continued his study of these (now confirmed) external galaxies, calculating the distances to many others. He also measured the velocity (v) at which these galaxies are moving away or toward the Milky Way by measuring the shift of emission/absorption lines in their spectra. In his 1929 study, Hubble reported that:

1. the vast majority of the galaxies he studied are *receding* from the Milky Way, i.e., have *redshifted* spectral lines; and
2. their velocity is proportional to their distance from us.

These two points have profound implications when combined with the Extended Copernican Principle (i.e., we – here meaning the Milky Way – hold no special place in the Universe). They imply that, no matter where you are in the Universe, (almost) all other galaxies will be receding from you at a velocity that is proportional to its distance from you. The only way this can be the case is if the Universe is expanding in all directions. Thus, this was the first clear sign that the Universe originated in a Big Bang (although it's not conclusive proof).

A further, perhaps more practical (but no less important) use of Hubble's 1929 result is that we no longer have to tediously measure the periods of Cepheids to measure distances to galaxies. Instead, we can make the far easier measurement of the galaxy's receding velocity (v) from its redshift (recall redshift: $z \approx v/c$ when $v \ll c$) and use Hubble's relation calculate the distance (r):

$$d = H_0 v \tag{2}$$

where H_0 is Hubble's constant, which is currently measured to be $67.6^{+0.7}_{-0.6}$ km s^{-1} Mpc^{-1}.

2 Extragalactic research today

Hubble conducted his groundbreaking research by studying just a handful of galaxies. Today, telescope surveys of the night sky have identified hundreds of millions of galaxies (of the trillions that are contained within the observable Universe). These galaxies display a vast diversity of shapes, masses, luminosities, stellar ages, metallicities etc. and reside in environments ranging from dense superclusters to isolated voids. The main goal of extragalactic research is to explain this huge diversity in galaxy parameters and the correlations between them.

3 What we need to explain

While a glance at a field of galaxies may give the impression that their properties are randomly distributed, there is, in fact, quite a lot of underlying order and a number of correlations between their various physical parameters. Much of extragalactic research is focussed on explaining what causes this order. Below, I briefly highlight some of the key properties of galaxies that we need to explain. We will consider each of these in more detail throughout the course.

- **Morphologies:** Perhaps the most obvious feature of galaxies is their shape, whether spiral, elliptical or irregular. But, why do galaxies have different shapes? Why aren't they all the same shape? And why those particular shapes? Why not cubes, pyramids, or dog-shaped?

- **Luminosity function:** Not all galaxies have the same luminosities. When we produce a histogram of galaxy luminosities – known as the galaxy *luminosity function* – we find that there are far more low luminosity galaxies than high luminosity galaxies. Furthermore, the galaxy luminosity function follows a characteristic shape, known as a Schechter function. Why don't all galaxies have the same luminosity? Why are there more low luminosity galaxies than high luminosity galaxies? And what causes the specific shape of the luminosity function?

- **Stellar populations:** Different galaxies are made up of different mixes of stellar populations. Furthermore, the mix of stellar populations depends strongly on the type of galaxy, with elliptical galaxies tending to have old, metal-rich populations, while spiral galaxies typically having younger stellar populations. Why don't all galaxies have the same mix of old and yound stars? And why do different types of galaxy contain different proportions of old and young stars?

- **Scaling relations:** As well as the link between stellar populations and galaxy morphologies, a number of other correlations exist between various galaxy properties. For example, there exists a tight correlation between the rotational velocity and the luminosity of spiral galaxies

(Tully-Fisher relation). There is a similar relationship between luminosity and velocity dispersion for Elliptical galaxies. Also, the colour of ellipticals (literally, whether they are bluer or redder) is also correlated with luminosity. There even exists a tight correlation between mass of the supermassive black hole that reside at the centres of galaxies, and the mass of their host bulges. Why aren't the various properies of galaxies simply randomly distributed amongst each other? What causes these correlations to exists?

- **Clustering:** Galaxies are not distributed evenly in space. Instead, they clump together in superclusters, clusters and groups. Furthermore, various galaxy properties are related to their cluster environment with, for example, large ellipticals preferring high density regions. Why do galaxies clump together? Why are they distributed in the way they are? And why do their various physical properties correlate with environment?

And all that is simply for today's galaxies! We can also ask how all of these different properties and correlations were different at different epochs of the Universe, to almost 13.6 billion years ago.

4 Studying galaxy evolution

If we wish to explain the physical properties of today's galaxies, it is important that we understand how they have evolved to their present state. There are two key means of achieving this:

- **The fossil approach:** Examining the detailed structures, stellar populations, abundance patterns of galaxies in the local Universe. For example, piecing together the star-formation histories of different galaxies to determine when their stars were put in place.

- **The high redshift approach:** Examining the properties of distant galaxies as a function of redshift/lookback time.

The high redshift approach works because of the finite speed of light. As we look to increasingly distant objects, the light from those objects has been travelling for longer and longer times. We therefore see those objects as they were when they emitted their light. For example, we see the Sun as it was about 8 minutes ago; the nearest star as it was about 4 years ago; the most distant stars in our galaxy at they were a few 10,000 years ago; the Andromeda galaxy as it was about 2.5 million years ago; the most distant galaxies as they were about 13 billion years ago.

If we can find galaxies in the distant Universe that we *believe* will evolve into the types of galaxies that we see today, then by studying those distant galaxies we gain insights into what today's galaxies likely looked like at earlier times. For example, we see that galaxies in the distant (and therefore early) Universe tend to form stars more rapidly than nearby (i.e., today's) galaxies. So, unless we live in a very peculiar region of the Universe, it is highly likely that nearby galaxies – including the Milky Way – went through an episode of rapid star-formation at earlier times. As it turns out, this is backed-up by the fossil record of nearby galaxies.

5 Evidence of galaxy evolution

There are a number of key pieces of evidence indicating that galaxies have evolved significantly over the past 13 billion years. Most of these have been established over the past half-century or so, and I briefly describe some of them below. First, however, we'll consider a philosophical argument that has been offered as evidence of an evolving Universe for almost 200 years.

5.1 Olber's paradox

Olber's paradox concerns the question of why the sky is dark. At first, it may seem obvious why the sky is dark. However, if the Universe is infinite in both size and age (and contains a roughly constant density of stars/galaxies), then the night sky should be bright.

This can be explained if we consider a square patch of sky say, 1 degree-by-1 degree. There will be very few "nearby" stars – at, say distance r – in this small patch, but because they are nearby, they will appear bright. As we increase r, the number of stars within that patch increases as r^2, but their individual brightness drops off as r^2. So, the sum of the light ($F_{\rm Tot}$) from all the N ($\propto r^2$) stars in the patch at a given r is constant:

$$F_{\rm Tot} = \sum_{i=0}^{N} \frac{L_i}{4\pi r^2} = N\frac{\bar{L}}{4\pi r^2} = Cr^2\frac{\bar{L}}{4\pi r^2} = k\bar{L} \tag{3}$$

where C is a constant of proportionality, $k = C/4\pi$ and $\bar{L}$ is the average luminosity (L) of the stars. Note that the second step comes from the definition of an average:

$$\bar{L} = \frac{\sum_{i=0}^{N} L_i}{N} \tag{4}$$

In theory, if we sum over all r in an *infinite* Universe, then we get an infinite amount of light in the patch. In practice, however, stars will start to block each other out, so in fact the whole sky will have the same surface brightness as a typical star. As such, if the Universe were infinite in age and size, we should expect the night sky to have about as the same surface brightness as the surface of the Sun! This paradox is solved if the Universe is either finite in size, age and/or evolving.[1]

5.2 Quasar number counts

Quasars are extremely bright, point-like astronomical objects. They are the brightest non-transient objects in the observable Universe. They are now known to be powered by matter accreting onto a supermassive black hole at the centre of a galaxy. Because of their extremely high luminosities, we can detect them from across the whole of the observable Universe. Indeed, for a long time, they were the most distant observable objects.

We will cover Quasars in a lot more detail later in the course. For now, it is suffice to say that because of their brightness, they provided our first insights into the very distant – and thus early – Universe. In doing so, they provided the first robust suggestions of an evolving Universe via their *number counts*.

Number counts are the flux equivalent of a luminosity function. Rather than a histogram of luminosities, "number count" is the term we use for a histogram of fluxes. It is somewhat more fundamental than a luminosity function, however, since it doesn't require an distance information (recall, flux, F, is simply what is measured by your telescope, whereas to calculate a luminosity we also require a distance, $L = 4\pi r^2 F$).

If the density of quasars within the Universe has remained constant with time (i.e., a non-evolving Universe), it is fairly easy to predict the numbers of quasars there should be within a given flux range (i.e., their number count). To do this, we'll use calculus to sum the number of

[1] If you're doing PHY406, this is just an alternative way to describe the same Olber's Paradox presented in that module.

quasars with a given observed flux, F, throughout the entire Universe. Consider a thin spherical shell of the Universe with radius r and thickness dr. The number of quasars, dN with *observed* flux F within this shell is:

$$dN = \rho(L) \times 4\pi r^2 dr \tag{5}$$

where $\rho(L)$ is the number of quasars of luminosity L per unit volume (i.e., this is the quasar luminosity function). However, to observe a flux F from a quasar of luminosity L, then r *must* be given by:

$$r = \sqrt{\frac{L}{4\pi F}} \tag{6}$$

Subbing this into Eqn. 5 gives:

$$dN = \rho(L) \frac{L}{F} dr \tag{7}$$

To integrate this to get N – the total number of quasars of observed flux F in *all* shells – we need to re-write dr in terms of dL and F by differentiating Eqn. 6:

$$\frac{dr}{dL} = \frac{1}{2} \frac{1}{\sqrt{4\pi LF}} \tag{8}$$

Subbing dr from Eqn. 8 into Eqn. 7 gives:

$$dN = \frac{1}{2} \frac{1}{\sqrt{4\pi}} \rho(L) \frac{L^{\frac{1}{2}}}{F^{\frac{3}{2}}} dL \tag{9}$$

Integrating both sides from $L = 0$ to $L = \infty$ (corresponding to $r = 0$ to $r = \infty$), and taking all non-L terms out of the integral, gives:

$$N = \frac{1}{4\sqrt{\pi}} F^{-\frac{3}{2}} \int_0^\infty \rho(L) L^{\frac{1}{2}} dL \tag{10}$$

and since it's a definite integral, it just integrates to a constant, irrespective of the form of $\rho(L)$ (provided that $\rho(L)$ is non-evolving, and thus independent of r), leaving:

$$N \propto F^{-\frac{3}{2}} \tag{11}$$

So, if the number density of quasars of a given L (i.e., the *luminosity function*, $\rho(L)$) does not evolve, then the number detected with a given flux, F, is proportional to $F^{-\frac{3}{2}}$. Usually, however, astronomers consider *differential* number counts:

$$\frac{dN}{dF} \propto F^{-\frac{5}{2}} \tag{12}$$

When we plot the actual numbers of *observed* quasars as a function of flux, we find that these numbers deviate significantly from a $F^{-\frac{3}{2}}$ law. The actual interpretation of this is not straight forward without further information (including the redshift distribution of the quasars), but it does imply that the quasar population *must* have evolved over the history of the Universe.

5.3 The "V over $V_{\rm Max}$" test

Later, more formal methodologies were adopted to test for an evolving galaxy population. One of the first of these was the $V/V_{\rm Max}$ test.

Consider a galaxy of luminosity L, detected in a galaxy survey with flux F and measured to have a distance from us of r (corresponding to a volume $V = \frac{4}{3}\pi r^3$). Lets say that the very lowest flux we can detect with our survey is $F_{\rm Min}$ (so, by definition $F > F_{\rm Min}$). So, the very furthest away *we could have* detected that galaxy (of luminosity L) is therefore:

$$r_{\rm Max} = \sqrt{\frac{L}{4\pi F_{\rm Min}}} \tag{13}$$

corresponding to a limiting volume of:

$$V_{\rm Max} = \frac{4}{3}\pi r_{\rm Max}^3 \tag{14}$$

So, for every galaxy in our survey we can calculate a $V/V_{\rm Max}$ value. In a non-evolving Universe, we would expect all galaxies to be evenly distributed, so:

$$\left\langle \frac{V}{V_{\rm Max}} \right\rangle = 0.5 \tag{15}$$

since, on average, we'd expect half of all galaxies to be within $\frac{V_{\rm max}}{2}$ and the other half to be between $\frac{V_{\rm max}}{2}$ and $V_{\rm Max}$. However, what we actually find is that $V/V_{\rm Max} > 0.5$ for the most luminous galaxies. This would only happen if there were a higher density of luminous galaxies close to the edge of the survey volume (i.e., close to r_{Max}), meaning their individual $V/V_{\rm Max}$ values are greater than 0.5. This result therefore implies there was a higher density of luminous galaxies in the distant, i.e., early, Universe compared to today.

5.4 Quasar luminosity function

As our ability to measure redshifts improved, we gathered distances for increasing numbers of quasars. With this information, astronomers were able to measure their luminosities and, in turn, determine the quasar luminosity function at different epochs. In doing so, they found that the luminosity function of quasars was, indeed, different at earlier epochs, thereby confirming the results inferred from quasar number counts. Again, it was found that the density of quasars was higher in the early Universe compared to today.

5.5 Butcher-Oemler effect

In 1978, Butcher and Oemler published a paper in which they had measured the colours of galaxies in two distant clusters of galaxies, one at redshift 0.39 (corresponding to 4.2 billion years ago) and another at redshift 0.46 (corresponding to 4.8 billion years ago). What they found was that there were a higher fraction of blue galaxies in the cores of the distant clusters compared to similar clusters in the local Universe (i.e., today). Blue colours in galaxies are normally associated with ongoing star-formation, so this result was interpreted as evidence of increased levels of star-formation in cluster cores in the early Universe compared to today. While at the time, they didn't know why this was the case (we've got a better idea now; see later lectures), it was clear evidence that galaxy clusters were different at earlier times; i.e., they have evolved.

5.6 Galaxy number counts

As telescopes grew in size and detectors became more senstive, astronomers were able to start detecting fainter and fainter *galaxies*, rather than just very bright quasars. As they reached the faintest galaxies – some of which will low luminosity, nearby galaxies, but most will be very distant, luminous galaxies – they found that their number counts were *not* consistent with a non-evolving Universe. Instead, they could only match the observed counts of faint galaxies if the density of blue, star-forming galaxies was higher in the early Universe compared to today.

6 Key learning objectives for L1

- What two groundbreaking discoveries did Hubble make that changed our understanding of galaxies and the Universe?
- List and understand the five key properties of galaxy populations that we need to explain.
- Know what the terms "luminosity function" and "number counts" refer to, and know the difference between them.
- Know, in general terms, what a quasar is and understand their importance to early investigations of galaxy evolution.
- Understand why studying distant galaxies and quasars can provide insights into the early Universe.
- Know what is meant by the "fossil approach" and the "high redshift approach".
- Be able to explain Olber's paradox.
- Know the five key pieces of evidence of galaxy evolution prior to the mid-nineties.

Lecture 2:
Challenges and recent advances

Dr. James Mullaney

February 2, 2018

1 Introduction

In the last lecture, we saw how we could exploit "lookback" time to gain a view of the Universe as it was at earlier times. In other words, using our telescopes effectively as a "time-machine" to look into the Universe's past. Unfortunately, it's not quite as easy as this, and in this lecture we'll cover how observing over such vast distances hampers our attempts to study galaxy evolution.

Unfortunately, the material in this lecture is a bit "dry", but it's really important that you understand it to grasp the caveats of our understanding of galaxy evolution. I'll try my best to make it interesting.

2 The challenges of studying the distant Universe

In this section, we'll cover the main difficulties we face in using observations of the distant Universe to study galaxy evolution. In the following section, we will cover how astronomers are overcoming these problems (although some will never be solved entirely).

2.1 Dimming with redshift

The first, most prominent and most obvious challenge in studying the distant (i.e., high redshift) Universe is that all sources are subject to geometrical dilution. In other words, the more distant an object, the fainter it appears. In a geometrically flat (i.e., Euclidian) Universe, all sources will dim at a rate of $1/d^2$ with increasing distance d. This is because as light travels from a source, it gets "shared-out" over a spherical surface with area $4\pi d^2$, so there's less light per unit area as d increases.

In addition to "classical" geometric dilution, at high redshifts other factors start to have a detrimental effect. Because distant galaxies are travelling away from us at high velocities, their photons are redshifted and thus reduced in energy by a factor of $1/(1+z)$ (where z is the measured redshift). Also, because of time dilation, photons from a distant galaxy arrive more slowly by *another* factor of $1/(1+z)$. So the energy per unit time per unit area (i.e., flux) received by our telescope decreases by a factor of $1/4\pi d^2(1+z)^2$. This means that for a galaxy at $z=2$, its flux is almost a factor of *9 times* lower than what we'd expect from classical geometric dilution!

Things get even more strange (and worse!) when we consider how the *surface brightness* of a galaxy (i.e., the measured flux per unit area) changes with redshift. This is because angular distance also has a $1/(1+z)$ dependence (proving this within the realm of cosmology, so is beyond the scope of this course). What this ultimately means is that the surface brightness of a galaxy decresease as $1/4\pi d^2(1+z)^4$. This means that the surface brightness of a resolved galaxy at $z = 2$ is *81 times* times lower than what we'd expect from classical geometric dilution of an unresolved source!

Together, this all means that distant galaxies *are even fainter than you'd expect*, which makes studying them in a systematic, robust manner *incredibly* hard.

2.2 The K-correction

Another difficulty associated with observing distant galaxies is literally caused by the redshifting of light due to their high receding velocities.

Imagine you wish to compare the luminosities of two galaxies – one at low redshift, say $z = 0.1$, and another at $z = 2$ – in a given optical filter, let's say the r-band (i.e., around wavelength $\lambda = 658$ nm). You've been really careful to take all the above effects due to redshift dimming into account when converting their observed fluxes into luminosities. However, an additional problem is that, because of redshift, the r-band filter samples two different *rest-frame* wavelengths in the two galaxies: 658 nm$/(1 + 0.1) = 598$ nm for the $z = 0.1$ galaxy, and 658 nm$/(1 + 2) = 219$ nm. In the case of the $z = 0.1$ galaxy, you're measuring the luminosity of the rest-frame red part of its spectrum, and in the case of the $z = 2$ galaxy, you're measuring its rest-frame ultraviolet luminosity. Because galaxies tend to be quite faint in the ultraviolet, this often has the effect of making the $z = 2$ galaxy appear dimmer than if you were measuring the red part of its spectrum.

One way to account for this is to apply a *correction* to the ultraviolet luminosity you measure in the $z = 2$ galaxy – known as a *k-correction*. However, to do this, you must know the ratio of red-to-ultravoilet luminosity for the $z = 2$ galaxy. Today, this is usually achieved by observing the galaxy in multiple bands and using spectral templates to interpolate to the desired band. However, in the past, astronomers had to rely on average K-correction factors for different types of galaxies.

Alternatively, you could observed the $z = 2$ galaxy in a band centred at 598 nm $\times (1 + 2) =$ 1794 nm, which is in the near-infrared region of the spectrum, as it would probe the rest-frame 519 nm part of the $z = 2$ spectrum.

2.3 Morphological K-correction

One feature of the K-correction is that it can actually effect the observed *shape* of a galaxy.

Most galaxies, and especially star-forming spirals, do not contain uniforms distributions of stellar populations. Instead, spirals often contain old stellar populations in their cores, and have clumps of star-forming regions containing very young stars in their arms. Between these clumps are older stars.

Imagine we're looking at the same two galaxies as in the previous subsection – one at $z = 0.1$ and another at $z = 2$. By some amazing coincidence, it turns out that they have exactly the same intrinsic structure – spirals with an old bulge at the centre and clumps of star-formation in their arms, with populations of older stars between the clumps. For the $z = 0.1$ galaxy we can see all this structure perfectly in our optical filters: the blue filters pick up the star-forming clumps, while the red filters pick up the central bulge and the "sea" of old stars between the clumps.

For the $z = 2$ galaxy, however, the red light from the bulge and inter-clump regions is shifted into the near-infrared, so we won't see it in our optical bands. Also, because old stellar populations don't emit very strongly in the blue part of the spectrum, there won't be much light from the bulge or inter-clumps shifted into the optical-red filter. Instead, the optical-red filter will be dominated by the shifted blue light from the clumps of young stars. So, all you'll see in your blue and red filters are small clumps of star-formation, which you could mistake as lots of small galaxies (since the inter-clump and bulge are too faint to see).

The way to solve this problem is to try to observe high-redshift galaxies in longer-wavelength filters to compensate for the red-shifting of the light. Ideally, you'd like to use filters which mean you observe the same rest-frame wavelengths at different redshifts, but sometimes this isn't possible (due to technology constraints, or atmospheric absorption).

2.4 Progenitor bias

The primary aim of extragalactic research is to determine how today's galaxies got to look the way they do. The justification behind studying galaxies in the high redshift Universe is that it gives us a chance to "look back in time" at galaxies in earlier stages of development. However, it is important to remember that we aren't literally looking at today's galaxies at earlier times (that would require a time machine!). Instead, we have to try to identify those galaxies in the early Universe that we *think* will evolve into certain types of galaxies today. For example, what type of galaxy observed in the distant, early Universe will eventually evolve into the type of massive, elliptical galaxies we seen in the local Universe today (this is, indeed, a big question in current astronomy research)? This is made more difficult because the Universe (and the galaxies in it) have changed a lot over the past 13 billion years. Making the wrong assumptions about progentior galaxies can strongly bias our interpretation of observations, and thus our conclusions.

As an example, we may *think* that the most massive galaxies at $z = 2$ evolve into the most massive galaxies today. It may, however, instead be lots of smaller galaxies at $z = 2$ that merge together to form the most massive of today's galaxies. If we simply assume the first case is true, we may erroneously focus on the most massive galaxies at $z = 2$ and generate (possibly incorrect) evolution theories about how they evolve.

2.5 Cosmic variance

On very large scales (i.e., Gigaparsec scales) the Universe is very homogeneuous; i.e., a given cube 10^9 pc on its side will look the same as another. On scales of $\lesssim 100$ Mpc (Mpc=Megaparsec), however, the Universe is very much non-homogeneous, clumping into clusters and groups with large voids in between. So, a given cube a few Mpc across may contain a very dense cluster of galaxies, wheres another randomly chosen volume of the same size may contain a void.

With the properties of galaxies known to depend on the density of their environmnent (i.e., cluster, group, void), it is important that we survey large enough volumes of the Universe to ensure that we sample galaxies across a wide range of environments.

For example, if we compared galaxies in the (nearby; $z \approx 0$) Virgo supercluster against those in a small-area survey that focussed on a high-redshift void (mostly including what we call "field" galaxies), then we wouldn't be comparing like-for-like. To ensure we're comparing like-for-like, we'd need to use a larger-area survey that contained at least one high-redshift cluster.

2.6 Assumed cosmological model

This one's a bit of an odd one...

In order to convert measured fluxes into intrinsic galaxy properties such as luminosity, mass or even diameter, we need to assume a cosmological model. These cosmological models include things like the shape of the Universe and how fast it is expanding (due to dark energy).

Cosmological models are described by a set of parameters, such as the Hubble constant (H_0), the matter density parameter (Ω_m; i.e., what fraction of the Universe is mass), the dark energy density parameter (Ω_Λ; i.e., what fraction of the Universe is dark energy). One way these parameters can be determined is by "calibrating" the Hubble diagram; using independent distance measurements and comparing them against redshift. For example, if we know the intrinsic luminosity of an object and measure its flux, we can obtain a redshift-independent distance measurement.

In the past, giant elliptical galaxies were used as "standard candles"; i.e., we assumed that all giant elliptical galaxies all had the same intrinsic luminosity. So measuring their fluxes gave us a distance measurement which could then be used to determine cosmological parameters. However, if other astronomers unknowingly used these cosmological parameters to measure the luminosities of giant elliptical galaxies, then they would measure them as all being the same (i.e., it would be circular). Cue headline:

All giant elliptical galaxies are the same luminosity at all redshifts shocker!!!

This highlights that we need to be very careful in making sure that any results we obtain when comparing distant galaxies to nearby galaxies are not simply a consequence of the cosmological parameters and how they were derived.

2.7 A lack of knowledge of today's galaxies

Finally, a bit of a surprising one...

We don't actually understand today's galaxies (i.e., nearby ones) as well as we'd like to. This represents somewhat of a problem if we're trying to understand how galaxies have evolved to their present state since we don't really fully know what the "present state" is.

Bear in mind that even "local" galaxies are very distant in real terms. We can't pick them up and look at them at different angles. Some parts of them are osbcured by dust. Often we can't resolve individual stars in even nearby galaxies. And many aspects we think we do understand are actually highly dependent on uncertain models (e.g., star-formation).

As such, while we understand local galaxies better than more distant ones, the former certainly don't represent a well-defined "end point" for evolutionary studies. Indeed, should our understanding of local galaxies change dramatically, it could force a consderable rethink of our evolutionary models.

As an example, it was only very recently found that the number of dwarf galaxies around the *Milky Way* and other nearby galaxies is much lower than our evolutionary models predicted there should be. As a consequence, our galaxy-evolution models are being re-thought to try to rectify this inconsistency with the real Universe.

3 Overcoming challenges

As telescope, detector and computing technology has developed, astronomers have increasingly mitigated the above challenges. Some have even been overcome entirely.

3.1 Telescope size

Perhaps the most obviously impressive development has been the dramatic increase in telescope size over the past century, starting with the $100''$ telescope on Mount Wilson in 1919. Today's largest telescopes have mirrors 10m in diameter, and thus have 16 times the collecting area of $100''$. This has helped us to observe more distant galaxies, helping to tackle the challenge of the rapid fall off in flux of galaxies at high redshifts.

3.2 Detector technology

A huge development in astronomy occured when astronomers progressed from taking observations by eye to using photographic plates in the mid 19^{th}. Not only are photographic plates less subjective, but they all allow us to *integrate* (i.e., collect) light over longer periods of time than the eye. Photographic plate are, however, hugely inefficient, so a second revolution occured in the 1980s when astronomers started using CCDs as detectors. Although perhaps not as immediately obvious, CCD technology has had a greater impact on modern astronomy than telescope size. Modern CCDs can be more than 90 times more efficient than photographic plates (compared to a 16 times increase in telescope collecting area in the last $\sim$100 years). As a result, today's top-end *amateur* CCD-equipped telescopes can be more senstive than the $100''$ when it was equipped with photographic plates.

The other key advancement in detector technology has been the development of detectors that can measure light across almost the whole electromagnetic spectrum, from the longest wavelength radio to the highest-energy gamma rays. By being able to observe in many different wavelengths, we mitigate the problem of K-correction, since we can often use a waveband that corresponds to the appropriate rest-frame wavelength we wish to compare. There is the caveat that some wavelengths are blocked by our atmosphere, but space-based missions – while expensive – have opened up those parts of the electromagnetic spectrum.

3.3 Wide-field astronomy

As well as being able to detect fainter galaxies, telescope technology has also developed the ability to observe over wide areas. For example, the SDSS telescope has a field-of-view of about 6 square degrees. This means it can observe a whole hemisphere of the sky in about 3500 pointings. That might seem a lot, but each pointing is only viewed for about 1 minute, meaning it can survey the entire observable sky in 3500×1 minute $= 58$ hours ≈ 8 nights (that's for one filter, it has five filters in total).

Such large fields-of-view help to mitigate the "cosmic variance" problem, as it allows us to sample large volumnes of the Universe in a short time. For example, it takes only about 45 pointings (i.e., 45 minutes) for the SDSS telescope to survey a volume of 1Gpc^3 out to $z = 1$, which, as we have seen, is large enough to sample the full diversity of galaxy environments from superclusters to voids. However, the problem remains that, since the SDSS is a relatively small (2.5 m) telescope by today's standards, it will only detect the brightest galaxies at $z = 1$, and miss the faintest ones. In other words, as with any astronomical survey, it is *incomplete*.

Another major development in wide-field astronomy has been the development of multiplex spectroscopy. Early spectrographs could only take the spectrum of one object at a time by placing a single slit on the object. During the 1990s, however, multi-slit and multi-fiber systems were developed to allow astronomers to take the spectra of lots of different objects in the same field-of-view at the same time. Multi-slit systems work by placing a "slit-mask" in the field-of-view which blocks out the light from most objects, but allows the light from a few to pass through to the spectrograph. Multi-slit systems can typically take the spectra of a few tens of objects at a time, and you have to be careful of "spectral overlap" (where two slits are too close to each other and their spectra overlap). Even more advanced and capable are multi-fiber systems (as used by SDSS for its spectrographic mode), where individual fiber-optics are places at the positions of all the objects you want to take the spectra of. The light then travels down the fiber to the spectrograph. These can typically take the spectra of hundreds of objects at the same time and avoid the problem of spectral overlap by arranging the fibers on the spectrograph-end so they don't overlap.

3.4 Independent cosmological parameters

Rather than relying on galaxies as standard candles or standard rulers, we now have galaxy-independent standards that we use to deteremine the values of cosmological parameters,

Perhaps the most well-known standard candles are Type-1a supernovae. These occur when a White Dwarf star accretes sufficient mass from an orbiting partner star (in a binary system) to ignite carbon-fusion in its core which rapidly leads to the star going supernova. The reason that we can use Type-1a supernovae as standard candles is that theory strongly suggests that their peak luminosities are always the same since they are always triggered by the same process. The other great benefit of Type-1a supernovae is that they are extremely bright, so can be oberved to high redshifts ($z \sim 2$) and thus provide a long baseline for calibration.

While Type-1a supernovae are a very good way of obtaining independent distance measurements and thus determining cosmological parameters, it is important to have other methods so we can check they are consistent. Another important independent distance measurement comes from "Baryon Acoustic Oscillations". As the name suggests, these are regular, periodic density fluctuations in baryonic matter of the Universe; a sort of "echo" of the Big Bang. The lengh of these fluctuations is about 500 million light years and, as such, can be used as a standard ruler. We can use surveys of galaxies (since they are, of course, made of baryons) to measure the length of these fluctuations at different redshifts to get an independent distance measurement, which then leads to independent cosmological parameters.

4 Learning objectives from Lecture 2

Well, you made it! Here are the learning objectives:

- Understand the problems associated with studying the high-redshift Universe to investigate galaxy evolution.
- Understand how technological advancements have helped to mitigate some of these problems, and eradicate others.

Lecture 3:
The theory of galaxy formation

Dr. James Mullaney

February 5, 2018

1 Introduction

Before looking in detail at piecing the evolution of galaxies together from observations, we'll start with considering how we simulate galaxy evolution, since much of our theories are based on such models.

2 Broad theories of galaxy formation

There are two broad theories of galaxy formation:

- **Monolithic collapse** (Eggen, Sandage & Lynden-Bell 1962)[1]: Here, a single, large cloud of gas collapses to form a single, massive galaxy. Density fluctuations in the cloud lead the denser regions to collapse to form a loose cluster, or halo, of stars. As the cloud collapses further, it conserves angular momentum; the dominant component of rotation wins-out over the others and a disk is formed. Gas continues to collapse within this disk to form more stars. In this scenario, all of the mass that eventually forms the galaxy is present from the outset in one big "lump".

- **Hierarchical galaxy formation** (White & Rees 1978): In this scenario, larger galaxies are built-up over time by the repeated merger of smaller galaxies. The smaller galaxies are created by the collapse of gas clouds to form collections of stars. These then coalesce to form a larger galaxy, dragging their gas content with them, enabling some star-formation to continue in the resulting larger galaxy. This idea of hierarchical merging fits well with our models of a Universe that is dominated by collisionless dark matter. It is currently the most widely-accepted model of galaxy-formation and evolution.

While the hierarchical model of galaxy formation is currently our best theory to explain the observed properties of galaxies, it needs a lot of "tweaking" in order to accurately reproduce what we see. Many of these tweaks *are not* physically motivated, and are instead set by hand to reproduce observations. For example, among of the most important tweaks are the efficiencies of various "feedback" processes, such as how light from young stars heats their surrounding gas, thus preventing it from forming further stars. The physics governing these feedback processes are still too complex to

[1] On researching this, I came across a report that suggests Olin Eggen was a bit of a kleptomaniac. After his death in the last '90s, scores of rare books that had gone missing from the Royal Greenwich Observatory three decades earlier were found in his office. He had always insisted that he never had them.

model, so theoretical astronomers have to estimate their efficiencies until their models reproduce the properties of galaxies in the real Universe. As such, although our models are very effective at reproducing observations, it *does not* mean we fully understand galaxy formation/evolution.

3 Methods of modelling hierarchical galaxy evolution

There are two main methods of modelling galaxy formation and evolution: **hydrodynamic** and **semi-analytic**. As you may expect, both have their benefits and pitfalls (otherwise, we'd only use one). In the following two subsections, we'll explore both these methodologies.

3.1 Full hydrodynamical simulation

Of the two methods, this is probably the easiest to understand conceptually. Hydrodynamic simulations try to model the motion of dark matter, stars and gas under the influence of physical forces. For dark matter and stars this is relatively easy since they are both *collisionless*; in other words, they only interact via their gravity (the chances of two stars colliding are extremely small due to their tiny size compared to the separation between them). As such, dark matter and stars can be modelled using "straight-forward" N-body simulations.

Gas, on the other hand, is *highly dissipational* (i.e., it dissipates gravitational potential energy easily in the form of radiation). Gas also readily absorbs energy, which affects its ability to collapse and form stars. As such the behaviour of gas is governed by many more processes than gravity: e.g., heating, cooling, pressure, etc. To get a sense of how much more complicated modelling gas is compared to considering only gravity (as for stars and dark matter), consider the complex motion of air from a hairdryer to the simple orbits of the planets. To be modelled fully, gas needs to be treated *hydrodynamically*.

There are two main types of hydrodynamic models: **Lagrangian** and **Eularian**. The easiest way to think about the difference between the Lagrangian and Eularian aproaches is to consider their most common examples: **Smoothed Particle Hydrodynamics (SPH)** (Lagrangian) and **Mesh models** (Eularian). In SPH, the gas is treated as a population of particles that interact with other gas particles via a *smoothing length*. The larger the smoothing length, the more the gas interacts with itself (in SPH codes, stars and dark matter are treated as particles with a zero smoothing length). Taken to the extreme, an *ideal* SPH model would consider every individual molecule or atom of gas as a particle. In reality, however, we typically have to assume individual gas particles that are many parsecs across and contain many solar masses of gas (our computers are a long way from being able to simulate every atom of gas in a galaxy).

In Mesh models, the modelled volume is split up into very small cells, and *continuity equations* are used to calculate how much gas enters a given cell from its neighbours, and how much leaves the same cell to its neighbours. If one is different from the other, the density and pressure of the gas in that cell must change. The smaller the cells, the more precise your model. However, decreasing the size of the cells in one dimension by a factor of two increases the number of cells by a factor of $2^3 = 8$, with a corresponding increase in the number of calculations required (and thus the total time it takes for the model to run).

In each type of model, the physical conditions of the gas (such as temperature, density, pressure) are calculated at each *timestep*. The shorter the timestep in the model, the more accurate it is, but the more calculations are needed (and thus, again, an increased running time). After each timestep, the properties of the gas represented by a given particle or contained within a given cell are compared to their neighbours and calculations are made to determine how it interacts with its neighbrouing particles or cells. For example, heat may be passed from one particle/cell to another, there may be bulk motion from one particle/cell to its neighbours etc. Also, "prescriptions" (which may be well-defined, or simply guesses) are used at each timestep to calculate how much the gas should be heated, cooled, and how much will have collapsed to form stars, or fallen into a black hole.

3.2 Semi-analytic models

Semi-analytic models (SAMs) take the philosophy of replacing the most complicated aspects of hydrodynamic simulations with simple analytic expressions. As such, they tend to be far quicker than hydrodynamic models, but rely on far more assumptions.

As with hydrodynamic models, SAMs are not restricted to astrophysics; they are also commonly used to model the Earth's climate. In general, they use relatively simple N-body simulations to model a core component of the physical situation, then use analytical expressions to model the detailed processes on top of this underlying core. In galaxy evolution, they exploit the widely accepted concept that the dominant form of matter in the Universe is dark matter. The argument goes that, since dark matter dominates, we can model the dark matter – which is relatively easy, since it is non-interacting – and then use analytic expressions to populate this dark matter with gas and stars.

One key benefit of the SAM approach is that the big N-body calculation, i.e., that in which the dark matter is modelled, only needs to be performed once. After that, any number of different analytic expressions can be used to generate populations of galaxies (in a relatively short time, since they are analytic), which can then be compared against observations of the real Universe. For example, one model could be that every "blob" (or, "halo") of dark matter contains a galaxy with a stellar mass (i.e., the sum of the mass of all its stars; M_*) that is 1% of the mass of the dark matter halo (M_{Halo}). The analytic expression for this would be:

$$M_* = 0.01 \times M_{\mathrm{Halo}} \tag{1}$$

As you can see, this is a really simple expression, but it would result in a population of galaxies with given masses. Even if our simulation contained a billion dark matter halos, it would only take a few seconds for our semi-analytic model to the masses of the galaxies. If, on comparison against the real Universe, we then realised that this was a bad model, we could easily try 2% (without having to run the whole N-body dark matter model again) and see if that were any better. Today's SAMs are very sophisticated, with anayltic expressions used to populate dark matter halos with gas, to control the cooling of this gas, to control how stars form, to control feedback processes, etc., but the principle remains the same.

In what follows, we will go through the steps needed to generate a more typical SAM of galaxies in the Universe:

3.2.1 Choosing initial conditions

Any simulation needs a starting point: a set of initial conditions at time $t = 0$ that the simulation can then evolve to the next time step. In cosmological SAMs, the initial condition is set by the distribution of dark matter 379,000 years after the Big Bang. This time is chosen as it is the point at which the Cosmic Microwave Background (CMB) was emitted. The CMB has been well-studied by satellites such as COBE, WMAP and Planck. As a result, the temperature fluctuations of the CMB are now very well-defined. These temperature fluctuations give a representation of the matter distribution in the early Universe, which cosmological SAMs use as their starting point. On measurement, the probability of a given density fluctuation δ is given by a Gaussian field:

$$p(\delta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\delta^2}{2\sigma^2}\right) \tag{2}$$

SAMs use this equation to populate their models with an initial distribution of dark matter (i.e., at each point in your $t = 0$ model allocate a density selected randomly from a Gaussian distribution). Since simulations are limited by computing power, they must have a limited resolution; in the most widely used cosmological SAM – the Millenium Simulation – this resolution corresponds to each dark matter "particle" having a mass of a billion solar masses.

3.2.2 Generate a merger tree

Once the initial conditions of our dark matter are set, we can start our simulations running. Since we're dealing only with dark matter at this stage, we only have to contend with two factors: (a) the expansion of the Universe due to the Big Bang and, later, dark energy (these are all specified by our cosmological parameters, which are now pretty well-defined) and (b) gravity.

At the end of the simulation (i.e., often, but not always, when $t =$ today), all of the clumps of dark matter (known as dark matter halos) are identified, and all the dark matter particles that end up in each halo are traced back through the simulation to their starting point. This creates what is referred to as a "merger tree", since at $t = 0$ there are lots of separate particles (i.e., branches) that first merge to form mid-sized halos (i.e., limbs) and then large halos (i.e., tree-trunks).

The merger-tree represents the end-point of the N-body dark matter simulation. It describes the full merger-history of all dark matter particles in the simulation, and is all we need if we want to use analytic expressions to create a population of galaxies. This is because, since we're assuming that dark matter dominates over everything else, we don't have to worry about the detailed motions or positions of the halos when using analytics to populate them with baryons (i.e., gas, stars, galaxies etc). Everything after this stage is the "analytics" part.

3.2.3 Cooling of gas in dark matter halos

With the merger trees in-hand (and remember, they provide everything we need for the analytics stage), we can use analytic expressions to population them with gas. First, we use a prescription to allocate gas to a halo; this is usually as simple as saying that each halo gets its "fair share" of baryons. In other words, each halo is given a mass of baryons proportional to its dark matter mass (usually $M_b \approx 0.15 M_{\rm DM}$; i.e., the ratio of baryonic matter to dark matter).

Next, the gas is assumed to fall toward the centres of their dark matter halos. As it does this, the gas gets shocked and heated to the virial temperature, given by:

$$T_{\rm vir} = \frac{1}{2} \frac{\mu m_H}{k} \frac{G M_{\rm Tot}}{r_{\rm vir}} \tag{3}$$

where m_H is the mass of the Hydrogen atom, k is the Boltzmann constant, $M_{\rm Tot}$ is the total halo mass, and $r_{\rm vir}$ is the virial radius (here, the radius within which a cloud of gas is destined to form a galaxy).

Then, the hot, shocked gas is assumed to cool from the inside outwards with a disk of cool gas forming due to the conservation of angular momentum. How quickly the gas cools is also defined analytically and can depend on many factors, such as the temperature and composition (i.e., metallicity) of the gas. It can therefore be quite complicated (and thus subject to a lot of tweaking). Of course, after the first round of star formation in the model, we can include analytic expressions that control how the metallicity of the gas changes due to stellar reprocessing, which will affect later gas cooling. Thus, the model can become highly self-interacting, or *dynamic*.

3.2.4 Star-formation

Since we now have cooling gas, we should think about the natural consequence: star-formation. Precisiely how gas clouds collapse to form (populations of) stars remains one of the biggest unanswered questions in astrophysics and is an active area of research (just ask Simon Goodwin, whose research focusses almost entirely on this area). Without a complete theory of star-formation, semi-analytic models rely on empirically-defined relationships between the amount of available cold gas and the rate of star-formation. The most well-known of these is the Schmidt-Kennicutt relationship:

$$\Sigma_{\mathrm{SFR}} = (2.5 \pm 0.7) \times \Sigma_{\mathrm{gas}}^{(1.4 \pm 0.15)} \tag{4}$$

where Σ_{SFR} is the star-formation rate per unit area (in $\mathrm{M}_\odot$ yr^{-1} pc^{-2}) and Σ_{gas} is the surface-density of gas (in $\mathrm{M}_\odot$ pc^{-2}). So, with our analytic prescription for how much cool gas there is in a galaxy, we can use Eqn. 4 to calculate the rate of star-formation and build-up our galaxy stellar masses in our simulation over time.

3.2.5 Feedback processes

When the first large, cosmological SAMs were run, it was found that they tended to make galaxies that were far more massive than we observe in the real Universe. This implied that there were some processes in the real Universe that was preventing galaxies from forming too many stars that were not being included in our models.

One of ways that gas can be prevented from forming stars is by heating it, so it stays too warm to collapse. Another way is to expel it from the potential well at the centre of the dark-matter halo. It was found that both these effects could be achieved by including "feedback" processes – so called because it is the actual act of gas cooling that causes these feedback processes to be triggered.

There are a number of potential feedback processes, but the two most important are due to supernovae and Active Galactic Nuclei (**AGN**). In the case of supernovae, once the first stars in the simulation have been formed, an analytic expression is used to determine what fraction will go supernova at a given time. These supernovae heat, and potentially expell, the surrounding gas, preventing it from forming new stars. It therefore provides *negative feedback* since the process of star formation actually acts to prevent further star formation (via their end-of-life supermovae).

In the case of AGN, gas cools and falls to the centre of the dark matter halo, where some of it will ultimately fall toward the supermassive black hole at the centre of the nascent galaxy. As it does so, the gas forms an accretion disk, which heats up and releases vast amounts of energy (we'll cover AGN in much more detail later in the course). In some SAMs, this energy is used to heat up and expell some of the surrounding gas, thus preventing it from forming stars (again, it is a form of negative feedback).

Feedback processes are some of the most uncertain and difficult-to-model features of models (SAMs and hydrodynamic). As such, they rank as some of the most argued-upon features of modern astrophysics. They almost certainly exist in the real Universe, but we don't have a very good grasp of how they work, and under what circumstances.

3.3 The benefits and pitfalls of Hydro models and SAMs

Now we've had a fairly comprehensive introduction to the two main types of cosmological galaxy evolution models, we can consider the benefits and pitfalls of each.

Hydro models

Benefits:

- Track the detailed evolution of dark matter, gas and star motions as galaxies evolve via gas accretion and mergers.

Pitfalls:

- The resolution is relatively poor ($\sim 20 - 100$ pc)
- At smaller scales than the resolution, we still have to make major assumptions about the physics ("sub-grid physics").
- While fewer in number, these sub-grid assumptions are akin to the analytic prescriptions in SAMs.
- As such, while it is less *degenerate* than SAMs, there is still room to tweak sub-grid parameters to match observations.
- Take a long time to run, so it is difficult to "try-out" lots of different sub-grid prescriptions to investigate how they alter the outcomes.

Semi-analytic models:

Benefits:

- Once the dark matter N-body simulation has been done, it is relatively inexpensive (i.e., quick) to try-out different analytic prescriptions.
- If you're not too interested in the detailed physics, they can give you a useful "mock universe" to plan observations etc.

Pitfalls:

- Don't track in detail the motion of gas or stars (since stars form from gas).
- Huge amount of free parameters (and more can be easily added), which can make them highly degenerate (i.e., the effects of changing one analytic parameter can be countered by changing another).
- Some of the analytics to describe the physics are highly uncertain (and some are just added to get the "correct result" without much physical intuition.

4 Learning objectives

In this lecture, we've considered the main *methods* of modelling galaxy formation and evolution. The key learning objectives are:

- Have and knowledge understanding of the main models for galaxy formation and evolution.
- Appreciate the differences between the semi-analytic and fully hydrodynamic approaches to modelling galaxy evolution.
- Have knowledge of the main steps involved in forming galaxies in SAMs: initial conditions, an "N-body" dark-matter simulation, merger trees, gravitational accretion and cooling of gas, shock heating, star formation and feedback.
- Have an understanding of the benefits and pitfalls of SAMs vs. hydrodynamic simulations.

Lecture 4:
Spectral synthesis and star-formation indicators

Dr. James Mullaney

February 10, 2017

1 Introduction

In this lecture, we will cover how we actually measure the properties of galaxies. This is critical to our understanding of galaxy evolution since, if we can't measure their properties, we can't build up a theory of how they evolve.

2 Measuring galaxy properties

In astronomy, we can only study objects from a distance, normally via the electromagnetic radiation they emit (and, as of 2015, gravitational waves). We can't handle them, look at them from a different angle etc. etc. So how do we measure their physical properties such as mass, composition etc.?

In the case of stars, we have the luxury of binaries that can help us to measure their masses, while the fact that stars are "simply" spherical balls of gas also helps (though, I'm sure Profs. Crowther and Dhillon would disagree!). In the case of galaxies, we don't have such "ideal" scenarios, so we're simply limited to analysing the combined light from their populations of stars.

But, how do we take the light from a galaxy and convert it into a mass, metallicity, star-formation rate, etc.? One of the most important and widely-used means is by spectral synthesis.

2.1 Spectral synthesis

The basic idea of spectral synthesis is to build a model that represents the light from all the stars in a galaxy. If that model accurately reproduces the spectrum (ideally) or colours (more usually) of the galaxy, then we can infer various galaxy properties from that model. For example, if the spectrum of a galaxy was perfectly reproduced by combining the spectra of a billion identical, sun-like stars, then we'd be able to say that the mass of the galaxy is a billion solar masses. In reality, however, it is far more complicated than that since, as you'd expect, galaxies are made up of stars spanning a wide range of masses, ages and metallicities. So, to produce these "synthetic" spectra astronomers have to combine the spectra of lots of different types of stars. Thankfully, we have a recipe to help us constrain the range of stars we include in our models.

2.1.1 Initial mass function

The first part of the recipe requires us to choose an Initial Mass Function (IMF). As with most such "functions" we've met so far, it's really more of a histogram. It is the histogram of stellar masses that a cloud of gas produces the moment it collapses to form stars. It tells us that for every single 10 $M_\odot$ star, there are about a hundred 1 $M_\odot$ stars, and about ten thousand 0.1 M$\odot$ stars. Why is this important? Because the spectrum of a star (and how it evolves, see later) depends hugely on its mass, so it is important that our model includes the relative numbers of stars of different masses.

Finally, while there is some uncertainty in the shape of the IMF and whether it was different at earlier times in the Universe (indeed, whole conferences are dedicated to this topic), it seems that the IMF is fairly universal. As such, astronomers typically assume a constant IMF when doing extragalactic studies. However, since there are various different IMF (e.g., the Salpeter IMF, the Kroupa IMF, the Chabrier IMF), it is important that if you want to compare physical properties you use the same IMF throughout. This is also true if you want to compare your results to another study – it is often the case that astronomers need to "convert" their results to a different IMF. Usually, however, this is simply a case of multiplying the physical parameters (e.g., mass, star formation rate) by a constant factor.

2.1.2 Stellar evolutionary tracks, or isochrones

The IMF gives us the relative numbers of stars of different masses at time $t = 0$. However, all but a tiny fraction of stars in a galaxy will have an appreciable age (i.e., millions to billions of years old). So, to be able to model the stellar population of a galaxy, we need to know how the stars (whose masses are given by the IMF) evolve over time.

Thankfully, how a star evolves is largely dictated by its mass, with a lesser dependency on metallicity. Since we know the mass of every star in our model (from the IMF), we can model how it will evolve across the Herzsprung-Russel (HR) diagram. Of course, more massive stars will travel across the HR diagram more quickly than low mass stars, since they "live fast, die young". A line joining together all the different stars on the HR diagram at a given time, t, is called an *isochrone*; i.e., *iso* - meaning the same, *chrone* - meaning time. At this stage, we can also specify a metallicity dependence.

Provided we know our stellar evolutionary tracks, we will know the positions on the HR diagram of all the stars on our model at any given time.

2.1.3 Stellar spectra

At this stage, we can imagine our population of stars at time t as being represented by a whole load of points on the HR diagram. There will be lots and lots of low mass stars that will have barely evolved, and a handful of high mass stars that will have evolved very quickly (indeed, some of the most massive may even have undergone supernova). To generate the synthetic spectrum of the population, we simply have to sum the spectra of all the individual stars represented by those points.

Thankfully, we have observed spectra for stars across almost all the HR diagram, so we can use real, observed spectra in most cases. In some parts of the HR diagram where there are few stars, however, we sometimes have to resort to synthetic spectra due to a lack of observed spectra.

2.1.4 Star formation histories

Up to this point, we've only considered a single, instantaneous burst of stars. However, it is unlikely that a galaxy will form all its stars in one single burst. Instead, a galaxy will form its stars over a protracted period of time. This will result in a different synthetic spectrum. For example, if we have an extended period of star-formation, then massive young stars (which are hot and therefore emit strongly in the UV and blue part of the spectrum) will continue to be produced. Therefore, the synthetic spectrum of a continuous burst will stay bluer for longer than an instantaneous burst.

How do we model a continuous period of star-formation? Well, all we need to do is take lots of instantaneous bursts going off one after another. We can even modulate the size of each burst (i.e., how many stars are produced in a given burst) according to a *star formation history.* For example, we could model a constant, continuous episode of star formation by a set of bursts, one after another, which all produced the same number of stars. Alternatively, we could model an exponentially declining episode as a series of bursts, each containing:

$$N_{\mathrm{Stars}}(t) = N_{\mathrm{Stars}}(t = 0)\exp\left(-\tau t\right) \tag{1}$$

stars at time t (where τ is a constant that described how quickly the rate of star formation falls away with time). Then, we simply treat each burst separately as described above.

2.1.5 A gardening analogy

Sometimes an analogy can really help to explain an idea; I like this gardening one I came up with for generating synthetic spectra (it doesn't contain caterpillars!).

Let's say I have a garden and some seeds. It's March, I'm just about to plant my seeds, and I want to have an idea of what my garden will look like in June, July and August.

We can think of the IMF as the relative numbers of different types of seeds. Say 100 daisies, 50 marigolds, 20 sunflowers. The equivalent of stellar evolutionary tracks would be how quickly each variety grows; sunflowers grow really quickly, marigolds more slowly. Finally, the stellar spectra would be the colours of the flowers: yellow sunflowers, orange marigolds, white daisies. Given the relative numbers of seeds, their growth rates and their colours, I can predict what colours I will have in my garden during each month of the summer. The whole, *fabulous* floral effect would be the synthetic spectrum.

Finally, if I want to, I can plant lots of different sets of seeds each week to prolong the flowering season. This would be the equivalent of the star-forming history.

2.2 Model fitting

With modern computers, we can genarate a population of stars and produce its synthetic spectrum extremely quickly (i.e., in a few milliseconds). As such, given an observed spectrum/photometry of a real galaxy, what astronomers typically do is model a whole range of different metallicities, ages, and star-formation histories and see what combination reproduces the data the best. This iterative trial-and-error is simply a form of model fitting. Indeed, astronomers often take a χ-squared minimisation approach to find the best-fitting synthetic spectrum to observed spectra or photometry.

2.3 Caveats to using spectral synthesis

In principle, spectral synthesis and model fitting is relatively straight-forward. As such, it's relatively easy to use spectral synthesis to measure galaxy properties. However, just getting a measurement is only part of the effort. We must also consider how reliable these measurements are. We must consider what could affect the reliability of results from spectral synthesis and model fitting. Here are considerations:

- **Dust**: Since blue light is more readily absorbed by dust, if there is intervening dust then it will change the shape of the observed spectrum. Dust can make a population of stars look older than it is, because the blue light from the young stars is absorbed.
- **Uncertainties in post-MS tracks**: Any uncertainties in the paths that stars take across the HR diagram as they evolve will result in uncertainties in the resultant synthetic spectrum. In particular, there is a huge debate on how post-asymptotic giant branch stars evolve (especially if they're pulsating) because they are *really* bright so can contribute a lot to a galaxy's total spectrum.
- **Incomplete spectral libraries**: While we have taken spectra for many millions of stars, there remains some parts of the HR diagram that remian poorly sampled. In particular, since most of the stars in our region of the Milky Way have similar metallicities, we only have a few spectra of non-solar metallicity stars (and these are biased to more luminous stars). Further, since very massive stars don't live very long, they tend to be rare, making it difficult to find ones to measure their spectra.
- **Uncertain IMF**: We've already touched upon this point. Uncertainties in the IMF introduce uncertainties in the range of stellar masses, which subsequently introduces uncertainties into measured physical properties.
- **Uncertain star-forming histories**: One can imagine that a galaxy can have multiple episodes of star-formation during its evolution, possible with different regions having different histories. While, in priciple, we could model such complex histories, in reality there is not enough data to discriminate between such complex models. As such, we typically assume simple models (single burst, continuous, exponentially falling), which introduces uncertainties.

3 Measuring integrated star formation rates

One of the most important measurements we can make of a galaxy is its star formation rate (SFR). This effectively tells us how quickly a galaxy is growing, and is thus a fundamental aspect to understanding how today's galaxies have formed. With SFRs, we can ask questions like: When did galaxies grow fastest? Are today's galaxies growing faster or slower than previously? What causes galaxies to grow more quickly?

By modelling the stellar content of a galaxy, spectral synthesis *will* give us a measure of the SFR of a galaxy. Often, however, we don't have to go to such lengths, as there are other ways to measure accurate SFRs that rely on observations in just one band (i.e., UV, infrared) or a single emission line.

All of the single-band/emission line measures of SFR rely on one key fact: **that the hottest, most massive stars die young**. The most massive stars ($> 5M_\odot$) will live for "only" around 100 million years (compared to 10 *billion* years for the Sun). So, if we can count the number of massive stars in a galaxy, then we can calculate the average rate of *massive* star formation in that galaxy over the past 100 million years (which is pretty instantaneous for a galaxy!). However, the average rate of massive star formation *is not* the same as the actual SFR; recall that the IMF tells us that for every massive star formed, there are many, many more lower mass stars produced. So, to calculate the true SFR, we take the number of massive stars, then multiply it by the ratio of the total mass of stars per the total number of stars with $> 5M_{\odot}$, i.e.:

$$\mathrm{SFR} = \frac{N_{\mathrm{M>5M_{\odot}}}}{10^{8}\mathrm{yr}} \times R_{\mathrm{Stars}}(\mathrm{IMF}) \tag{2}$$

where, $N_{\mathrm{M>5M_{\odot}}}$ is the *observed* number of stars with masses greater than 5 $M_{\odot}$ and $R_{\mathrm{Stars}}(\mathrm{IMF})$ is given by:

$$R_{\mathrm{Stars}}(\mathrm{IMF}) = \frac{\text{Total mass of all stars}}{\text{Number of } > 5\mathrm{M_{\odot}} \text{ stars}} \tag{3}$$

which is calculated from our chosen IMF. As you should be able to see, the SFR is in units of $M_{\odot}$ yr^{-1}.

So, to calculate a galaxy's SFR (assuming a given IMF), all we need to do is measure how many stars with $M > 5$ $M_{\odot}$ it contains. How can we do that?

3.1 The UV luminosity of a galaxy

Massive stars are very hot, so they produce a lot of ultraviolet radiation. In fact, in the absence of an AGN, massive stars are pretty much the only sources of UV radiation in a galaxy. As such, measuring the UV luminosity of a galaxy provides a measure of the number of > 5 $M_{\odot}$ stars in a galaxy. With this in mind, astronomers have calculated a simple conversion from UV luminosity to total SFR:

$$\mathrm{SFR}\ (\mathrm{M_{\odot}\ yr^{-1}}) = 4 \times 10^{-41} L(\mathrm{FUV})\ (\mathrm{erg\ s^{-1}\ A^{-1}}) \tag{4}$$

As with all the other SFR indicators in this section, this takes into account the increased luminosity of high mass stars relative to the Sun, and for the corrections mentioned in §3 and summarised in Eqns. 2 and 3, i.e., it is the "true" rate of star formation in the galaxy, averaged over the past ~ 100 Myr.

The main problem with using the UV emission to measure SFRs is that it is readily affected by dust. Attempts can be made to correct for dust extinction, but they rely on various assumptions and even a small amount of dust can strongly affect measured UV flux.

3.2 The Hα luminosity of a galaxy

Because hot, massive stars emit strongly in the UV part of the spectrum, they emit a lot of photons with wavelengths shortward of 912 Å. Photons shortward of this wavelength are energetic enough to ionise Hydrogen. When this ionised hydrogen recombines with free electrons, the electrons cascade through the atomic levels, emitting photons as they drop to lower energyies. One of the strongest lines produced by this process is the Balmer-alpha (i.e., Hα) emission line at 6563 Å, which corresponds to an electron transition from $n = 3$ to $n = 2$. Measuring the luminosity of the Hα line therefore provides a measure of the number of young stars in a galaxy, from which we can calculate the SFR:

$$\mathrm{SFR}\ (\mathrm{M_{\odot}\ yr^{-1}}) = 7.9 \times 10^{-42} L(\mathrm{H\alpha})\ (\mathrm{erg\ s^{-1}}) \tag{5}$$

Ionising Hydrogen requires particularly energectic photons, which only the most massive, hottest stars are capable of producing. Indeed, stars with masses below about 10 $M_\odot$ are unable to produce sufficient numbers of high energy photons to ionise Hydrogen. This means that Hα traces only the very hottest, most massive stars that live for *only* around 10 Myr. As such, the Hα line gives a measure of the SFR averaged over only the past 10 million years. It is pretty much the most "instantaneous" measure of SFR that we can get.

There are a few of caveats that come with using Hα to measure SFR. Firstly, like the UV, Hα is affected by extinction due to dust. However, it is not as badly affected by the UV and is easier to correct-for using the "Balmer decrement" (the intrinsic ratio of the Hα to Hβ lines is 3:1, so if we measure, say, a 4:1 ratio, we know it is being extinguished by dust, and can correct for it). A second caveat is that some of the ionising photons may escape the gas cloud without hitting a H-atom, meaning we'll measure a lower SFR rate. Thirdly, since AGN are also prodigious producers of Hα, it is difficult to use this to measure the SFR of galaxies hosting AGNs (although these are the minority of galaxies).

3.3 The infrared luminosity of a galaxy

Most stars form out of gas and dust. When the massive stars start to shine, their heat warms the dust that surrounds them and the dust re-radiates this warmth in the form of infrared radiation. As with Hα and UV emission, it is only the most massive (> 5 M$_\odot$) stars that are hot enough to warm the surrounding dust to produce this infrared radiation. As such, by measuring the infrared luminosity arising from a galaxy, we can get a measure of the number of young stars it contains, and thus its SFR:

$$\text{SFR (M}_\odot\ \text{yr}^{-1}) = 1.8 \times 10^{-44} L(\text{FIR})\ (\text{erg s}^{-1}) \tag{6}$$

Since stars with masses greater than about 5 M$_\odot$ are capable of heating the dust, the infrared provides a measure of a galaxy's SFR averaged over the past ~ 100 Myr.

The benefit of using infrared wavelengths is that, unlike UV and Hα is it largely unaffected by dust obscuration. However, since infrared wavelengths are readily absorbed by the Earth's atmosphere, we need to use expensive space missions to observe at these long wavelengths. The constraints on telescope size that this introduces, combined with the long wavelengths, mean that infrared observations typically suffer from poor spatial resolution.

4 Learning objectives

In this lecture, we have covered how we actually measure the physical properties of galaxies. This is crucial if we want to be able to build-up a coherent theory of galaxy exolution. The key learning objectives you should take from this lecture are:

- Have an understanding of how spectral synthesis modelling can be used to determine the masses, SF histories and metallicities of galaxies.

- Have an understanding of the main ingredients and uncertainties in spectral synthesis modelling.

- Appreciate of the main techniques used to determine the integrated SFR of galaxies and their pros and cons

Lecture 5:
Studying galaxy evolution via the fossil record

Dr. James Mullaney

February 17, 2017

1 Introduction

Until now, we've largely considered investigating galaxy evolution by exploiting the light travel-time to study galaxies in the early Universe. We will continue to cover this "high-redshift" approach later in the course. In this lecture, we will consider the alternative (but complementary) approach: the fossil record. Here, we study nearby galaxies in a lot of detail to try to understand how they have been "built".

2 Early type – or, Elliptical – galaxies

Elliptical galaxies are the big, boring ones without spiral arms. When Hubble came up with his galaxy "tuning fork" diagram in the early 20^{th} century, it was thought that elliptical galaxies evolved into spiral galaxies (i.e., left-to-right on the diagram), which is why they are also known as "early-type" galaxies. We now know that elliptical galaxies *do not* evolve into spirals but, unfortunately, the name has stuck and so it's still used extensively today.

2.1 Stellar populations of elliptical galaxies

As well as their shape, elliptical galaxies in the local Universe share similar properties. They tend to be quite massive (stellar masses of $\sim 10^{11}$ $M_\odot$ or above) and and are red in colour, which suggests an old stellar population (i.e., the massive, hot, blue stars have had time to evolve and die, leaving just the less massive, cool red ones). When we analyse their spectra using spectral synthesis (see Lecture 4), we indeed find that the spectra of the majority ($> 70\%$) of local ellipticals are best modelled by a population of old stars with ages > 8 billion years. This suggests that today's massive ellipticals have remained mostly as they are today for over half the age of the Universe.

2.2 The colour-magnitude relation for elliptical galaxies

When we study elliptical galaxies in more detail, we find there are other consistencies within the class. While elliptical galaxies are, in general, redder than their spiral counterparts, it's not the case that they are all exactly the same colour. Indeed, some are "redder" than others.[1] It turns

[1] In astronomy, we define a colour by the relative amounts of light in two bands, which means we can actually quantify how "red" or "blue" a galaxy is.

out that the redness of an elliptical galaxy is not random. Instead, it is tightly correlated with its total luminosity such that more luminous elliptical galaxies are redder.

Since the luminosity of a galaxy is related to how massive it is (i.e., in general, the more stars a galaxy contains, the brighter it is likely to be), this also means that more massive elliptical galaxies are redder. Astronomers have confirmed this by plotting the colour (i.e., "redness") of elliptical galaxies against their *velocity dispersion*, or σ, and find that redder galaxies have higher velocity dispersions. Elliptical galaxies are supported against gravity by the velocity of their stars (more on this later), so their velocity dispersion gives a direct proxy measure of their mass; the higher their velocity dispersion, the greater their mass.

The tightness of the relation between colour and mass/luminosity for elliptical galaxies in a given cluster suggests that they were all formed at roughly the same time (in that particular cluster). The argument goes that if elliptical galaxies in a given cluster were formed at different times, then they would show much more diversity in their colours (i.e., there'd be some that formed their stars recently, so would be blue, and some that formed their stars a long time ago, so would be red). However, if they're all virtually the same age, then why haven't they all got *the same* colour; why is there a colour dependence on mass/luminosity? That can be explained in terms of metallicity. More massive galaxies hold onto their reprocessed gas more easily than low mass galaxies because their gravity is stronger (they have greater "binding energies"). Due to their stronger gravities, it is more difficult for stellar winds and supernovae to push this enriched gas out of more massive galaxies. Reprocessed gas is metal rich, so when it re-collapses to form more stars, it forms metal-rich stars, which are redder than metal-poor stars of the same mass.

In summary, the slope of the colour-magnitude/luminosity/mass relation for elliptical galaxies is due to metallicity effects. The tightness of the correlation is because all ellipticals in a cluster formed at roughly the same time.

2.3 Light profiles of Elliptical galaxies

As well as measuring the total flux and luminosity of galaxies, we can also measure their light *profiles*, i.e., how their surface brightness changes as a function of distance from their centre.

2.3.1 Extended light profiles

When we model the light profiles of elliptical galaxies, we find that the light profiles *outside their central regions* fall off as roughly $R^{-\frac{1}{4}}$:

$$I(R) = I_e \exp\left(-7.67\left(\left(\frac{R}{R_e}\right)^{\frac{1}{4}} - 1\right)\right) \tag{1}$$

where I_e is the surface brightness at the effective radius of the galaxy, R_e, defined as the radius which contains half of the total light from a galaxy.

On closer inspection of the light profiles of elliptical galaxies, astronomers found that a more general form of Eqn. 2 gave a better fit to the light profile:

$$I(R) = I_e \exp\left(-b\left(\left(\frac{R}{R_e}\right)^{\frac{1}{n}} - 1\right)\right) \tag{2}$$

which is known as a *Sersic* profile. Typically, $n = 4$ to 6.

2.3.2 Core light profiles

When we look at the very centres of elliptical galaxies, astronomers find that some depart from the Sersic profiles. They do this in such a way that their light profiles *flatten* within their centres (within about 1 arcsec for nearby examples, corresponding to the central few 100 pc). These are known – rather ambiguously – as *cores*. Thus, ellipical galaxies fall into two groups, those with *cores* and those without (i.e., their light profiles keep rising toward the centre). In general, galaxies with cores have higher luminosities than those without.

Today, it is thought that the reason for the deficit of light in the centres of ellipticals with cores is a consequence of how they formed. It is thought that massive ellipticals are formed by major, gas-poor – or, dissipationless – mergers (i.e., they don't contain much gas, so the kinetic energy of the merger is kept within the galaxy in the form of stellar kinematics, rather than being dissipated away). During such mergers, the central supermassive black holes at the centres of the two merging galaxies fall toward the central regions of the final galaxy. As they do so, they "throw-out" stars via gravitational sling-shot. This is known as "scouring" and is an effective way of removing stars from the central regions of a galaxy post-merger.

2.4 Disky vs. Boxy ellipticals

As well as core ellipticals, early type galaxies also separate into two separate classes according to their morphologies. Some ellipticals have a slightly "rugby-ball" shape and are known as "disky", whereas others are "squarer" in appearance and are known as "boxy". Quantitatively, we distinguis between the two by performing an angular Fourier series fit to the light profile of elliptical galaxies and measuring the fourth term:

$$\Delta r(\theta) \approx \sum_{k \gtrsim 3} a_k \cos k\theta + b_k \sin k\theta \tag{3}$$

If $a_k > 0$, then it is a disky elliptical, otherwise it is a boxy elliptical.

When we consider the two types of galaxies separately, it turns out they differ in more ways than their morphologies. While all elliptical galaxies rotate more slowly than spiral galaxies, disky ellipticals tend to rotate more quickly than their boxy counterparts. Indeed, boxy ellipticals are almost entirely supported by random motions, rather than bulk rotation. Also, disky ellipticals tend to be less luminous (and thus less massive) than their boxy counterparts. Boxy ellipticals also tend to have "cores".

All this suggests that the two types of elliptical galaxies formed via different paths. It is now thought that disky ellipticals formed at high redshifts ($z \gtrsim 1$) as a result of the merger of two gas-rich galaxies. Since gas is dissipational, it would have lost energy, and collapsed to form a rotating disk, producing lots of stars in the process. Once the gas had finished producing stars, what is left is a rotating elliptical - rotating because of the imprint of the gas, and elliptical because the orbits of the original stars would have undergone a degree of randomisation during the merger.

Boxy ellipticals are also thought to have been produced by merging galaxies, but in this case are thought to have been produced by gas poor ("dry"), dissipationless mergers. The stars were already in-place prior to the mergers (and thus are very old; estimated to have formed at $z > 3$). In this situation, there is very little gas to collapse into a rotating disk, so pretty much all the energy of the merger goes into randomising the orbits of the stars. This creates a galaxy that is entirely supported by the random motions of their stars, and is boxy due to the lack of rotation. As explained above, the cores are due to "scouring" due to black holes falling toward the cores of the merger remnant. Unlike in disky ellipticals in which new stars are formed by the gas-richer merger, in boxy ellipticals the "scoured" stars are not replaced by star formation.

3 Spiral galaxies

The other main type of galaxy on the Hubble diagram are spiral galaxies. Since the Milky Way galaxy is a spiral, we can learn a lot about spiral galaxies by studying our own Galaxy.

3.1 The strucure of the Milky Way

While the stars in the Milky Way are close compared to other galaxies, establishing its structure is hampered by the fact that we sit inside it. By measuring the distances to various stars within the Milky Way we can, however, map-out its structure, in the same way that if you're sat *inside* a wharehouse, you'd have a pretty good idea of the *outer* dimensions of the wharehouse simply by looking around you. In the case of the Milky Way, however, this is hampered by dust along our line-of-sight, which blocks out the light from some regions. This problem can be mitigated by observing at longer wavelengths which more easily penetrate the dust clouds.

When we look at our Galaxy from within using penetrating infrared wavelengths, we see that the Milky Way is a disk galaxy with a bulge. Judging by the prominence of the bulge (and comparing to other external galaxies) it is likely that the Milky Way is a "late type" spiral galaxy (i.e., it is to the right of the Hubble tuning fork), probably an Sc or SBc galaxy. The latter classification (i.e., SBc) is based on the shape of the Milky Way's asymmetric bulge, which sugests it has a weak bar.

But that's just the overall appearance. Studying the fossil record of the Milky Way involves much more detailed analyses of the stellar content of the Galaxy, such as:

- **Kinematics**: This is the motion of stars within the Milky Way relative to the Sun or, more generally, the Galactic Centre.
- **Metal Abundances**: Since stars generate metals as they evolve and, eventually, die in supernovae, the metal content of stars provide information on the ages and star-forming histories of stellar populations.
- **Ages**: Calculating the ages of indvidual stars is hard, but for populations, we can use the relative numbers of a population across the HR diagram (akin to the technique using in spectral synthesis) to calculate accurate ages of populations.
- **Precise positions**: Mapping out the above properties as a function of position relative to the Galactic centre provides clues as to how the Milky Way (or, for that matter, other galaxies) was built-up.

By considering all the above, we now know that the Milky Way is made up of different populations of stars which make up four different components of the Galaxy: the *Thin Disk*, the *Thick Disk*, the *Bulge*, and the *Halo* populations.

3.1.1 The Thin Disk

The thin disk is made up of relatively young, metal rich stars (known, confusingly, as Population I stars. At the location of the Solar System, this disk rotates *in bulk* around the centre of the Milky

Way at a speed (relative to the centre) of about 220 km s^{-1}. This disk is about 50 kpc in diameter (outer edge to outer edge), but has a scale height of just 325 pc (i.e., $\rho(z) \propto e^{-z/325\ \mathrm{pc}}$, where ρ is the density of stars and z is distance perpendicular to the disk). The ratio of thickness to diameter of the thin disk is roughly the same as a vinyl LP or a CD; it is *very* thin!

The Milky Way is still continuously forming stars at a rate of about 1 $\mathrm{M}_\odot$ yr^{-1}, most of which takes place within the thin disk. The thin disk contains a lot of gas and dust from which these stars are being made. Today this disk is thought to be what's left of the dissipational collapse of a large gas cloud that formed the fledging Milky Way about 8 billion years ago.

3.1.2 The Thick Disk

In addition to the Thin Disk, there is a Thick disk which, as you may expect, has a larger scale height ($\approx$1000 pc), but which has roughly the same diameter. The thick disk is made up of older, more metal-poor stars (Population II) than the thin disk. Thick disk stars orbit slightly more slowly around the Milky Way's centre – at around 180 km s^{-1} – compared to thin disk stars. As they rotate, however, they also oscillate up and down throughout the thick disk, so at any given time, some of the thick disk stars are *within* the thin disk. However, the distance between two stars in the thin disk is so large that the likelihood of collision between thin and thick disk stars is tiny.

It is thought that the thick disk is what is left over from when the very young Milky Way collided with a smaller galaxy around 8 billion years ago (i.e., just as the natal gas cloud was collapsing to form the thin disk). The stars in the then collapsing thin disk were "thown out" by this collision, which formed the "puffed up" thick disk. Because kinetic energy was tranferred from the galaxy collision to partially *randomize* the stellar motions, this process is known as "heating" (elliptical galaxy have very "hot" stellar kinematics because they are almost entirely help-up by random motions). Because there is little gas within the thick disk (other than at the point it is cospatial with the thin disk), it does not form stars, so the thick disk stellar population provides a snapshot of what the thin disk was like very early-on in the Milky Way's history.

3.1.3 The Bulge

The bulge of the Milky Way is roughly spherical in shape, with a radius of about 3 kpc. It is formed from metal-rich, Population I stars, the rotation velocity of which is proportional to their metallicity. The stellar density of the bulge is significantly higher than the rest of the galaxy with, on average, 5×10^4 stars per cubic parsec.[2] That's just the average, the stellar density profile of the bulge falls off as $\approx r^{-2.2}$, so it is even more dense in the centre.[3]

The bulge is thought to have a very complex formation history. Some of the stars in the bulge are likely to have formed early on in the history of the Milky Way from the natal cloud of gas. Then, later, the successive accretion of satellite galaxies likely heated the bulge up (helping to make it the shape it is) and introduced more metal-rich stars.

[2] This means that the average distance between two stars in the bulge is about 0.03 pc, compared to the 1.3 pc between the Sun and our nearest star, Proxima Centauri. Yet, 0.03 pc is still about 120 times larger than the radius of Pluto's orbit. Yep, space is *big*.

[3] You should be able to take this density profile and use the above numbers to calculate the average density of the central 10 pc.

3.1.4 The Halo

Finally, we have the low-density spherical halo of stars that surrounds the entire Milky Way. This consists of very old, metal poor stars, that orbit the milky way very slowly (at about a charectristic velocity of about 40 km s^{-1}. The density of the halo drops off as r^{-3}.

It is believed that many of the halo stars are the remnants of the tidal disruption of satellite galaxies that ultimately merged with the Milky Way. These stars were formed early-on in these satellite galaxies, then were tidally stripped as the galaxy orbited and, eventually, fell into the Milky Way. There is also the prospect that some of the halo stars were formed out of the natal gas cloud that would eventually collapse to form the Milky Way.

3.2 Streams around the Milky Way

As we have seen, the merger of satellite galaxies can explain a lot of the bulk features of the Milky way. Today, there is significant evidence that the Milky Way has cannibalised many smaller, satellite galaxies over the past few billions of years. The most striking of such evidence are the "Fields of Streams", which are streams of stars that loop around the Milky Way.

As a smaller satellite galaxy orbits and spirals into a larger satellite like the Milky Way, the differential gravitational pull on one side relative to the other causes it to get stretched-out and ripped apart. The resulting "stellar debris" forms impressive loops around the galaxy, which are known as "tidal streams" (tidal forces are those created by differential gravitational pulls; the tides on the Earth are caused by the Moon's gravitational pull being different on one side of the Earth compared to the other).

3.3 Is the Milky Way normal?

If we're going to use the Milky Way as an example of how spiral galaxies have formed, we'd better check that it is typical of the population. It turns out, in fact, that the Milky Way is somewhat atypical of Spiral galaxies since, as far as we can tell, it falls below the Tully-Fisher relationship (i.e., the correlation between orbital velocity and the luminosity of a spiral galaxy). It is thought that this may be because the Milky Way has actually undergone *fewer* merger events compared to other spirals, which has maintained its high rotation speed relative to its mass (i.e., other spirals are "hotter" and have more support from random motions). Despite this, the Milky Way isn't *drastically* different from other spirals, so it can give us a decent "first-order" view of how spiral galaxies have formed.

4 The local galaxy population

Going beyond the Milky Way, we can get a better understanding of galaxy evolution *in general* by surveying galaxies in the local Universe. When we do this, we find that the galaxy population forms a bi-modal distribution in terms of their colours. There is a population of red galaxies – known as the "red sequence" when plotted on a colour-magnitude plot – and blue galaxies –known as the "blue cloud". Red-sequence galaxies are mainly ellipticals and tend to be more massive than blue cloud galaxies (which are mainly spirals). Their colours give a clue to their star-forming activities – galaxies in the blue cloud are star-forming, whereas red sequence galaxies tend to be gas-poor and "dead".

4.1 The population fossil record

With the availability of spectra for tens of thousands of nearby galaxies, astronomers can perform spectral synthesis on the whole local population to work out their star-forming histories. In doing so, they have been able to piece-together the average star formation rates of the Universe throughout its history.

For example, say we perform spectral synthesis for a single galaxy – galaxy A – and find it formed a quarter of its stars in a burst 1 Gyr ago, another quarter of its stars were formed in a burst 5 Gyr ago, and the remaining half was formed continously over the past 10 Gyr. We could then plot the star formation rate of this galaxy A as a function of time – it would have a two sharp spikes of star-formation superimposed on a continuous low-level. Then, we could consider the next galaxy – galaxy B – and add its history of star-formation to our plot. Then move onto the next, and the next, each time adding the star-forming histories to our plot. Eventually, if we did this for all the nearby galaxies, we'd have a plot of the total star-forming history of the local Universe. If we then assume that we don't live in a special place in the Universe (i.e., the Extended Copernican Principle) then we can apply this result to the whole Universe, i.e., we can say it is a reasonable measure of the star-forming history of the Universe.

When astronomers do this, they find that the star-forming history of the Universe peaked at about redshift 2, corresponding to roughly 10 billion years ago. Since this time, it seems that the rate of stellar production in the Universe has slowed significantly – by a factor of > 10. We'll explore why this is the case later in the course.

Finally, since we have pieced together the global star forming history from individual galaxies, we can also explore the star-forming histories of galaxies split into various different subcategories (e.g., what's the global star-forming histories of spirals vs. ellipticals? or for galaxies of different mass?). A very important result that we get when we do this is that the rate of star formation for today's most massive galaxies peaked at earlier times than less massive galaxies. In other words, today's massive galaxies started to form *first*, with less massive galaxies forming later. This process is known as **Cosmic Downsizing**.

5 Learning objectives for Lecture 5

- Knowledge of what detailed observations of nearby galaxies (the fossil record) tells us about the evolution of both spiral and elliptical galaxies
- An appreciation of how less detailed (statistical) studies of large samples of galaxies in the local Universe aid our understanding of galaxy evolution
- Familiarity with the concept of cosmic downsizing

Lecture 6:
Survey Astronomy

Dr. James Mullaney

February 16, 2017

1 Introduction

In the previous lecture, we learned that by studying nearby galaxies in great detail we can gain insight into how these galaxies were formed. In this lecture we'll consider the opposite approach: how surveys of thousands – or even millions – of galaxies can provide clues to their evolutionary history.

2 The philosophy of galaxy surveys

The guiding principle of extragalactic surveys is that by measuring the properties of whole populations of galaxies, we can gain insight into how they have formed and evolved. You should think of it as an alternative approach, yet highly complementary, to detailed studies of individual galaxies.

An analogy I use is studying the causes of heart disease. One way scientists can study the causes of heart disease is via dissection, i.e., the detailed study of an individual's heart to see what has caused the blockage of arteries, such as fatty build up. However, another way to investigate the causes of heart diease is to survey a large group of people – some with heart disease, others without – to assess what lifestyle choices may cause heart disease. Both are perfectly valid ways of studying the causes of heart disease and give consistent answers: that a poor diet rich in saturated fats (which causes the fatty blockaged) plays a role in causing heart disease.

The principle is the same in astronomy. For example, if the fossil record tells us that a lot of stars were put in place about 10 billion years ago, then we should see lots of strongly star-forming galaxies when we survey the Universe at $z = 2$ (corresponding to about 10 billion years ago). However, one of the great benefits of survey astronomy over the fossil record is that we can study galaxies in a statistical sense, identifying correlations and patterns that we otherwise wouldn't be able to uncover by studying just a few galaxies in detail.

2.1 What are galaxy surveys?

In the loosest terms, a galaxy survey can be described as simply a sample of galaxies that satisfies given selection criteria. For example, we could survey all the galaxies above a given stellar mass within 100 Mpc of the Milky Way. Or, we could survey all galaxies above a given luminosity limit within a given redshift.

For the purposes of this lecture, however, we will focus on "Blank-field surveys". These are samples of galaxies that are detected in flux-limited observations of a contiguous patch of sky. They are obtained by literally pointing a telescope at a blank piece of sky, making a set of obervations to a given integration time, and detecting as many galaxies as possible in that patch of sky. The sensitivity of the telescope plus the total length of integration dictates the *flux limit* of the survey.

How large a patch of sky you choose to observe, and for how long to integrate, is determined by your science goals and is known as your *survey strategy*.

3 Survey design

Due to the large amount of resources required to undertake a galaxy survey, a lot of thought goes into *designing* the survey. Things to consider are: the area and depth of the survey (i.e., the survey strategy); where to survey on the sky; what wavelength you want to survey in; is photometric data sufficient, or do we require spectroscopic information as well?

3.1 Survey strategies

As with any type of survey (election polls, consumer surveys etc), what we are ultimately trying to achieve with a blank field survey is a representative sample of galaxies. The more representative that sample is of the whole galaxy population, the better. Ideally, we'd survey the whole night sky to extremely *deep* (i.e., sensitive, in terms of flux limt) levels. However, such an approach requires an unfeasible amount of resources, especially in terms of oberving time. Thankfully, we're helped-out by the Extended Copernican Principle, since it tells us that (on large scales) "the Universe looks the same in all directions". So, to obtain a representative sample of galaxies doesn't *necessarily* require us to survey the whole sky, just a patch of sky large enough to capture the full diversity of galaxies. This is just the same as polling people: pollsters don't ask the entire voting population of the UK their opinions, but rather aim to survey a sample that is *as representative as possible* of the entire population.

But, how do astronomers decide on what "large enough" is, and how *deep* (i.e., sensitive, in terms of flux limit) their survey needs to be? That depends on what science questions you want your survey to address. Since very luminous galaxies are extremely rare, we have to survey very large *areas* of sky to obtain a representative sample; indeed, some of the most luminous galaxies and quasars are *only* seen in all-sky surveys. However, since they are very luminous, we don't need a very sensitive survey to detect them, so all-sky surveys only need to be relatively shallow to identify lots of very bright galaxies/quasars.

If, by contrast, the aim is to study more typical, less luminous galaxies, then such a large survey area isn't needed since there are lots and lots of "normal" galaxies even within a small patch of the sky. Instead, the problem is now that these galaxies are much less luminous, so require a deeper survey.

An analogy I like to help illustrate the area/depth trade off is that of pebbles on a beach. There are only a handful of very large (say > 50 cm diameter) pebbles on a beach, so you'd need to look at the whole beach to get a representative sample of them. However, since they're large, they're easy to spot and study from a distance. By contrast, there are hundreds of thousands of $< 1\ cm$ diameter pebbles, so you may only need to survey $1\ m^2$ of beach to get a representative sample, but you'd only be able to study them in detail from a few centimeters away.

Finally, if you are interested in studying the full range of galaxy environments and large-scale structure (i.e., clusters, voids, etc.), then you'd need a wide-area survey again. This is because the

Universe is only truly homogenous on large scales, so you need a wide-area survey to cover these scales and sample all types of environment.

In summary:

- **Large area, shallow surveys:** Are good for studying rare, luminous galaxies and sampling the full range of scale structure, but less good for studying more typical, fainter galaxies.

- **Small area, deep surveys:** Are good for studying fainter, more typical galaxies, but won't contain many luminous examples, nor span the full range of galaxy environments.

Between these extremes is a wide range of trade-offs between area and depth.

3.2 Where to survey

Since the Universe is homogeneous, we should, in principle, be able to choose any patch of sky to conduct a survey on. Unfortunately, it's not as straightforward as that, and a lot of thought goes into where on the sky extragalactic surveys are conducted.

The first and foremost consideration is that the aim of extragalactic surveys is to obtain samples of galaxies *external* to our own. As such, astronomers must choose regions of the sky that are not impeded by the Milky Way in the foreground. Otherwise, the light from external galaxies would be heavily obscured by the dust (and at X-ray wavelengths, gas) within the Milky Way.

A secondary consideration is observability by our telescopes. For example, some regions of the night sky near the celestial poles can be observed continuously by the Hubble Space Telescope as it orbits the Earth, whereas other regions are, at times, blocked by the Earth. Thus, if the design of your survey requires a lot of Hubble observations, then it may be more efficient to survey a patch within the continuously-observable regions. By contrast, regions of the sky near the celestial equator can be observed by ground-based telescopes in both the Northern and Southern hemispheres, so if the design of your survey requires a lot of ground-based observations, then it may be more efficient to survey a patch closer to the celestial equator.

Over the years, a number of well-observed patches of sky have been surveyed and now there are a selection of "well-surveyed" regions that get repeatedly observed by existing and new telescopes (e.g., COSMOS, Lockman Hole, GOODS-S, GOODS-N). Since these well-observed patches of sky now have a huge amount of multiwavelength data available for them, they are frequently chosen as the go-to patches to re-survey with new telescopes.

3.3 Multiwavelength surveys

A major benefit of having a set of well-surveyed patches of sky is that we can concentrate our attention on building-up data from across the full observable electromagnetic for these regions.

For example, some of the most well-surveyed patches of sky, such as the GOODS and COSMOS fields, have been observed extensively at radio, sub-mm, infrared, optical, ultra-violet and X-ray wavelengths. As such, we can combine all this data to obtain as complete a picture as possible for the galaxies in those fields. It would be far less effective if one patch of the sky had been observed in radio wavelengths, another in the optical, and another at X-rays, for example, since we wouldn't multiwavelength coverage of the same galaxies.

The reason this is important is that different parts of the electromagnetic spectrum provide information on different physical properties of the galaxies. For example:

- *Radio*: Star formation rate (SFR), Active Galactic Nuclei (AGN) power.
- *Sub-mm*: (Obscuration-independent) SFRs, gas content, dust content.
- *Far-infrared*: (Obscuration-independent) SFRs.
- *Mid-infrared*: (Obscuration-independent) SFRs and AGN powers.
- *Near-infrared*: Stellar masses, although these are significantly more precise if we have mass-to-light ratios from the...
- *Optical*: Mass to light ratios for stellar masses, star-forming histories (from spectral synthesis), AGN power (caveat dust obscuration).
- *Ultraviolet*: SFRs (caveat dust obscuration).
- *X-rays*: AGN power (and to a lesser extent SFRs).

Thus, by having full, multiwavelength data available for detected galaxies in a survey, we can investigate how different galaxy properties relate to one another. For example, how does SFR relate to stellar mass, or is there a connection between AGN power and SFR?

The other major benefit of surveying the same patches of sky in multiple wavelengths is to overcome problems associated with k-correction, which we first came across in Lecture 1. As we observe more and more distant galaxies, their spectra get shifted further and further redward. So, by observing two galaxies at different redshifts in the same band (i.e., wavelength), we actually sample different *rest-frame* wavelengths. For example, a v-band observation of a $z = 0$ galaxy samples its rest-frame 5400Å emission, whereas observing a $z = 2$ galaxy in the same band samples rest-frame 1800Å, which complicates comparison. However, if we also survey in the H-band near-infrared wavelengths (centred at 16300Å), then this will sample the $z = 2$ galaxy at $16300/(1+2) = 5433\mathring{A}$, which is *very* close to the rest-frame v-band. This therefore allows us to compare like-for-like rest-frame v-band luminosities. By surveying at *all* observable wavelengths, it maximises that likelihood that we will be able to compare galaxies at different redshifts at the same rest-frame wavelengths.

One major drawback of multiwavelength surveys, however, is that they have different spatial resolutions. For example, the Hubble Telescope's point spread function (PSF) at optical wavelengths is < 0.1 arcseconds. By contrast, Spitzer's PSF at infrared wavelengths was as high as tens of arcseconds, meaning potentially tens of sources detected with Hubble could lie within the Spitzer PSF. Trying to figure out which of those tens of Hubble sources corresponds to a single Spitzer source is a major challenge (indeed, there could be more than one Hubble source contributing to the total Spitzer flux).

4 Biases in surveys

As with most types of survey, the goal of a blank-field galaxy surveys is to provide an unbiased sample of galaxies. However, no survey is completely unbiased, and that holds true for galaxy surveys. The dominant bias present in blank-field surveys is the flux limit of the survey. Quite simply, blank field surveys will only identify galaxies that are brighter than (i.e., have fluxes higher than) the flux limit of the survey (which depends, among other things, on the telescope sensitivity and exposure time). As such, astronomers have to be extremely careful to take this "selection bias" into account when interpreting results from blank field surveys. Astronomers often have to ask themselves: “Is that correlation we see between two parameters real, or simpy due to the flux limit of the survey?”.

Such selection bias in galaxy surveys is similar to that in election polls. For example, it is well known that if pollsters rely solely on internet polls, then they will disproportionately under-represent older voters, since they tend to use the internet less than younger voters. Similarly, our blank-field surveys under-represent faint galaxies. Just as pollsters attempt to correct for biases to predict election outcomes, astronomers attempt to correct for biases in blank-field surveys.

Correcting for selection bias is particularly challenging when using data from surveys taken at multiple wavelengths. Unfortunately, reaching the same (relative) flux limits at all wavelengths is currently impossible due to cost and technological limitations. For example, today’s optical telescopes are extremely sensitive, meaning we have incredibly deep survey data at optical wavelengths. By contrast, our far-infrared surveys are much shallower, largely due to the limitations introduced by requiring space-borne telescopes to observe at these wavelengths. Since infrared emission is produced by star formation, this means we are biased toward detecting strongly star-forming galaxies at infrared wavelengths. If we did not correct for this bias, we might think that *all* galaxies were strongly star-forming (when, in fact, they aren’t).

5 Spectroscopic Surveys

To this point, we’ve only considered *photometric surveys*, i.e., taking an image of the sky in a certain waveband/frequency/photon energy. The majority of extragalactic surveys are, indeed, photometric due to the relative ease of conducting them (i.e., “point your telescope, take a picture, detect the sources”). Photometric surveys have, however, a number of significant drawbacks. In particular:

- **Photometric redshifts**: These are redshifts derived by shifting galaxy tempates (of the type used in spectral synthesis) to broad band photometric data. We’ll cover these in more detail in section 6; here, it’s suffice to say they’re a lot less precise than spectroscopic redshifts.
- **No kinematic information**: Unlike with spectroscopy, photometric data doesn’t provide any information on kinematics of galaxies (i.e., how they move, rotate, etc.)
- **Little or no information on gas physics**: Photometric data only really provides information on the combined stellar light of a galaxy (plus some information on the AGN, if a galaxy contains one). It contains very little information on the gas content of the galaxies, which is a crucial ingredient for star-formation (and thus galaxy evolution).

All of these problems can be resolved (to a greater or lesser degree) by taking the spectra of galaxies. However, until about 20 years ago, it was only really possible to take the spectrum of one galaxy at a time, which made obtain the spectra of a large sample of galaxies *very* time consuming and, thus, expensive. Recently, however, we have witnessed the rise of *spectroscopic surveys*. This is where multiple spectra of multiple galaxies (in the same field-of-view) can be obtained simultaneously, dramatically reducing the amount of time needed to obtain spectra for large samples of galaxies.

The most famous spectroscoic survey is the *Sloan Digital Sky Survey*, which has taken the spectra of over a *million* galaxies. It does this by placing fibre-optics at the positions of stars and galaxies in its field-of-fiew which carries the light down to a grating, which then disperses it onto the detector. Today, most large telescopes have some kind of such "multiplexing" capabilities (although not all are fibre-fed), meaning that many of the most well-studied "blank-field" surveys have excellent spectroscopic coverage. The main drawback of spectroscopic surveys, however, is that the targets for spectroscopic follow-up (usually) have to be pre-selected from photometric data. This means it is typically the brightest, or most "interesting", galaxies that are chosen for spectroscopic follow-up. Obviously, this introduces a bias which can be difficult to account-for.

Recently, pre-selection bias has been mitigated (to some degree) by the introduction of large-area (1-arcmin^2) Integral Field Units (IFUs), such as MUSE on the Very Large Telescope. These IFUs take the spectrum at every single point within the field-of-view, irrespective of whether there is a star or galaxy there, or whether it's just blank sky. While the fields-of-view are still quite small, it does offer the prospect of selection-free spectroscopic surveys in small patches of the sky.

6 Photometric redshifts

As we saw in the previous section, one of the key benefits of multiplexed spectroscopic surveys is that they provide highly accurate spectroscopic redshifts for large numbers of galaxies. However, even the most capable multiplexing systems cannot target all galaxies spectroscopically. Why is this the case? Well, firstly, there is the problem of the limited availability of fibers or slits: there are simply too many galaxies to target each one. Secondly, there is the problem of source brightness: you can usually only get the spectra of the brighter galaxies in a survey. This is because when a spectra is taken, the light from the galaxy is *spread out*, and the more it is spread out, the less light there is per pixel on your detector, meaning a lower signal-to-noise per pixel. It's impossible to take meaningful spectra of the faintest galaxies in your survey.

To overcome these problems, astronomers have developed a technique to derive redshifts from photometric data, rather than spectroscopic data. To describe this technique, let's first consider how we obtain a normal spectroscopic redshift. Here, we spread the light from a galaxy over lots of individual wavelength bins, each with a width of, say, $\Delta\lambda = 1$Å. Since $\Delta\lambda$ is small, we're be able to *resolve* individual spectral lines, meaning each line will be covered by lots of individual wavelength bins. Next, by comparing the observed wavelengths of lines to the rest wavelengths of the lines, we can calculate the redshift (i.e., $z = \lambda^{\rm obs}_{\rm Line}/\lambda^{\rm rest}_{\rm Line} - 1$). So far, so familiar. Now, consider we increase $\Delta\lambda$ to 10Å. Each wavelength bin is now 10 times wider, so we'll lose some *spectral resolution*. Now we may not be able to fully resolve individual spectral lines: each spectral line might only be covered by one wavelength bin. We'd still be able to identify the lines, but because they're not well resolved, there will be a larger uncertainty on $\lambda^{\rm obs}_{\rm Line}$, meaning a larger uncertainty on the redshift. If we now increase $\Delta\lambda$ to 100Å, we probably won't resolve *any* individual emission lines. However, we'll still be able to see the overall *shape* of the continuum. In particular, we may even be able to see *breaks* in the continuum caused by the absorption of photons by neutral hydrogen. If we can see the overall shape, then we can try to fit it with a galaxy template, shifting the template to and fro in wavelength until we get a good fit. Once we find a good fit, we the amount we've had to shift the template by gives us the redshift of the galaxy.

If instead of a spectrum with bins of $\Delta\lambda$=100Å we have lots of different *photometric filters* of width $\Delta\lambda = 100$Å, we have exactly the same situation: lots of very wide bins of flux. By fitting the overall *shape* of the spectrum or SED that these photometric points trace out with galaxy templates – shifting the template in wavelength until we get a good fit – we can derive the redshift of the galaxy. Of course, since the wavelength bins are wider, the uncertainty in the shift is larger than if we can resolve individual lines. Also, the precision decreases dramatically with fewer and fewer filters. This explains why photometric redshifts are less precise than spectroscopic redshifts. Having said that, provided we have photometric data from lots of filters (> 10 filters is not uncommon in blank field surveys) today's photometric fitting codes can typically measure photometric redshifts to within an accuracy of $\Delta z/z \sim 0.1$ or better.

The large numbers of accurate photometric redshifts available for galaxies in blank field surveys has had a dramatic effect on statistical studies of galaxies. While an individual photometric may not be particularly accurate, if we have lots and lots of them (and we do), then we can reliably use them to derive statistical properties of galaxies, particularly distributions such as mass and luminosity functions.

7 The future of extragalactic surveys

Because of the huge impact galaxy surveys have had on our understanding of the Universe, they are factored-in as a major component of all new observing facilities. Indeed, some telescopes are built specifically to conduct surveys. As such, they will continue to play an increasingly important role in revealing how galaxies have formed and evolved.

8 The Pros and Cons of galaxy surveys

Finally, I though I'd wrap up with a summary of the pros and cons of galaxy surveys:

Pros:

- Samples of galaxies unbiased by pre-selection (but see "con" about flux bias).
- Because the "blank fields" have been surveyed by lots of different observing facilities, we have a lot of multiwavelength data for the galaxies within those surveys. This helps considerably when determing galaxy properties and overcoming k-corrections.
- Surveys are an excellent sources of targets for more detailed, follow-up studies.
- The deepest surveys provide the most sensitive view of the Universe to-date.

Cons:

- Although less biased than studies of pre-selected galaxies, "blank-field" surveys are biased toward brighter galaxies because of the flux limit of the survey.
- While multiwavelength data often exists, it can be a challenge to match between different wavelengths.
- A major drawback of surveys is that they often lack the "detail" of more targetted observations.
- The smallest-area surveys can be particularly badly affacted by cosmic-variance (which was described in Lecture 2).

9 Key learning objectives for L6

By the end of this lecture you should have an understanding of:

- why we conduct extragalactic surveys;
- , the different survey strategies, and why we use them;
- how survey fields are selected for observations;
- the multi-wavelength aspect of surveys, and what physical propertied are measured by different wavelengths;
- the pros and cons of survey science;
- photometric redshifts.

Lecture 7:
The star-forming history of the Universe & Lyman Break Galaxies

Dr. James Mullaney

February 27, 2017

1 Introduction

In the previous lecture, we covered the general philosophy and techniques of survey astronomy. In this lecture, we will cover two important results that are rooted in data from extragalactic surveys: the cosmic history of star formation, and Lyman Break Galaxies.

2 The cosmic history of star formation

Galaxies are made of stars, so asking "when did today's galaxies form?" is akin to asking "when did today's galaxies form their stars?".[1] In lecture 5, we looked at how we could use the fossil record to determine this: effectively combining the star forming histories of all local galaxies to determine when their stars were put in place. We can also, however, try to answer this question using the redshift approach. We can study galaxies at different redshifts (i.e., at different *lookback times*) and measure how much star formation was taking place at different times throughout the history of the Universe. We do this by combining the methodologies outlines in lectures 4 and 6: using UV or infrared surveys to survey the star formation rates (SFR) of galaxies at different times during the history of the Universe. From this, we can determine – on average – when galaxies formed their stars.

2.1 The Madau diagram

One of the first attempts to systematically measure the star-forming history of the Universe using blank field surveys is described in Madau et al. (1996). As such, the resulting plot of SFR vs. redshift is known as the *Madau diagram* or *Madau plot*. Madau used rest-frame UV emission (from the Hubble space telescope) to measure the average SFRs of galaxies out to a redshift of $z \sim 5-6$ (i.e., over the past $\sim$ 12 billion years (or $\sim$ 90% of the age of the Universe). What Madau et al. found was that the *SFR density* – that is, the average SFR per unit volume – appeared to peak at around $z \sim 1-2$, corresponding to about 6 to 8 billion years ago. Since that time, the SFR density of the Universe has slowly declined until today when it is about one tenth of what it was at its peak. It's like galaxies in the Universe had a growth spurt around 6-8 billion years ago, and has now settled into middle age. Further, since heavier elements (i.e., remember, in astronomy anything heavier than He is a *metal*) are formed in the cores of stars and when stars go supernova,

[1] We'll leave *how* stars are formed to Simon's Star Formation course.

the peak of SFR at $z = 1 - 2$ also means that the rate of metal production also peaked at these times.

It's interesting to consider just how rapidly galaxies were forming their stars when they were at their peak SFRs. From the Madau plot, we see that the SFR density of the Universe peaked at around 0.1 $M_{\odot}$ per year per Mpc^3. That may not sound like much, but bear in mind that most of the Universe is empty. As such, some galaxies were producing well over a hundred stars *per cubic kpc* per year.

Because of the impact that Madau et al. (1996) had on the field of SFR history, the terms "Madau diagram" or "Madau plot" have now become synonymous with all subsequent plots of SFR density vs. redshift (no matter who publishes them).

2.2 Uncertainties in the Madau diagram

Of course, as with any measurement – especially first attempts – there were considerable uncertainties associated with the earliest attempts to measure the SFR density of the Universe. In particular, by using rest-frame UV light to measure SFRs, the earliest Madau diagrams suffered significantly from uncertainties due to dust obscuration (recall, the UV is *strongly* attenuated by dust). Furthermore, as we saw in Lecture 2, there are significant difficulties associated with measuring light from high redshift galaxies. In particular, Hubble images are strongly biased toward detecting high surface brightness galaxies (remember the $1/(1 + z)^4$ law for surface brightness from Lecture 2), meaning they could easily miss any diffuse regions of star formation and thus underestimate the SFR density. Finally, we also have the usual problems associated with converting UV fluxes to SFRs: how to convert UV luminosity to numbers of high mass stars (which depends on uncertain models of stellar evolution) and from there to total numbers of stars (which depends on uncertain IMFs, especially in the high redshift Universe).

As our telescopes have become more sensitive at more wavelengths (especially infrared wavelengths), our ability to mitigate these uncertainties has improved (consider material in L2). As such, since its first depiction in 1996, the Madau diagram has evolved somewhat, especially at the high redshift end. While it remains the case that the SFR density of the Universe peaked at around $z = 1 - 2$, the most recent Madau plots show less of a steep decline *at higher redshift* than first thought. In other words, the SFR density of the Universe was fairly constant (if slightly increasing) from between $z = 6$ to about $z = 1 - 2$, after which it has dropped significantly. As we shall see later in the course, the reason for this drop-off in SFR density over the past 6-8 billion years is due to the declining availability of gas – the raw material of star formation – in the Universe as it gets used up to form stars.

3 Lyman Break Galaxies

Identifying galaxies at high redshift – as required if we want to use this approach to study galaxy evolution – can be extremely challenging. Not only do we have to contend with the faintness of these galaxies due to their distance from us (as outlined in Lecture 2), but once you have a sample of galaxies, how do you reliably identify the few that are at the highest redshifts out of the millions of more local galaxies? Even with spectroscopic surveys, it would be highly inefficient to target all galaxies with the hope of finding the small handful that were at the highest redshifts.

One approach we saw in the last lecture was to use photometric redshifts. However, that approach requires observations in *many* different bands and so can be costly (in terms of telescope time). In this section, we'll look at a way we can identify high redshift galaxies using as little as two wavebands. This technique relies on exploiting a "break" in a galaxy's spectrum – typically where the flux lonward of the break is higher than shortward of the break. One of the strongest breaks in the spectra of galaxies is the *Lyman* break, so galaxies that are identified by exploiting this break are known as *Lyman Break Galaxies* (we like to be original in astronomy!).

3.1 Identifying Lyman break galaxies

Before we consider how we actually identify Lyman break galaxies, we'll first look into what the "Lyman break" actually is. Consider a galaxy containing a population of massive, hot stars. Because they are hot, they release a strong continuum of UV light. Usually, these stars will be sitting in a pool of surrounding gas (containing mainly Hydrogen) from which they have been born. This means that, almost immediately after the UV photons leave the surface of the stars, those with wavelengths shorter than 912Å (and thus capable of ionising H) hit a neutral H atom and is *absorbed* by it, producing a H ion. The H ion will (eventually) recombine with an electron but – importantly – it is unlikely to recombine directly into the lowest energy level. It is far more likely that it will recombine to a higher energy level, emitting a lower energy photon than the original ionising photon followed by a series of other transitions. As such, it is *highly unlikely* that the original < 912Å will be re-emitted. What all this means is that, because of the ready absorption of < 912Å photons by Hydrogen, there is a sudden drop in flux shortward of this wavelength – known as a "break". And because it is caused by absorption by the first, or *Lyman*, level of Hydrogen, it is known at the *Lyman Break*.

As mentioned already, astronomers can use this break to identify high redshift galaxies. To explain how this works, however, we'll first consider a galaxy at $z = 0$. In this low redshift example, we'd detect the galaxy at wavelengths *longward* of 912Å, but not *shortward* of 912Å, since these short wavelengths photons are absorbed by the hydrogen gas in the galaxy. For galaxies at higher redshifts, however, the break *shifts* to *longer* wavelengths, from the UV to optical wavelengths. For example, for a galaxy at $z = 3.2$, the break will be at $912 \times (1 + 3) = 3830$ Å. This *redshifted* break at 3830Å now lies between the U and B wavebands (centred at 3650Å and 4450Å), but the same principle still applies. A bright, star-forming galaxy at $z = 3.2$ would be detected in the B band, which samples *longward* of the break, but it would not be detected in the U-band, which samples *shortward* of the break. Such galaxies are referred to as "U-band *dropouts*" because they are said to have *dropped-out* of the shorter-wavelength U-band. Because of the wavelength separation between the U and B filters, these bands are sensitive to dropouts between $z \approx 3650/912 - 1 = 3$ and $z \approx 4450/912 - 1 = 3.9$.

By using longer wavelength bands, astronomers can identify Lyman-break galaxies at higher and higher redshifts. For example, using the B and V (5500Å) bands would identify B-band dropouts between $z \approx 4$ and $z \approx 4.9$, while the V and R (6580Å) bands would identify V-band dropouts between $z \approx 4.9$ and $z \approx 6.1$. As a test, what redshift could an I-band dropout have (centred at 8060Å; the next longest wavelength is the z-band at 9000Å)? What type of dropout would the highest-redshift galaxy, at $z = 11.09$ be? (You'll need to look up the central wavelengths of common telescope filters. *Clue*: Check out the near-infrared)

Since their discovery, a large number of Lyman break galaxies (hereafter, LBG) have had their redshifts confirmed spectroscopically using large (8-10 m class) telescopes. The dropout technique makes this feasible – out of the thousands or even millions of galaxies in the deep field surveys, only a few thousand might be LBGs, which can (relatively) easily be followed-up with multi-fiber or multi-slit spectrographs. However, only the brightest ($m_R < 25.5$ LBGs can have their redshifts confirmed spectroscopically. Most fainter ones will remain unconfirmed until larger, more sensitive telescopes are commissioned.

3.2 The spectral characteristics of Lyman break galaxies

For the LBGs that astronomers *have* been able to obtain spectra for, not only do we know their (confirmed) redshifts, we can also measure some of their physical properties. Typically, the spectra of LBGs show a combination of both nebular emission and absorption lines, as well as (weak) absorption lines associated with stellar photospheres. Some also show strong Lyα (at rest-frame 1216Å) and CIV emission from ionised gas surrounding the hot, young stars in these galaxies. However, these lines often appear asymmetric as the photons in the blue wings of the lines are absorbed by Hydrogen gas in the galaxy (which becomes excited to the $n = 2$ level).

The spectra of LBGs appear similar to galaxies in the local Universe that are undergoing episodes of rapid star formation – known as *starburst* galaxies. However, LBGs tend to have spectra that are metal-poor compared to galaxies in the local Universe, suggesting large amounts of *pristine* gas (i.e., unprocessed by star-formation) in LBGs. This makes sense when considering that we're observing LBGs as they were when the Universe was young before lots of reprocessing in stars had "contaminated" gas with metals.

3.3 The morphological characteristics of Lyman break galaxies

Once astronomers have identified LBGs using the dropout technique, they can also study their morphologies (i.e., shapes) in the bands *longward* of the break. Because of their high redshifts, this typically requires the high spatial resolutions only afforded by the Hubble Space Telescope (or adaptive optics, in which the distorting effects of the atmosphere are counteracted using deformable mirrors placed in your telescope optics.)

Detailed morphological studies of LBGs have reveals that the tend to be physically smaller than local galaxies of the same luminosity, with half-light radii similar to the bulges of local spiral galaxies or small ellipticals (although, as ever, we need to be wary of missing low surface brightness features). While some LBGs appear relatively smooth in structure (like today's disk galaxies), many show signs of being highly clumpy and irregular in nature, which is suggestive of mergers and/or interactions in the high redshift Universe. You will recall from Lecture 2 that we need to be mindful of morphological K-corrections if we see such "clumpy" structures in high redshift galaxies. However, LBGs show little evidence of suffering from this problem, with both the rest frame optical (probed by the near-infrared bands) and rest-frame UV (probed by the optical bands) showing similar morphologies.

3.4 The star-forming properties of Lyman break galaxies

By the very nature of how they are identified, LBGs must have a strong UV continuum. There must be a lot of UV photons just longward 912Å (which is still the UV) in order for a LBG to be seen to drop out of the shorter waveband. Since only massive, hot, *young* stars produce a strong UV continuum, and since these stars only live for a short time, then LBGs must have recently undergone a recent bout of star formation.

While LBGs must be star-forming, actually measuring their star formation rates (SFRs) can be problematic. This is because their SFRs are usually measured via their rest-frame UV flux, or sometimes via their nebular emission (if a spectrum is available), both of which can be strongly affected by dust obscuration. Prior to correction for dust, LBGs are measured to have SFRs of a few tens of solar masses per year. After correcting for the effects of dust, however, we find they have SFRs of around 100 $M_{\odot}yr^{-1}$. This is a very high SFR by today's standards - the SFR of Milky Way is about 1 $M_{\odot}yr^{-1}$; only a handful of galaxies undergoing intense starbursts due to major gas-rich mergers in the local Universe have SFRs even approaching 100 $M_{\odot}yr^{-1}$.

Since LBGs are strongly star-forming, we can include them in our measurements of the star-forming history of the Universe and plot them on the *Madau plot*. Since we have identified lots of LBGs, we have good statistics for them, meaning the average SFR density that we derive for them is well-defined and has comparatively small error bars. When we include (dust corrected) LBG SFRs to the Madau plot, we find that they agree well with SFR densities derived using other techniques: the SFR density of the Universe peaked at around $z = 1-2$, and was relatively constant before that time.

3.5 The masses of Lyman break galaxies

As well as measuring their SFRs from UV observations, we can also measure the stellar masses of LBGs using spectral synthesis (recall spectral synthesis fitting can be performed on photometric data, it doesn't require spectroscopic observations). When we do this, we find that the stellar masses of LBGs are typically between $10^9 - 10^{11}$ $M_{\odot}$ (some of the spread is introduced by the uncertainty in the model fits). Interestingly, however, when we measure the *dynamical* mass of LBGs via the (assumed) gravitational motions of their stars, we find that they are about a factor of ten *lower* than the total stellar mass measured from spectral synthesis. Clearly, the gravitational mass can't be lower than the mass of all the stars (if it were *higher* it could simply be explained in terms of dark matter). This suggests that the dynamical masses of LBGs are an underestimate, possibly introduced by their motions having a non-gravitational component (e.g., mergers).

3.6 Evidence of outflows in LBGs

When the spectra of LBGs are obtained, it is often observed that their emission and absorption lines are broadened and/or shifted relative to their rest-frame wavelengths. Since shifts in emission or absorption lines are produced by gas moving at high velocities relative to the galaxy, this is seen as strong evidence of inflowing and outflowing gas in LBGs. While the inflowing gas is likely feeding the star-formation, the outflowing gas is thought to be driven by strong winds, powered either by supernovae or stellar mass loss.

Such winds are thought to be extremely important in shaping galaxies, not least in the early Universe. By expelling gas from the galaxy, such outflowing winds are one of the key feedback mechanisms employed in hydrodynamic and semi-analytic models of galaxy evolution. Indeed, winds driven by stellar mass loss and/or supernovae are widely thought to have suppressed the formation of smaller galaxies. This is why the *galaxy* mass function is less steep than the dark matter halo mass function at low masses (see Lecture 3).

3.7 The volume density and clustering of LBGs

Finally, we will consider the distribution of LBGs in space. Since LBGs are detected in blank field surveys, we can calculate how many there are per unit volume of the Universe. To do this, we measure the *luminosity function* of LBGs which, if you recall, provides the number of galaxies (in this case LBGs) per unit volume in a given luminosity bin.

As with most luminosity functions in astronomy, the LBG luminosity function is well fit by a Schechter function: roughly a broken power law with a break at a given luminosity, L^*. When we measure the luminosity function of LBGs, we find that the volume density of LBGs with luminosities close to L^* is roughly the same as the volume density of L^* galaxies in the local Universe. This similar volume density suggests that LBGs may well have evolved to form many of today's massive galaxies. This is further backed up by observations that LBGs cluster together in a similar way to today's galaxies.

4 Where are LBGs today?

Based on their masses, volume density and clustering properties, there is reasonable evidence that LBGs have evolved to form many of the intermediate to high mass galaxies in the local Universe. However, by now, the stars we see forming in LBGs at $z > 3$ will be over 10 billion years old, so will represent some of the oldest stars in today's galaxies. Further, the characteristic masses and sizes of LBGs are similar to those of the cores and central bulges of local elliptical and spiral galaxies, respectively. However, their morphologies are far more disturbed by mergers than nearby bulges or ellipticals. As such, it is thought that what we are seeing when we observe LBGs are the central bulges of spiral and elliptical galaxies in the process of being formed.

5 Learning objectives for Lecture 7

We've covered a couple of important topics in this lecture. Here are the learning objectives:

- Know what is mean by the star-forming history of the Universe.
- Know what the Madau diagram is, including its shape, how it is affected by extinction due to dust, and other challenges associated with defining it.
- Know what a Lyman Break Galaxy (LBG) is.
- Know how LBGs are identified (i.e., be able to decribe the dropout technique).
- Understand the spectroscopic, morphological, star-forming, mass, and clustering properties of LBGs.
- Have an idea of the importance of LBGs in the build-up of today's galaxies.

Lecture 8:
Studying galaxy evolution in the IR/sub-mm

Dr. James Mullaney

February 27, 2017

1 Introduction

In the previous lecture, we covered the use UV wavelengths to measure the star-forming properties of high redshift galaxies. We also looked at Lyman Break Galaxies, which are those galaxies that are *identified* exclusively via a *break* in their rest-frame continuum.[1]. However, as we've seen extensively already, the UV is heavily affected by absorption by interstellar dust, which needs to be accounted-for using uncertain correction factors. In this lecture, we'll consider the use of the longer infrared (1 μm – 500 μm) and sub-mm (500 μm – 1 mm) wavelengths to measure the star-forming properties of galaxies. These long wavelengths, especially $\gtrsim 30\ \mu m$, are almost impervious to absorption by dust.

2 What produces infrared emission in galaxies?

Infrared emission is produced by warm stuff. That's the case whether it's a (living) human body, coals on a fire, or a galaxy. In the latter case, the infrared is largely emitted by dust in the galaxy that is being heated by shorter wavelength, UV photons. These UV photons hit the dust particles, their energy goes into heating the dust particle, which then re-radiated the energy at infrared wavelengths. Eventually, this absorption of UV photons, heating and re-emission at infrared wavelengths reaches an equilibrium and the dust becomes a black body (i.e., with a continuum spectrum described by the Planck equation).

So, to emit in the infrared, a galaxy needs two things: dust, and a source of UV photons to heat the dust. It's still not entirely clear where dust comes from, but it's thought that supernovae (and thus star-formation) play a key role in dust production. There are two main sources of UV photons in a galaxy: the first and most common are hot, young stars, the second is emission from the accretion disk of an AGN. Typically, UV photons from populations of young stars will "heat" the dust to about 20-100 K (yes, it's still pretty cold, but much warmer than the $\sim$ 3K Cosmic Microwave Background). By contrast, AGN tend to produce more high energy UV photons, so are thought to heat the dust to warmer temperatures (although, as Clive will tell you, there is some contention on this issue).

[1]**Important note**: The presence of a breaks in galaxy continua is relied-upon heavily when measuring photometric redshifts. A break is a dramatic feature in the *shape* of an SED, so it helps a lot when "shifting" the synthetic spectrum in wavelength to find a photometric redshift

3 Early observations of the Universe at infrared wavelengths

Since most of the infrared part of the spectrum ($5 - 500$ μm) is absorbed by the Earth's atmosphere (largely by water molecules; the $1 - 5$ μm) wavelengths less-so), observations of the Universe at these wavelengths normally require space-based observatories. The first astronomically useful space-borne infrared telescope as **IRAS**, which was launched in 1983. This conducted the first infrared survey of the whole sky (at 12, 25, 60 and 100 μm). Remarkably, it remains the *only* all-sky survey at 60 and 100 μm).

During its lifetime, **IRAS** identified around infrared 350,000 sources. Many of these remain unclassified, but it is known that around 75,000 are extragalactic sources. Due to infrared astronomy being in its infancy when *IRAS* was launched, it suffered from poor sensitivity by today's standards. As such, it only detected the brightest infrared galaxies; all but the nearest ones are there extremely luminous at infrared wavelengths. Such galaxies are known as either *luminous infrared galaxies*, or LIRGs (with infrared luminosities $L_{\rm IR} = 10^{11} - 10^{12}$ ergs s^{-1}), and *ultra luminous infrared galaxies*, or ULIRGS $L_{\rm IR} = 10^{12} - 10^{13}$ ergs s^{-1}.[2].

On discovery of these bright infrared galaxies by IRAS, astronomers observed them in follow-up observations at optical wavelengths. What they found was that many of these infrared bright galaxies were in the process of undergoing *major mergers* of gas-rich galaxies.[3] In such cases, intense episodes of star-formation is triggered by the merger process compressing the gas within the merging galaxies, causing it to collapse to form stars at a very high rate. A LIRG is typically forming stars at a rate of 10-100 $M_\odot$ per year, whereas a ULIRG forms stars at a rate of 100-1000 $M_\odot$ per year. Because of their very high star formation rates, such galaxies are known as *starburst* galaxies.

Due to its low sensitivity, however, IRAS was not very good for exploring the high redshift Universe, so provided little insight into the evolution of galaxies at infrared wavelengths. Instead, this had to wait until the development of sub-mm astronomy in the mid-90's.

4 Sub-mm astronomy

As the name suggests, sub-mm astronomy exploits the wavelengths of light shortward of 1 mm, just before you enter the infrared regime at around 500 μm. It is the part of the spectrum between the radio and infrared. One of the greatest benefits of sub-mm astronomy over infrared astronomy is that the Earth's atmosphere is relatively transparent at these wavelengths, especially at very high altitudes at very dry sites.[4] As such, they are not as limited in terms of aperture size as infrared telescopes; indeed, the James Clarke Maxwell (JCMT) sub-mm telescope on Mauna Kea is 12 m in diameter. Despite its size, however, it still suffers from poor spatial resolution due to the long wavelengths it detects (remember, angular resolution$= 1.22\lambda/D$, where λ is wavelength and D is aperture size). For example, the angular resolution of JCMT is roughly 15 arcsec compared to $\sim$1 arcsec of ground-based optical telescopes.

[2]There are also HyperLIRGS, with $L_{\rm IR} > 10^{13}$ ergs s^{-1}, but these are extremely rare

[3]A major merger is usually defined as a galaxy merging with another galaxy with at least one third of the mass of the first. By contrast, a *minor merger* is usually defined as when a galaxy merges with another galaxy with a mass *less* than a third of the first.

[4]There are only a few places around the world that are high and dry enough to conduct sub-mm astronomy. Two of particular note are the summit of Mauna Kea on Hawai'i, and the Atacama Desert in Chile.

4.1 The positive K-correction of sub-mm observations

Another great benefit of sub-mm astronomy derives not from practical considerations, but the very shape of the infrared/sub-mm SED of galaxies. Recall from Lecture 2 the K-correction that needs to be applied to observations of high redshift galaxies due to the shifting of shorter wavelengths into the observed band. At optical wavelengths, this K-correction is said to be negative because it works against us: galaxies tend to be *less* luminous at shorter optical wavelengths, so when these fainter, shorter wavelengths are redshifted into our observing bands they are more difficult to detect. By contrast, at infrared and sub-mm wavelengths (longward of about 100 μm), galaxies tend to be *more* luminous at shorter wavelengths. As such, when these *brighter* shorter infrared wavelengths are shifted into our sub-mm observing bands, they are *more easily* detected. This *positive* K-correction almost exactly counteracts the $1/r^2$ cosmological dimming, meaning that a $z \sim 1$ galaxy is as easy to detect at 850 μm as a $z \sim 3, 4, 5$ galaxy! Not only that, but it can actually be *easier* to detect distant galaxies in the sub-mm than ones in the local Universe.

4.2 High redshift astronomy at sub-mm wavelengths

Because of the positive K-correction, sub-mm wavelengths are ideally suited to studying high redshift – and therefore distant, early – galaxies. However, early (1990s to the mid-2000s) sub-mm studies of high redshift galaxies were severely hampered by the poor resolution of the telescopes. When the sub-mm point spread function (PSF) is around 15 arcsec in diameter, it is extremely difficult to pinpoint which galaxy seen at optical wavelengths it corresponds to. For example, in the deep fields – where the first sub-mm surveys were conducted – Hubble could detect ten or more galaxies within a circle of diameter 15 arcsec. Which (if any) of those Hubble galaxies actually correspond to you sub-mm detection? Indeed, there could be more than one contributing to the total sub-mm flux of your sub-mm source. This causes significant problems for measuring the redshifts of your sub-mm galaxies. If you don't know which galaxy it is, you can't even use photometric redshifts, let alone target them for spectroscopic follow-up. It is possible, however, use the overall *shape* of the sub-mm SED to get a very crude redshift (often by just using a flux ratio to define the shape), but this only gives a redshift that's accurate to within $\Delta z \sim 1$, corresponding to roughly half the age of the Universe!

4.3 Using sub-mm wavelengths to measure SFRs

Under normal circumstances, the large uncertainties in the redshifts of sub-mm galaxies would scupper our chances of measuring the (rest-frame infrared) luminosities (see Lecture 4 for how we use infrared wavelengths to measure SFRs). And if we don't know their luminosities, we can't convert these into meaningful physical parameters such as star formation rates. Again, however, the positive K-correction saves us. Because the positive K-correction counteracts the $1/r^2$ geometric dimming, it means that a galaxy of a given luminosity *will have roughly the same sub-mm flux*, irrespective of its redshift (to within reasonable limits, and certainly to within $\Delta z \sim 1$). As such, as long as we have *some idea* of the redshift (from sub-mm flux ratios, for example) and a measured flux, we can convert it into a fairly confident infrared luminosity and, therefore, SFR (yep, it feels like cheating, but it actually works!).

With SFRs in hand, astronomers could work out the SFR density (i.e., the SFR per unit volume in the Universe) arising from sub-mm galaxies (SMGs) and plot this on the Madau plot (which we saw in L7). Although only based on a small number of sub-mm detected galaxies detected in the deep fields (remember, this was during the 1990's when sub-mm astronomy was still in its infancy) – and thus subjected to significant uncertainties – it appeared that SMGs were responsible for roughly the same contribution to the SFR density optical/UV-detected galaxies. Thus, sub-mm observations roughly doubled our estimates of the amount of star formation (per unit volume) taking place in the high redshift Universe.

Most importantly, sub-mm observations showed us that a significant proportion of star-formation activity in the Universe had been *completely missed* by optical/UV observations due to absorption by to dust. That's the equivalent of measuring global birth rates, then discovering India and China!

4.4 Pin-pointing the positions and redshifts of SMGs

As we have seen, during the mid-90's it became increasingly clear that SMGs were an important component of the star-forming history of the Universe. As a consequence, it became increasingly important that astronomers were able to pinpoint their positions more accurately, not least to obtain accurate redshifts.

Since all sub-mm galaxies are strongly star-forming, it was realised that observations at other wavelengths that are also sensitive to star-formation could also help to pinpoint them, provided they were higher resolution. However, as we've seen, we can't use the optical/UV, as it is blocked by dust. The (partial) solution came in the form of radio observations. This is because when stars undergo supernovae, they release a lot of radio emission due to synchrotron emission in the supernova ejecta. If a galaxy is strongly star-forming, it have a high rate of supernovae, and thus a strong radio luminosity. The benefit of radio wavelengths over UV/optical is that they are *completely impervious* to dust (if a radio wave can get through the walls of your house, it can easily penetrate interstellar dust). The other great benefit of radio wavelengths is that we can reach very high angular resolutions via radio interferometers (in fact, very long baseline radio interferometers provide the highest angular resolutions of *any* telescope).

By targeting sub-mm galaxies with radio interferometers, astronomers were able to pinpoint the positions of some of them to within about 1 arcsecond. This is easily accurate enough for follow-up optical spectroscopy, from which spectroscopic redshifts could be obtained. Unfortunately, not all sub-mm galaxies are detectable at radio wavelengths due to the relative sensitivities of our telescopes at different wavelengths (recall the biases introduced by multiwavelength surveys discussed in Lecture 6). Similarly, even if we have a radio position, that is no guarantee that our optical telescopes will be able to obtain a redshift since, by their nature, sub-mm galaxies are heavily obscured at optical wavelengths. As such, by 2005, only around 50% of the 150 brightest sub-mm galaxies had secure spectroscopic redshifts.

4.5 SMGs in context

Acknowledging the above problems, with the redshifts that we *did* have, it was clear that sub-mm galaxies were, indeed, more common at high redshifts. Their numbers peak at around redshift $z \sim 2.4$. With precise redshifts, we could finally obtain precise infrared luminosities and, thus SFRs, which revealed that sub-mm galaxies typically had SFRs over 100 $M_\odot$ per year. It was also confirmed that the SFRs of sub-mm galaxies are at least 100 times *higher* than what would be estimated via the rest-frame UV (when they are even detected at these short wavelengths and after trying to correct for dust obscuration), meaning that most of the star formation is hidden by vast columns of dust.

Sub-mm galaxies therefore have SFRs that are roughly similar to ULIRGs today. However, ULIRGs in the local Universe are extremely rare - roughly one per ten million Mpc^3. By contrast, sub-mm galaxies are roughly a thousand times more common at $z \sim 2.5$ than ULIRGs are today. This correxponds to a much higher SFR density, as seen when the contribution from sub-mm galaxies are included on the Madau plot.

It is interesting to compare SMGs against the other main type of star-forming galaxies at high redshift: Lyman break galaxies (LBGs). After correcting for dust extinction, LBGs and sub-mm galaxies contribute roughly equal amounts to the total SFR density of the Universe at $z \sim 2 - 2.5$. However, unlike LBGs, whose contribution continues to remain significant to higher redshifts ($z \sim 5$), the contribution from sub-mm galaxies drops away at higher redshifts. As such, the sub-mm "phase" of the Universe is, by comparison, fairly short lived, lasting from $z \sim 3$ to $z \sim 1$, or roughly 4 billion years.

Finally, since the density of SMGs peak at $z \sim 2.4$, and with today's most massive galaxies already in place by $z \sim 3$, it is unlikely that we are witnessing in SMGs the build-up of today's most massive galaxies. That role seems to be being played-out by LBGs at higher redshifts. Indeed, it's still not entirely clear what SMGs evolved to become in the current Universe.

5 Spitzer's contribution to infrared astronomy

As we have seen, for many years after IRAS completed its all-sky survey, infrared astronomy was dominated by observations at sub-mm wavelengths (and that's not a contradiction: sub-mm observations of high redshift galaxies samples the rest-frame infrared). This changed in 2003 with the launch of *Spitzer*, which was significantly more sensitive (by roughly a factor of 1000) at $3.6 - 160$ μm than any telescope that had gone before. Unlike IRAS, however, it did not perform an all-sky survey. However, it was the first infrared telescope to conduct blank-field surveys comparable in depth to those provided by Hubble in previous years.

Due to its excellent sensitivity at infrared wavelengths, Spitzer was capable of detecting tens of thousands of *high redshift* galaxies in the deep field, *despite* the negative K-correction.[5] Most of the sources that Spitzer detected in the deep fields were either LIRGs (at redshifts between 0.8 and 1.5) or ULIRGs (at $z > 1.5$).

The greatest benefit that Spitzer has over previous infrared/sub-mm observatories was, however, its angular resolution, allowing its tens of thousands of infrared-detected galaxies to be pinpointed to with a few arcseconds. This allowed the Spitzer-detected galaxies to be matched to galaxies detected at other wavelengths in the deep fields (recall, deep survey fields tend to be observed in many different wavelengths) which, correspondingly, enabled accurate photometric redshifts to be determined. For the first time, therefore, astronomers were able to determine accurate infrared luminosities – and thus SFRs – for *tens of thousands* of galaxies across large redshift ranges.

5.1 The evolving infrared luminosity function

With the availability of accurate luminosities and redshifts for many, many galaxies, astronomers used Spitzer to determine – for the first time – how the infrared luminosity function has evolved over the history of the Universe. This is a really important result: since the infrared traces star-

[5] It should be noted that, unlike sub-mm telescopes, Spitzer sampled the side of the dust's black body spectrum *shortward* of the peak and, as such, was subject to negative K-correction.

formation, what they were really plotting is how the histogram (recall, a luminosity function is just a histogram) of galaxy star formation rates have evolved over the history of the Universe.

In a result that echoed that of sub-mm observations, what astronomers found using Spitzer was that the number density (i.e., number per unit volume) of LIRGs and ULIRGs in the early Universe was *orders of magnitude* higher than in the current Universe. So, the peak in the Madau diagram wasn't caused by there being a higher density of galaxies of roughly the same SFR as today, but by today's galaxies growing much more quickly in the early Universe than they are today. It's as though most of today's galaxies went through a major growth spurt around 10 billion years ago.

5.2 Caveats with infrared astronomy

Hopefully by now you will have a sense of how important infrared and sub-mm astronomy is for our understanding of galaxy evolution. Before summarising this lecture, however, I would like to raise some caveats associated with infrared astronomy.

Firstly, recall all the problems associated with sub-mm observations. All these lead to significant selection biases that lead to large (systematic) uncertainties on our results. Despite Spitzer's high angular resolution solving some of these problems, it is still the case that many of the redshifts associated with infrared sources are photometric, rather than spectroscopic, and thus come with large uncertainties.

Secondly, throughout all of this Lecture, I've breezily mentioned that we can calculate infrared luminosities from observations sometimes at single infrared wavelengths (e.g., a single sub-mm or Spitzer band). However, implicit to this are a number of assumptions, in particular the shape of the SED and the conversion from infrared luminosity to SFRs. There is a whole host of poorly understood physics which determine the shape of the infrared SED and thus the conversion between a single band observation to an infrared luminosity. And there are less well-understood physics in converting infrared luminosities to star-formation rates. As such, there are a number of (often skipped-over) systematic uncertainties in measuring SFRs from infrared observations (although it's probably no worse than any other wavelength!).

6 Learning objectives for Lecture 8

In this lecture we covered the study of galaxy evolution at infrared and sub-mm wavelengths. Because these wavelengths probe the SFRs of galaxies, they provide a direct insight into how galaxies have grown over the history of the Universe. Here are the learning objectives:

- LBG, SMG and infrared-selected galaxies each contribute similar amounts of the global star formation density at z 2.
- For SMG and infrared-selected galaxies ¡10% of the star formation is directly visible at UV wavelengths.
- The SMG and infrared-selected galaxies together contribute $>$ 50% of the star formation density at high redshifts (based on the far-IR luminosities).

Lecture 9:
The evolution of early-type galaxies

Dr. James Mullaney

March 6, 2017

1 Introduction

Now that we have covered the many ways astronomers study galaxy evolution – through simulation, deep multiwavelength surveys, and the fossil record – we'll explore how we currently believe the main types of galaxies in today's Universe have evolved to their current state. In this lecture, we'll cover early type (i.e., elliptical) galaxies, while in the following lecture we'll consider the evolution of spiral galaxies.

2 Ellipticals in the context of all galaxies

Recall from Lecture 5 that elliptical galaxies in the local Universe tend to be fairly massive, with masses typically between 10^{10} and 10^{12} $M_\odot$. Indeed, the most massive galaxies in today's Universe are elliptical galaxies that reside in the centres of dense galaxy clusters. By contrast, spiral galaxies (which we'll consider in the next lecture) dominate the numbers of galaxies with masses below about 10^{10} $M_\odot$ galaxies.[1]

Roughly 50% of the stellar mass in today's Universe is contained within elliptical galaxies, which also means that they contain about half of all stars by number. If we also include the *bulges* of disk galaxies in that number (which are sort of like mini-ellipticals) then that proportion goes up to as high at 70%. By that measure, so-called *spheroids* (which groups together ellipticals and spiral bulges) are the most important group of objects in today's Universe.

3 A recap of ellipticals from the fossil record

Another thing we covered in Lecture 5 was that elliptical galaxies tend to be quite red in colour, but also that their colour is tightly correlated with their mass: more massive ellipticals tend to be redder than less massive ellipticals.[2] We also saw how there are different types of elliptical galaxies that separate roughly according to mass:

- Boxy ellipticals tend to be the more the massive of the two, and have "cores". They are almost entirely supported by the random motions of their stars.

[1] In that respect, the Milky Way and Andromeda galaxies are relatively rare, being both massive and spiral galaxies.

[2] You should be able to explain why this is the case.

- Disky ellipticals tend to be the less massive of the two, and tend not to have “cores” - their stellar density continues to rise toward their centres. They are supported by a combination of rotation and random motions.

It is widely regarded that these differences arise due to different formation and evolutionary histories, with disky ellipticals forming via the merger of gas-rich galaxies and boxy ellipticals forming via the merger of gas-poor galaxies. Judging by the fossil record of disky ellipticals, it seems that many of their stars have been formed relatively recently (since $z \sim 1$), during an episode of intense star-formation triggered by the gas rich merger. By contrast, there is very little evidence of any recent star-formation in massive, boxy ellipticals.

4 The disky elliptical formation link to ULIRGs

If disky ellipticals are formed via the major-merger of gas-rich galaxies, then does this mean that the ULIRGs we observe in the local Universe are ellipticals in the making? We saw in the last lecture that ULIRGs are produced when two gas rich galaxies collide, their high star formation rates a result of the merger process compressing the gas and causing it to collapse to form stars. But, even though we see these collisions taking place, it doesn’t necessarily mean that they will form elliptical galaxies. After all, a major galaxy merger will typically last for tens of millions of years, so we can’t “wait and see” what will happen.

One fairly strong piece evidence that major mergers in the local Universe will ultimately lead to elliptical galaxies are the motions of stars (i.e., stellar kinematics) within ULIRGs. These are comparable in magnitude to those of elliptical galaxies and have a significant random component. Crucially, the level of rotational support vs. support from random stellar orbits is closer to that of disky ellipticals, rather than boxy ellipticals. This makes sense in terms of our understanding of how disky ellipticals are formed. The rotation is left-over from the dissipational collapse of the gas within the merging galaxies into a rotating, star-forming disk. The random orbits are those of stars that existed in the two merging galaxies *prior* to the merger. Since stars are not dissipational, any rotational velocity they had prior to the merger is converted into random motions post-merger. Finally, even the morphologies of ULIRGs in the final stages of major mergers are similar to those of disky ellipticals.

A significant amount of evidence therefore suggests that local ULIRGs represent the transformation of gas-rich spirals into disky ellipticals via major mergers. As such, galaxy evolution is not confined to the high redshift, early Universe. However, more massive disky elliptical galaxies tend to have older stellar populations compared to low or moderate mass disky ellipticals. This means more massive disky ellipticals must have formed via gas-rich mergers in the early Universe, whereas less massive disky galaxies formed from more recent mergers. This is in agreement with the “downsizing” description of the Universe in which the most massive galaxies formed first. Finally, the star-forming histories of local disky ellipticals are consistent with a large fraction of their stars being formed within a very intense, very short period of star-formation – exactly what results from a gas-rich major merger.

So, the formation and evolution of disky ellipticals can be summarised as:

- The most massive disky ellipticals were formed by major, gas rich mergers in the early Universe, as evidenced by their old stellar populations and rotation (due to dissipative gas collapse).

- Less massive disky ellipticals also formed via the major merger of gas rich galaxies, but at later times. This is another case of *cosmic downsizing*.

In both cases, a large fraction of the stars in disky ellipticals were formed in a short period of time due to the intense levels of star-formation induces during the merger. This is supported by the measured star-forming histories of disky ellipticals.

5 The formation and evolution of boxy ellipticals

As we have seen (from Lecture 5 and above), our current understanding of elliptical galaxies is that major galaxy mergers have played a crucial role in their formation. For disky ellipticals this means major, gas-rich mergers. But what about the other type of elliptical galaxies – the boxy ellipticals that dominate the numbers of high mass elliptical galaxies?

Boxy ellipticals show very little evidence of rotation, which indicates that their mergers were highly non-dissipational. This means they must have contained very little gas, as gas is *very* dissipational. As we know, this strongly suggests that boxy ellipticals are formed by gas-poor (also known as "dry") mergers. Since gas-poor mergers do not form stars (since gas is needed to form stars), this means that all of the stars that make up a massive boxy elliptical galaxy were already in place prior to the merger. All the merger process did was randomise, or "heat-up" the orbits of the stars that already existed pre-merger. Since no stars are produced in the merger process, the stellar populations in boxy ellipticals should have simply been evolving "passively" for the past few billion years.

Do we see evidence of this passive evolution? To investigate this, astronomers need to find populations of massive, boxy ellipticals at higher and higher redshifts. At any redshift, the most massive galaxies always live in the most dense environments: massive clusters and superclusters. So, to find large numbers of boxy ellipticals at different redshifts, astronomers search for increasingly distant clusters of galaxies. To determine the evolutionary state of these massive ellipticals, astronomers measure their rest-frame colours. If they have evolved passively since their formation, then boxy ellipticals at high redshifts should be bluer than local boxy ellipticals by a predictable amount.

When astronomers plot the colour-magnitude diagram for massive ellipticals in dense clusters at different redshifts, they do indeed find that earlier ellipticals are bluer by *precisely* the right amount to be explained by passive evolution. This has been confirmed up to at least $z \sim 1$, indicating that massive, boxy ellipticals have evolved passively (i.e., have not formed any significant numbers of stars) for *at least* the past $\sim$6 billion years (or roughly half the age of the Universe). Further, neither the slope nor the scatter around the colour magnitude diagram has changed significantly over this time. This suggests that the stellar populations in massive ellipticals were already well-evolved even before $z \sim 1$.

5.1 Identifying elliptical galaxies at $z > 1$

Finding dense clusters of galaxies at higher and higher redshifts becomes increasingly difficult for both practical (they're faint) and physical (they become less dense, and thus more difficult to detect) reasons. As such, beyond about $z > 1$, astronomers use other techniques to identify elliptical galaxies.

Since elliptical galaxies don't contain lots of hot, massive stars producing UV photons (i.e., they're not strongly star-forming), we can't use either the Lyman break technique to identify them, nor exploit the benefits of the positive K-correction in the sub-mm. Instead, astronomers use the fact that the old stellar populations of massive ellipticals will look extremely red due to their age. As such, they can use colour selection to find very red objects, often exploiting rest-frame near-infrared observations, which are particularly sensitive to populations of old, cool stars which dominate in massive ellipticals. Because of their colours, such galaxies are known as *Extremely Red Objects* (EROs). Although it is an effective means of identifying old populations, astronomers need to be careful to avoid galaxies that are, instead, red due to dust absorption. Indeed, roughly half of EROs are, in fact, reddened star-forming galaxies.

On studying the stellar populations of non-star-forming EROs at $z \sim 1.5-1.8$ via spectroscopy, astronomers find that their stellar populations are roughly 3.5 Gyr old. This is quite incredible, considering that the age of the Universe at $z \sim 1.8$ is only 3.7 Gyr! Clearly, this suggests that stars in early EROs must have formed very soon after the Big Bang. Even more incredibly, some of these galaxies at $z \sim 1.5 - 1.8$ have masses of 10^{11} to 10^{12} $M_\odot$, indicating that they must have formed hundreds of billions of stars in the space of just a few hundred million years! Interestingly, however, their sizes are very compact, having a half-light radius of about a 1 kpc compared to the $\sim$ 5 kpc of similar mass galaxies today.

5.2 A consistent evolutionary theory for boxy ellipticals

Collating all this observational evidence, astronomers have developed a consistent evolutionary theory to explain the appearance of massive (boxy) ellipticals at different redshifts. Their old stellar populations (even at high redshift, when the Universe was only a few Gyr old) indicates that the stars in massive ellipticals were formed at a very early time - just a few hundred million years after the Big Bang. However, these were not formed in-situ. Instead, it is thought that they were formed very early-on by the collapse of lots of comparatively small clouds of gas. Initially, these clouds of gas would have collapsed dissipationally, forming rotating stellar disks.

In the dense regions of a natal cluster, these early rotating disks of stars (which would have quickly exhausted their gas supply) would have soon merged together in dry major mergers.[3] This explains why we see ellipticals in the early Universe that contain old stars – the stars are old because they formed in these dense natal clusters from collapsing gas clouds soon after the Big Bang, and they are elliptical due to the dry merger of all these smaller, natal galaxies in rapid succession. An important consequence of successive, major dry mergers is that the galaxies will "puff-up" in size, which explains why today's massive ellipticals are larger (in physical extent) than similar mass galaxies in the early Universe.

So, to summarise, it is thought that massive ellipticals formed by:

- Clouds of gas collecting in the densest parts of the early Universe (which would become today's clusters and superclusters).
- These clouds collapsing to form stars in small "natal" galaxies just a couple of hundred million years after the Big Bang.
- The successive non-dissipative mergers of these small, now gas-poor galaxies which randomise the orbits of these stars (destroying any diskiness and rotation) and "puffs them up" in size.

[3] Remember, "major" simply refers to the *ratio* of galaxy masses, not the absolute mass. The merger of two small galaxies (e.g., each of 10^8 $M_\odot$) would still be referred to as a major merger, as the ratio of galaxy masses is 1:1

- Since the mergers are gas poor, little or no new stars are formed after the initial collapse of the early gas clouds.
- After a few billion years, all of the small galaxies that are going to merge have already done so, meaning that the masses of the most massive ellipticals stay largely the same from $z \sim 1$.

6 The evolving elliptical galaxy luminosity function

The different evolutionary paths of massive (boxy) and less massive (disky) ellipticals can be summarised nicely in terms of the evolving elliptical galaxy luminosity function.

When astronomers measure the luminosity functions of elliptical galaxies at various redshifts, they find that the number density of the brightest (and thus most massive) elliptical galaxies in the Universe has barely changed over the past ~ 6 billion years. As such, the numbers and masses of the most massive (and thus boxy) elliptical galaxies in the Universe has remained largely the same: these galaxies were already in place in the early Universe and have barely evolved since.

By contrast, the low-luminosity end of the elliptical galaxy luminosity function has evolved considerably. Today, there are many more low mass (and thus disky) elliptical galaxies than there were in the early Universe. This implies that gas-rich mergers have been slowly building up the numbers of disky ellipticals over the past 13 billion years to "catch up" with the numbers of massive boxy ellipticals that were formed much earlier on.

Thus, the evolution of elliptical galaxies is perfectly consistent with the idea of *downsizing*: that the most massive galaxies formed first in the early Universe, followed by more moderate, and then low-mass galaxies. It's just that the two different types of elliptical galaxies (boxy vs. disky) have formed via different routes.

7 Lecture 9 learning objectives

Some of this lecture is a recap of what we covered in Lecture 5 on the fossil record. As such, there has not been as much new material to read as usual. You may now want to take another brief look through the Lecture 5 notes to ensure that it all makes sense to you. Since massive ellipticals contain a large fraction of all the stars in the Universe, understanding how they have formed and evolved is a major component of our understanding of galaxy evolution overall. Here are the learning objectives for this lecture:

- Knowledge and understanding of the fossil record for early-type galaxies in the local Universe
- Understanding of the methods used to detect early-type galaxies in the distant Universe
- Knowledge of the main results obtained for early-type galaxies at high redshifts
- Understanding of the main evolutionary trends with redshift for both high and low mass early-type galaxies
- Understanding of the concept of cosmic downsizing
- Be able to describe the formation and evolution of both boxy and disky ellipticals.
- Understand how the elliptical galaxy luminosity function supports these our understanding of their evolution.

Lecture 10: Morphological evolution and spiral galaxies

Dr. James Mullaney

March 5, 2017

1 Introduction

Previously, we covered the topic of the formation and evolution of early-type galaxies. As mentioned in the last lecture, these elliptical galaxies contain roughly 50% of the stellar mass of today's Universe. In this lecture, we'll cover the formation of the other major class of galaxies in the local Universe: spiral galaxies.

2 The Milky Way's fossil record

Recall from lecture 5 that we can use the fossil record of the Milky Way to get a first-order impression of the formation of spiral galaxies. In that lecture, we learned that the Milky Way is formed of many different components, each with different star-formation histories. This points to a complex formation history consisting of both early gas collapse, current gas streaming and successive mergers.

The thin disk of the Milky Way is thought to be what's left of the gas cloud in which the fledgling Milky Way formed around 8 billion years ago. This thin disk is still being replenished by gas streaming onto the Milky Way from its immediate vicinity, providing a fuel supply for continuous star formation. By contrast, the thick disk is thought to be the result of a merger that the early Milky Way encountered when it was beginning to form (again, around 8 billion years ago), which "heated up" the orbits of the stars that had already formed by that stage. Finally, the bulge is made up of stars both from the original collapse of the early gas cloud that formed the Milky Way, and from subsequent galaxy mergers (which, again, heated the stellar orbits to create the spheroid).

However, that's just for one galaxy – our own – which could be freak event. To determine whether the Milky Way's formation history is typical of other spiral galaxies, we must study the morphological evolution of many other galaxies.

3 The morphological evolution of galaxies

We can exploit the redshift technique to study the morphological evolution of spiral galaxies. However, until recently, it was extremely difficult to do this beyond about redshift 1.5, since we need excellent spatial resolution and sensitivities to low surface brightness galaxies to resolve and measure the morphologies of distant (i.e., early) galaxies. However, redshift 1.5 corresponds to about 9.4 billion years ago, so even only going out to this "modest" redshift corresponds to studying morphological evolution across 70% of the age of the Universe.

3.1 Quantifying morphological evolution

While taking care to overcome surface brightness dimming and morphological K-corrections, astronomers can exploit the excellent seeing of the Hubble Space Telescope to study how the morphologies of galaxies have evolved over the past ~ 10 billion years. However, morphological studies are fraught with problems associated with human subjectivity. While we can easily quantify the luminosity or size of a galaxy, how do we quantify its shape in a non-subjective manner? How can you put a number on how "spirally" a galaxy is? This is a significant problem for morphological studies, with astronomers disagreeing on whether a given galaxy shows spiral structure or not (there's a few borderline cases in the lecture slides, if you want some examples). As such, it can be difficult using human classification alone to determine the degree of morphological evolution over time.

In an attempt to overcome the subjectivity problems associated with morphological classification, astronomers have attempted to identify ways to parameterise galaxy morphology, with varying levels of success. One of the most popular of these is the CAS set of parameters, which uses a simple set of measurements that can be easily made from images of galaxies. These are:

- **Asymmetry index (A)**: rotate the image of your galaxy of interest through 180 degrees, then subtract the rotated image from the original, unrotated image. Sum the absolute values of the intensities in the subtracted image, then divide by the total intensity of the original.

- **Structure index (S)**: smooth the image of your galaxy of interest then subtract the smoothed image from the un-smoothed image. Sum the intensities in the subtracted image, then divide by total intensity of the original.

- **Concentration parameter (C)**: Measure the radii that contain (i) 20% and (ii) 80% of the light of your galaxy of interest. Divide the radius containing 80% of the light by the radius containing 20%, take the $\log_{10}$ of this ratio and multiply the result by 5.

When astronomers measure these parameters and plot them against one another for various populations of galaxies, we find that they separate-out along lines of visual classification: irregular/peculiar/merger galaxies separate from spiral galaxies which further separate from elliptical galaxies.[1] As such, they provide a quantitative means of separating galaxies, reducing the subjectivity associated with visual classification alone.

3.2 The morphological evolution of galaxies

With a quantitative means to measure galaxy morphologies in-hand, astronomers can investigate how these morphologies have evolved with redshift. One of the first attempts to do this was by measuring the galaxy number counts of different galaxy types and comparing them against those expected from non-evolution models. In doing so, astronomers found that the numbers of irregular/peculiar/merger galaxies are much higher in the early Universe compared to today.

[1] Irregular/Peculiar/Mergers are often lumped together because, morphologically, it is often difficult to distinguish between mergers and irregular "clumpy" galaxies.

However, this trend is reversed for luminous spirals – there are many more luminous spirals in today's Universe compared to at $z > 0.5$.

From this and subsequent studies, astronomers have now developed a consistent picture of how the morphologies of galaxies have evolved since $z \sim 1.5$:

- At redshifts below about 0.3 (i.e., up to about 3.5 billion years ago), luminous "grand design" spirals – like the Milky Way, Andromeda, the Whirlpool galaxy – exist. As such, the Hubble diagram exist in its full form.
- At around $z \sim 0.5$ (i.e., around 5 billion years ago), barred spirals become rarer and the spiral arms of all massive spirals are much less well-defined. The splitting of Hubble's tuning fork among barred and unbarred spirals is much less obvious.
- At earlier times ($z > 0.6$, or > 6 billion years ago) the proportion of merger and irregular galaxies relative to spirals increases significantly. By around $z \sim 1$ around a third of all massive galaxies cannot by placed on today's Hubble diagram (in other words, they have irregular/peculiar/merger morphologies).

4 The morphological evolution of field galaxies

We saw in the last lecture that the most massive ellipticals have lived in the densest regions of the Universe throughout all of cosmic time. By contrast, other types of galaxies have predominantly lived in the less dense regions of the Universe between the clusters. Galaxies within these more sparse regions are known as "field" galaxies, with the more massive ones dominated, in terms of number, by spiral galaxies. So, if we want to know how spiral galaxies have evolved, we therefore need to study massive field galaxies out to high redshifts. One of the most comprehensive attempts to do this has been the Canada-France Redshift Survey (CFRS).

4.1 The Canada-France Redshift Survey

To determine the environments of galaxies in 3 dimensions requires both their position on the sky (i.e., x, y position) and their redshift. The more precise the redshift measurement, the more reliably astronomers place a galaxy at a given location, and thus in a given environment. Thus, the reliable identification of large samples of field galaxies benefits greatly from spectroscopic redshifts for lots of galaxies.[2] The development of multiplexing spectrographs on large telescopes in the mid-90s greatly advanced the study of field galaxies by providing redshifts measurements for large samples of galaxies.

Using the multi-slit spectroscopic instrument on the Canada-France-Hawai'i telescope on Mauna Kea, Hawai'i, Lilly et al. (1995) took the spectra of around 600 field galaxies. This survey is "complete" to all galaxies within the survey area with redshifts < 1, with B-band absolute magnitudes brighter than -20.4, and masses between $3 - 30 \times 10^{10}$ $M_\odot$. In other words, it samples the population of moderate mass field galaxies out to $z \sim 1$; in other words, the galaxies that would eventually become today's massive spirals.

The CFRS survey has had a significant impact on our understanding of how field galaxies have evolved. For example, the lower redshift sections of the Madau diagram (i.e., $z < 1$) are largely

[2]This is less of a requirement for dense environments, as these can be identified as overdensities in x-y space, then confirmed with redshift measurements of just a handful of cluster galaxies.

based on star formation rates measured via the [O II]$\lambda 3726\mathring{A}$ emission line in CFRS spectra.[3] One the greatest impacts that the CFRS survey has been, however, on our understanding of the morphological evolution of field galaxies.

4.2 The morphologies of CFRS galaxies

With spectroscopic redshifts from the CFRS in hand, astronomers could reliably identify field galaxies for more detailed follow-up. With this in mind, Hammer et al. (2005) obtained high resolution Hubble observations for a sub sample of 185 CFRS galaxies in order to measure their detailed morphologies. These detailed morphological studies confirmed the results from number counts described above: that the population of massive field galaxies contained a significantly greater proportion of irregular/peculiar/merger galaxies at $z > 0.4$ compared to today, while the relative numbers of spiral galaxies at these redshifts are significantly reduced compared to locally. For example, while around 70% and 7% of $\sim 10^{11}$ $M_\odot$ galaxies in today's Universe are spirals and irregular/peculiar/mergers, respectively, roughly 43% and 34%, respectively, are at $z > 0.4$. Furthermore, the number density of LIRGs (see lecture 8) at $z > 0.4$ among $\sim 10^{11}$ $M_\odot$ galaxies is roughly 30 times higher at $z > 0.4$ than it is today, with 64% of LIRGs at these high redshifts having irregular/peculiar/merger morphologies. As such, the CFRS provides clear evidence of *significant* morphological evolution of field galaxies since $z > 0.4$.

As the changing proportions of LIRGs suggests, the CFRS also provided evidence of strong evolution in the star-forming properties of $\sim 10^{11}$ $M_\odot$ field galaxies. Most of the spectra of CFRS galaxies show clear evidence of recent star-formation, again confirming that the star formation density of the Universe was significantly higher at $z > 0.4$. Importantly for this lecture, however, is that the *bulges* of $z > 0.4$ galaxies are *bluer* than those of similar mass galaxies today. This can be interpreted as evidence of increased levels of star formation within the central regions of $z > 0.4$ galaxies relative to galaxies in the local Universe. Of course, these enhanced levels of star-formation led to the production of heavy elements, which explains why galaxies in the CFRS have roughly half the metallicity of today's field galaxies (i.e., the metals were still in the process of being made in the $z > 0.4$ field galaxies).

Before constructing a theory for the morphological evolution of field galaxies (which, remember, are dominated by Spiral galaxies today), let us, in true "Look through the keyhole"-fashion, consider the evidence:

- The numbers of $\sim 10^{11}$ $M_\odot$ field galaxies are dominated by irregular/peculiar/merger galaxies at $z > 0.4$, whereas today they dominated by spirals.
- The proportion of $z > 0.4$ field galaxies that show evidence of rapid, recent star formation is significantly higher than for field galaxies today.
- There is clear evidence of enhanced levels of star formation in the central regions of $z > 0.4$ galaxies compared to today's spirals.
- The metal abundances of $z > 0.4$ field galaxies is roughly half that of local galaxies.

[3] Like Hα, [O II]$\lambda 3726\mathring{A}$ is only produced by high energy photons, and thus provides a measure of the numbers of massive, young stars in a galaxy. The reason [O II]$\lambda 3726\mathring{A}$ is often used in preference over Hα in the distant Universe is because Hα gets shifted out of the optical bands at $z > 0.4$, whereas [O II]$\lambda 3726\mathring{A}$ remains observable in optical spectra until $z \sim 1.5$.

5 The formation of local spirals

To explain the above observations, astronomers have developed a theory of how today's field galaxies have evolved:

5.1 An early merger phase

The irregular/peculiar/merger morphologies of early field galaxies suggests that most have undergone some form of major, gas-rich merger during the past 8 billion years or so. During this time, the disk is suppressed by the random motions of the stars while the dissipational gas falls toward the centre of mass of the system. This merger is associated with a sudden burst of star-formation, but which only lasts for a few hundred million years. This explains the LIRG nature of a lot of these early irregular/peculiar/merger galaxies. The merger phase is then followed by...

5.2 A compact galaxy phase

Many of the stars that formed during the intense episode of star-formation triggered by the gas-rich merger have now fallen toward the central regions of the galaxy. These stellar populations are still comparatively young, which explains the blue colours of the bulges of early field galaxies. The star formation rate falls significantly over the next couple of billion years from its peak during the merger. Any further gas remaining from the merger that has not already fallen into the central regions may dissipationally collapse to start to form a fledgling rotating disk around the central bulge.

5.3 Growth of disk phase

Stars form from the collapsing disk of gas left over from merger, but at a much slower rate than before. Additional gas is then accreted by the galaxy (due to its high mass) from the surrounding regions. Irrespective of the trajectory along which this accreted gas falls onto the galaxy, because of the dissipational nature of gas collapse it will ultimately become part of the rotating disk. This gas feeding is responsible for the observed low level of star formation in spiral galaxies since $z \sim 0.4$. The galaxy may then go through a series of successive *minor* mergers which, since too small to disrupt the disk, will actually build up the mass of stars and gas in the disk.

5.4 Plus downsizing

Finally, it should be noted that while the above applies to field galaxies in general, precisely *when* each phase occurred during the history of the Universe changes with galaxy mass in a way consistent that is with downsizing. For example, while moderate mass (i.e., $\sim 10^{11}$ $M_\odot$) field galaxies grew most of their mass at $z < 1$, it is thought that the most massive field galaxies (such as Andromeda) will have started to form at even earlier times (i.e., $z \sim 2$).

6 Massive disks forming at $z > 2$

Of course, there has been significant advancements in observing facilities since the first CFRS studies. The development of 8m class telescopes and, in particular, adaptive optics has helped us to obtain high spatial resolution integral field spectra for large numbers of galaxies at high redshifts.

Such spectra enables us to map-out the kinematics of galaxies at high redshifts, which can be used to identify rotating disks or major mergers *kinematically*.

These surveys have reported evidence of rotating disks among moderately massive ($\sim 10^{11}$ $M_\odot$ galaxies at $z > 2$. Indeed, at these redshifts roughly two thirds of the irregular/peculiar/merger galaxies appear to be "clumpy" galaxies that are in the process of forming disks of stars, with the remainder being mergers. As such, it has been suggested that these are today's massive spirals (like the Milky Way) and disky ellipticals forming at high redshifts.

7 Lecture 10 learning objectives

In this lecture, we have covered the evolution of the other main type of large galaxy in today's Universe: spiral galaxies. These dominate the numbers of field galaxies in the local Universe, so by studying the morphological evolution of field galaxies, we can build up a theory to explain the build-up of today's spirals. The learning objectives for this lecture are:

- Revise the fossil record of the Milky Way.
- Understand how astronomers quantify galaxy morphologies (i.e., CAS).
- Have an appreciation of the how Hubble number counts and the CFRS have provided evidence of the morphological evolution of field galaxies .
- Be able to describe our current understanding of how today's spirals have evolved, together with the observational evidence that backs this up.

Lecture 11:
AGN discovery and observed properties

Dr. James Mullaney

March 17, 2017

1 Introduction

Up to this point in the course we have considered "normal" galaxies whose light is produced exclusively by stars. There is, however, a class of galaxy in which a significant proportion of the total energy output is being emitted instead by material falling into a central supermassive black hole. Because of the extreme amounts of light being produced at their centres, these are known as "Active Galactic Nuclei" (AGN) and are currently thought to have had a major influence on how today's galaxies have formed and evolved. In this lecture, we'll cover how AGNs were first discovered and discuss their main properties.

2 The discovery of AGN

The earliest known account of the study of an AGN is that of Fath in 1908, in which he obtained the spectra of a number of "spiral nebulae" (this was before Hubble had demonstrated that such nebulae were, indeed, galaxies external to our own). Fath noted that while most of the spiral nebulae in his sample displayed spectra consistent with a population of unresolved stars, one – NGC 1068 – had a "composite spectrum, showing both bright [emission] and absorption lines". In the years that followed, a few other observers noted the peculiar "planetary nebulae-like" spectra of some other spiral galaxies (notably NGC 4051 and NGC 4151), but little more was made of these somewhat rare systems.

It wasn't until 1943 that AGNs were studied in any kind of systematic way. At this time, Carl Seyfert obtained the spectra of six nearby galaxies that all showed evidence of a bright, stellar-like nucleus. Seyfert noted that all of these galaxies showed evidence of strong emission lines superimposed on an otherwise normal galaxy spectrum. Most interestingly, however, was that in some cases (e.g., NGC 3516, NGC 7469), the permitted Hydrogen emission lines (i.e., Hα and Hβ were extremely broad (i.e., they covered a wide range of wavelengths). Seyfert attrbuted these broad lines to Doppler shifts, corresponding to velocities of up to 8,500 kms^{-1}. By contrast, in other cases (e.g., NGC 1068) the permitted Hydrogen emission lines had similar widths as the comparatively narrow (corresponding to up to $\sim$ 3000 kms^{-1}) forbidden emission lines. [1] As we shall see, the differences in the widths of the permitted emission lines continue to be used today to group AGNs into two different types:

[1]Unfortunately, Carl Seyfert died in a car accident before his important work on AGN was fully appreciated. He did, however, present the daily weather bulletin the local news whilst employed as full-time astronomy professor in Tenessee, which makes him awesome.

- Type 1 AGNs: Display broad (> 2000 km s^{-1}) permitted emission lines (e.g., Hα, Hβ), but narrow (typically < 1000 km s^{-1}) forbidden (e.g., [O III], [N II]) emission lines in their optical spectra.
- Type 2 AGNs: Display permitted and forbidden emission lines with roughly the same comparatively narrow (typically < 1000 km s^{-1}) widths.

In addition to differences in their emission lines, Type 1 and Type 2 AGNs also tend to have different continuum features. Type 1 AGNs typically have very blue continua that are quite devoid of absorption lines. Type 2 AGNs, by contrast, have much redder continua which display evidence of stellar absorption features. Indeed, the spectra of Type 2 AGNs typically look like those of normal galaxies, but with strong narrow permitted and forbidden emission lines.

3 AGN in the radio era

While Seyfert's work is now widely regarded as as groundbreaking in terms of the first systematic study of AGNs, it was largely ignored by other astronomers at the time. In fact, it took until the development of radio astronomy before AGNs would start to more widely studied by astronomers.

During the Second World War there was a lot of research conducted in the areas of radio communication and radar. Once the war ended, some of the engineers and scientists that had worked on radio communications during the war turned their attention to trying to determine where the persistant background "noise" that was picked up by their receivers came from. While it was soon realised that a significant fraction of radio noise came from the Sun, there remained significant "noise" from other non-terrestrial sources. The problem, however, was that the angular resolution of early radio telescopes was extremely poor (i.e., many degrees on the sky), making it very hard to pinpoint the positions of the radio sources to better than a few tens of degrees. With this level of resolution, radio astronomers were only able to locate a bright sources to be within an entire constellation, so the first radio sources were referred to by the constellation in which they were detected (e.g., Cygnus, Cassiopeia).

By the mid 1940s, radio astronomy technology had developed sufficiently to begin to pinpoint the positions of radio sources beyond the solar system. This was enabled by the development of radio interferometry techniques, the first of which exploited the "sea-cliff" technique. This involved situating a radio reciever at the edge of a sea-facing cliff. As the radio source rose above the horizon, some of its radio waves would travel directly toward the receiver, while others would be reflected off the surface of the sea before reaching the receiver. This effectively creates two radio sources that interfere with each other, creating a radio interference pattern. The end result is an interferometer with a baseline twice the height of the cliff, greatly increasing the angular resolution of the telescope without the need for connecting cables and correlators between two separate receivers.

Using this technique, early radio astronomers began to be able to pinpoint the position of bright radio sources on the sky. This revealed that the very bright source in Cygnus previously detected in low-resolution studies was compact, extending less than 8 arcminutes on the sky. With such levels of angular precision, by 1949 astronomers were able to associate the radio sources with those detected at optical wavelengths. In doing so, it transpired that Cygnus A was, in fact, associated with galaxy merger at a distance of 252 Mpc, meaning it has a total radio luminosity of $> 10^{45}$ ergs s^{-1} – more than the luminosity of all the stars in the merging galaxies combined. With the ratio of radio to optical luminosity of Cygnus A so much higher than that of a star, it must mean that the radio emission must be coming from another type of source. Later, in the 1960s, it was realised that this alternative source of radio energy was associated with galaxies showing optical properties the same as those reported by Seyfert around 20 years earlier - i.e., AGNs.

4 AGN classification

With the realisation that many of the brightest radio sources in the sky were associated with AGNs, interest in these objects increased dramatically. However, with some of Seyfert's original optical AGNs *not* associated with bright radio sources, it was soon appreciated that not all AGNs were powerful radio emitters. Instead, it seemed that in addition to the Type 1/Type 2 classification, there should also be a radio "loud" and radio "quiet" distinction.[2] Further, both radio loud and radio quiet AGNs can either be Type 1 or Type 2 AGNs; in other words, there is no obvious connection between an AGN's optical classification and its radio classification. So, there are:

- Radio loud:
  - Type 1: Broad Line Radio Galaxy
  - Type 2: Narrow Line Radio Galaxy
- Radio quiet:
  - Type 1: Seyfert 1
  - Type 2: Seyfert 2

Added to this, however, is another completely arbitrary distinction based on the *optical luminosity* of AGNs, with the most optically luminous AGNs known as *Quasars*. This unfortunate situation has arisen because quasars are so bright that when they were first discovered, their host galaxies could not be seen due to the "glare" of the light from the AGN. Indeed, they looked like very bright stars but with peculiar spectra (hence their name, which refers to "quasi-stellar objects"). As a consequence, their connection to Seyfert AGNs (which are all in easily-seen galaxies) was not realised at first, and by the time it was, the name had stuck. So, now we have:

- Radio loud:
  - Type 1:
    - High optical luminosity: Radio Loud Quasar
    - Low optical luminosity: Broad Line Radio Galaxy
  - Type 2:
    - High optical luminosity: Radio Loud Type 2 Quasar
    - Low optical luminosity: Narrow Line Radio Galaxy
- Radio quiet:
  - Type 1:

[2] "Loud" and "Quiet" typically refer to the ratio of radio luminosity to optical luminosity of an AGN, rather than the absolute radio luminosity. Despite being adjectives of sound volume, this is referred to as a measure of the radio *loudness* of a source, rather than its radio "volume" (since the latter has a double meaning in physics).

    * High optical luminosity: Radio Quiet Quasar
    * Low optical luminosity: Seyfert 1
  - Type 2:
    * High optical luminosity: Radio Quiet Type 2 Quasar
    * Low optical luminosity: Seyfert 2

What a nightmare!

5 AGN Unification

As the tale of the discovery of quasars indicates, as the different types of AGNs were being discovered, it wasn't at all clear to astronomers that they were, indeed, the same type of object. Are, for example, Type 1 AGNs completely distinct from Type 2 AGNs, or are they somehow manifestations of the same object?

The anwer to this specific question came in the mid 1980s as a result of the study of Type 2 AGNs in polarised light. When light is reflected, it becomes polarised, so when it was discovered that Type 2 AGNs showed evidence of Type 1-like broad emission lines in their optical spectra, it was interpreted that what was being detected was the reflected light from an otherwise "hidden" region. This led to the suggestion that the region responsible for the broad lines in Type 1 AGNs – i.e., the *broad line region* (BLR) – was indeed present in Type 2 AGNs, but was simply hidden from our view. However, the obscuring material could not be blocking *all* the light from this BLR, since some of it was being observed in polarised light. The interpretation was that surrounding the BLR of *all* AGNs is a "torus" of obscuring dust; in the case of Type 1 AGNs, we're looking down the hole of the torus, directly at the BLR, whereas in the case of Type 2 AGNs, the torus is side-on and thus blocks our line-of-sight to the BLR.

Today, this "dusty torus" is a major component of the "Unified AGN model", which explains the observed properties of AGNs in terms of a specific geometry:

- At the very centre of an AGN is a supermassive black hole.
- Surrounding the supermassive black hole is an accretion disk formed from gas and dust sprilling toward the black hole. As this material falls toward the black hole, it travels faster and increases in temperature due to friction. As it heats up, it radiates this heat in the form of light, sometimes outshining its entire host galaxy (as is the case of quasars).
- Slightly further out from the accretion disk is the BLR. The BLR consists of clouds of gas that are being illuminated and ionised by the light from the accretion disk. Being close to the black hole, they orbit it at high velocities, hence the emission lines they produce are strongly doppler broadened. The are also dense, and so only emit permitted emission lines.
- Surrounding the accretion disk and BLR is the dusty torus which, in the case of Type 2 AGNs, blocks out the light from *both* the accretion disk and the BLR.
- Beyond the dusty torus is the so-called *narrow-line region* (NLR). This is simply ambient gas in the host galaxy that is being illuminated and ionised by the AGN. This gas is far more rarified than that in the dense BLR, and hence emits in both permitted and forbidden emission lines. Since it is larger than the dusty torus, we observe emission from the NLR in both Type 1 and Type 2 AGNs.

Within the unified model, Type 1 and Type 2 AGNs are manifestations of *the same* object. The reason for their different appearances is simply due to the orientation of the dusty torus relative to our line of sight.

6 The multiwavelength continua of AGNs

An important consequence of the structure of an AGN is that they can emit strongly at almost all observable wavelengths, with each component of the unified model emitting at different wavelengths. With its high temperatures, the accretion disk emits strongly at optical through to ultraviolet wavelengths which can be seen directly in the case of Type 1 AGNs. Furthermore, some of the gas surrounding the accretion disk is heated to millions of degrees, which upscatters ultraviolet photons to X-ray energies. As such, AGNs are the strongest sources of X-rays in the Universe. Further out, the dusty torus is heated by light from the accretion disk which is then re-radiated at infrared wavelengths. At even longer wavelengths, in radio-loud AGNs, magnetically entrained jets propogate from the central regions of the accretion disk, emitting powerful synchrotron radiation as they do so.

7 AGN identification

Since AGNs are strong emitters of light spanning almost the whole observable electromagnetic spectrum, many different techniques emploiting different wavelengths of light have been developed to identify AGNs among "normal" galaxies. In this section, we'll consider some of the most commonly used approaches to identify AGNs. It is important to note, however, that the various techniques come with their own biases; for example, some are only sensitive to a particular type of AGN (e.g., optical Type 1 or Type 2), whereas others will only identify particularly luminous AGNs.

7.1 Optical selection

With Type 1 AGNs displaying strong blue optical continua, we can use this property to identify AGNs using broad band optical photometry (i.e., using imaging taken with different filters). Basically, what astronomers do is search for point like sources with blue rest-frame optical colours. Of course, this has to be able to account for K-correction, so different combinations of filters are used to identify the blue continua indicative of quasars at different redshifts. However, the AGN continuum must dominate strongly over the continuum of the rest of the galaxy for this selection to work. As such, this technique is biased in favour of the most optically luminous AGNs whose light dominates ovet the host galaxy. Of course, it only also only applies to finding Type 1 AGNs, since the blue optical continuum from the accretion disk is blocked by the torus in Type 2 AGNs.

7.2 Radio selection

This is an easy one. Roughly 10% of all AGNs are radio loud, so we can use radio surveys to identify them. However, since AGNs are not the only astronomical sources that produce radio emission we need to be careful to account for other types of sources that could contaminate our radio sample. The biggest type of contaminant are star-forming galaxies (since star-formation is accompanied by supernovae, whose remnants are strong radio emitters) and supernovae remnants in the Milky Way. However, almost every source with radio luminosities greater than 10^{23} W Hz is a radio loud AGN, so it is fairly easy to exclude contaminants via a radio luminosity cut. However, this does mean that radio selection is biased toward radio luminous AGNs (since AGNs with low radio luminosities will be thrown out with the star-forming galaxies). Since radio wavelengths are immune to dust obscuration, radio selection picks out both type 1 and type 2 radio AGNs.

7.3 Infrared selection

Being exposed to lots of high energy photons from the accretion disk, the dust surrounding the AGN gets very warm (indeed, the dust nearest the accretion disk gets so hot it actually evaporates). As such, AGNs are strong sources of infrared emission and have characteristic infrared colours. We can therefore use infrared colour selection to identify AGNs. However, as we have seen, low mass stars and star-formation can also produce a lot of near and mid infrared emission, respectively, so the emission from the AGN must dominate over the emission from the host galaxy in order for it to be identified via infrared selection. As such, infrared selection is biased in favour of AGNs above a certain AGN luminosity to galaxy mass, or AGN luminosity to SFR ratio. However, since infrared is relatively impervious to dust, infrared selection identifies both Type 1 and Type 2 AGNs.

7.4 Emission line selection

The accretion disk at the centre of an AGN – the "central engine" – produces large amounts of high ionising UV photons. Moreso even than the hottest, most massive stars. This intense UV flux is capable of ionising gas in the narrow line region to very high levels – emission lines from species with five or six missing electrons are not uncommon in the spectra of AGNs. This means that AGN spectra contain strong emission lines from highly ionised species, such as [O III]. Because the relative strengths of different emission lines in AGN spectra differ from those of non-AGNs, we can use emission line ratios to identify AGNs. The most commonly used emission line ratios are [O III]/Hβ and [N II]/Hα, with AGNs having high values of each of these ratios compared to non-AGN galaxies. Since each pair of emission lines are close to each other in wavelength (i.e., [O III] at 5007 Å, Hβ at 4861 Å, and [N II] at 6548 Å, Hα at 6563 Å) the effects of dust on each ratio cancels out.

As with all the AGNs selections highlighted here, however, emission line selection does have its pitfalls. Firstly, obtaining spectra is time-consuming, and only a fraction of all the galaxies in an imaging survey will have been observed spectroscopically. Further, as with all AGN selection techniques, the AGN features must dominate over any features arising from the host galaxy. This means that only AGNs with strong emission lines relative to those arising from the host galaxy will be selected by emission line selection. However, if measured carefully, emission line selection can identify both Type 1 and Type 2 AGNs.

7.5 X-ray selection

As mentioned previously, AGNs are the strongest emitters of X-rays in the known Universe, with some AGNs emitting $> 10^{46}$ ergs s^{-1} in X-rays alone. Since no other objects produce such large amounts of X-rays, X-ray observations are a relatively "clean" way of identifying AGNs. Further, since X-rays are emitted by the central regions of the accretion disk itself, they give us a very good measure of the "instantaneous" luminosity of the AGN.

The key pitfall of X-ray emission, however, is that it can be obscured by large columns of gas and dust. However, X-rays are less susceptible to absorption than the UV and the shorter wavelengths of optical light, and the highest energy photons (i.e., energies about about 10 keV) can penetrate very dense clouds of gas and dust. Since X-rays aren't totally impervious to obscuration, however, they are more likely to identify Type 1 AGNs relative to Type 2 AGNs (but like-for-like they find relatively more Type 2s than optical selection).

8 Lecture 11 learning objectives

In this lecture we took a turn away from normal galaxies to study AGNs. As we shall see, astronomers now think that AGNs have played a major role in shaping today's galaxies, so we'll explore them further for the next six lectures. For now, however, here are the learning objectives from this lecture:

- Know how AGNs were first discovered and studied at optical wavelengths during the first half of the 20th century.
- Have an appreciation of the role that radio astronomy has had in raising the interest in AGNs.
- Know the various different classes of AGNs, and understand the differences between them (especially the difference between a Type 1 and Type 2 AGN).
- Be able to describe the AGN Unified Model, and explain how it accounts for the different AGN types.
- Know the main ways we identify AGNs, and the benefits and pitfalls of each approach.

Lecture 12:
AGNs and supermassive black holes

Dr. James Mullaney

March 20, 2017

1 Introduction

Last lecture, we took our first look at Active Galactic Nuclei (AGNs). We learned how they were discovered, the different types of AGNs and how they are "unified", and how different parts of the electromagnetic spectrum are used to distinguish AGNs from "normal" galaxies. In this lecture, we'll consider where all the energy in an AGN comes from by taking a closer look at black hole and accretion disk system that forms the so-called "central engine".

2 Central engine characteristics

As we saw in the previous lecture, the first AGNs to be studied in any kind of detail were moderate luminosity "Seyfert-like" AGNs. While these first studies highlighted the curious observed properties of the nuclei of these galaxies, they didn't make any real attempt to explain their cause. This all changed when the distances to the first known quasars were measured and the consequent appreciation of their extreme luminosities. With the brightest known quasars having luminosities equivalent to 10,000 Milky Way-like galaxies, you can see why astronomers quickly began to question what could be the generating so much power.

Clearly, any theory to explain what powers a quasar must be able account for their extreme luminosities. However, it was soon realised that the power was not the only peculiar aspect of AGNs that any model of energy production must be able to explain. Take their optical spectra, for example. Type 1 AGNs in particular display spectra that are unlike any other type of object; no other celestial body displays emission lines with widths corresponding to thousands or even tens of thousands of km s^{-1}. What was causing such extreme velocities? In addition, any theory of energy production must be able to explain the very large radio luminosities of AGNs. Indeed, as the angular resolution of radio telescopes improved, it was discovered that the radio emission of some AGNs is being emitted by powerful radio jets that extend over many tens or even hundreds of kpc. Any model of the central engine must be able to explain what is launching such powerful jets (which, recall, can be more luminous at radio wavelengths than the *bolometric* output of all the stars in the host galaxy).

An interesting and important consequence of the discovery of *extended* radio jets was that it revealed that some AGNs had been generating power for long enough to launch jets that are millions of light-years long. Clearly this implies that some AGNs had been generating vast amounts of energy for millions of years, and were thus not the type of short "explosive" events normally associated with more familiar extreme phenomena such as supernovae. Clearly, any theory of energy production must therefore be able to explain how this energy output can be maintained for so long.

2.1 AGN variability

Of all the extreme properties of AGNs, perhaps the most puzzling – yet ultimately revealing – feature for astronomers was their rapid variability. Soon after the first discovery of quasars it was realised that their optical flux varied on timescales of a few days to years. While they always stayed extremely luminous, their brightness would change be a factor of a few within a relatively short period of time.

The reason why the rapid variability of quasars is so interesting is because how rapidly something varies reveals something about its size. For example, it would be impossible for anything as large as a whole galaxy to vary in brightness in the space of a year. This is because it takes time for the "information" regarding that change to be transmitted from one side of the galaxy to the other. The fastest any information can travel is the speed of light, so the fastest something as large as a galaxy can vary in brightest would be a few tens of thousands of years. So, if a quasar changes brightness over a period of just a few days, then it implies that its *maximum* size is a few light-days across. This means that the central engine of some AGNs is, at maximum, not much bigger than our solar system – an absolute minuscule size in galaxy terms. So, early AGN astronomers were faced with the incredible concept that more power than that emitted by all of the stars in the Milky Way is coming from a region of space not much bigger than our solar system!

3 Suggested AGN energy sources

There's an old saying that goes something like: "if the only tool you have is a hammer, you'll treat every problem as if it's a nail". Up until the middle of the 20th century, the only thing astronomers had ever known was stars and planets (galaxies are, after all, just collections of lots and lots of stars), so it was tempting to try to solve the "problem" of the AGN power source by appealing to stars. However, as we shall see, all of these theories had extreme shortcomings.

3.1 A 10^6 solar mass star

Among the first suggestions for what could be powering bright quasars was that it was a single extremely massive star. As we have seen, the most massive stars in a galaxy are very hot and are disproportionately bright for their mass (i.e., a star ten times more massive than the Sun is more than ten times more luminous). Perhaps inspired by the stellar-like appearance of quasars, some early AGN astronomers suggested that AGNs were, in fact, a single star with a mass of over a million solar masses. Such a star would, indeed, be extremely bright and would also emit strongly at ultraviolet wavelengths and could therefore explain the blue continua of AGNs. A very massive star would also satisfy the compactness requirements, and its strong gravity could be used to explain the broad emission lines through gravitational doppler broadening.

Despite accounting for some AGN properties, the massive star theory falls down on many key points. Firstly, such a massive star would be extremely short lived, possibly only lasting a few tens of thousands of years – not long enough to maintain the extended radio jets seen in some AGNs. Speaking of which, there is no known star that produces such collimated radio jets; how would a massive star do so? Further, stellar masses are known to form a continuum distribution (i.e., the initial mass function), so if there are million solar mass stars powering AGNs, where are all the slightly smaller, hundred thousand solar mass stars?

3.2 A massive star cluster

Another star-based early suggested model for the central engines of AGNs was a very dense central stellar cluster consisting of thousands or even millions of massive (i.e., > 20 $M_\odot$) stars. Being massive, these stars would quickly undergo supernovae one after another, which would explain both the high luminosity and variability of AGNs. Also, the combined radio emission of the supernovae remnants could explain the high radio luminosities of AGNs.

While it would be technically possible for this number of massive stars to be packed into a volume the size of the solar system, this model has a number or drawbacks. Firstly, it isn't clear how such a setup would create the collimated jets we see being launched from AGN. Most importantly, however, is that the spectral features of AGNs simply do not look like those of supernovae remnants

3.3 A central supermassive black hole

With all the attempts to use stars to explain the properties of AGNs looking extremely contrived (and yet still failing to explain key observed AGN features), a few theorists started to consider other non-stellar possibilities. As we now know, the most successful of these was the idea that a supermassive black hole is, ultimately, responsible for the AGN phenomena. With black holes the densest form of matter in the Universe, they are ideal candidates to meet the "compactness" requirement. However, since black holes by definition emit no light, how can they account for the most fundamental property of AGNs: their extremely high luminosities?

As we mentioned in the previous lecture, the key to an AGN's luminosity is not the black hole itself, but the accretion disk that forms around the black hole as it consumes surrounding gas and dust. As the material falls toward the black hole, it releases its gravitational potential energy in the form of light. In fact, this release of gravitational energy as material falls onto a supermassive black hole is an incredibly efficient process, with roughly 10% of the rest mass of the falling material being converted into light. This compares to $< 1\%$ efficiency for nuclear fusion in stars. Indeed, the release of energy due to gravitational collapse is the second only to matter/anti-matter annihilation in terms of its efficiency in converting rest mass into radiated energy. As such, even a relatively modest accretion rate of about one solar mass per year is sufficient to power a quasar of bolometric luminosity $\sim 5 \times 10^{45}$ erg s^{-1}.

So, an accreting supermassive black hole can explain both the power output and variability of AGNs (the latter because of its highly compact nature). But what about those radio jets seen in a subset of AGNs? Can an accreting black hole explain those as well? Well, one popular theory is that as the accreting material gets hotter and hotter, it starts to become ionised. As such, it forms a fast-moving plasma, which generates a strong magnetic field (as all moving charges do). It is thought that if the black hole is spinning, it will tangle the magnetic field lines up into a spiral pattern. Any electrons and plasma caught within this magnetic field will be accelerated to close to the speed of light, forming tightly collimated jets.

Finally, as we saw in the previous lecture, a supermassive black hole plus accretion disk system can explain the very broad emission lines seen in the spectra of Type 1 AGNs. As a reminder, these broad lines are being emitted by dense, ionised gas orbiting close to the black hole, which explains their extremely high velocities.

4 Finding evidence for supermassive black holes

While it seems that accreting supermassive black holes can explain very well many of the observed properties of AGNs, this does not constitute a proof that it is correct. For that, we would ideally identify evidence of supermassive black holes in the centres of galaxies. Before doing so, however, it would be useful to have some kind of indication of the expected masses of a supermassive black hole, just so we've got some idea of how challenging measuring them will be.

4.1 The Eddington Limit

One way of obtaining a rough lower limit for the mass of an accreting supermassive black hole is by assuming that they are accreting close to their Eddington luminosity. You should recall from previous lecture courses that the Eddington luminosity for an accreting object (whether a star or a black hole) is reached when the gravitational force pulling the material inward is balanced by the radiative force from the photons emitted by that accreted material. Equating these two forces gives:

$$L_{\mathrm{Edd}} = \frac{4\pi G M m_p c}{\sigma_T} \tag{1}$$

where G is the gravitational constant, M is the mass of the central object (in our case, a supermassive black hole), m_p is the proton mass, and σ_T is the Thompson cross-scattering area. Entering the numerical values for the constants, and converting the units gives:

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

We have already seen that the brightest quasars accrete at a rate of roughly one solar mass per year, producing a bolometric luminosity of about 5×10^{45} erg s^{-1}. Plugging this luminosity into the above equation gives a black hole mass of $5 \times 10^{45}/1.3 \times 10^{38} \approx 4 \times 10^{7}$ M$_{\odot}$. However, this assumes the black hole is accreting at its Eddington luminosity, whereas it is likely to be accreting at a rate somewhat lower than this. As such, this gives a *lower limit* to the mass of the black hole. Clearly, this is much more massive than a conventional, stellar mass black hole (which may reach a few tens of solar masses at most).

4.2 The Soltan Argument

Another way of getting an estimate of the mass of all supermassive black holes in today's Universe is to (a) sum up all the material that has ever fallen into a supermassive black hole and (b) assume that this mass is distributed evenly among all massive galaxies. This clearly assumes that all massive galaxies contain a central supermassive black hole, and thus again represents a lower mass limit. This methodology was first proposed by Soltan in 1982, and so it referred to as the Soltan Argument.

Part (a) sounds incredibly difficult: How can we possibly measure all of the mass that's ever fallen into a supermassive black hole? In practice, however, it's relatively straightforward (given some reasonable assumptions and some careful measurements). Black holes grow by accreting material, and as we have seen, the amount of light they emit is directly proportional to their accretion rate (assuming a given accretion efficiency). This means that the AGN luminosity function (i.e., the histogram of AGN luminosities per unit volume) provides a measure of the accretion rate density of the Universe. The nice thing about black holes, is that once matter has fallen into the black hole, it can't get back out.[1] As a consequence, integrating all these accretion rates over time gives the total amount of mass that accretes into a black hole in a given time interval. All we need to do, therefore, is measure the evolving AGN luminosity function and integrate it with respect to time to give total amount of mass (per unit volume) contained within today's black holes. Thankfully, since AGNs are bright, we are able to measure their luminosity function across large swathes of cosmic time and perform this calculation.

When astronomers integrate the evolving AGN luminosity function over cosmic time, they estimate that they get for the total amount of mass contained within black holes per cubic Mpc is about 5×10^5 $M_\odot$. Dividing this by the number density of massive galaxies in today's Universe (i.e., about 0.01 galaxies per Mpc^3 gives an average black hole mass of about 5×10^7 $M_\odot$ per massive galaxy. Again, this is far more massive than a typical stellar-mass black hole.

4.3 Weighing supermassive black holes

The above arguments make a strong case for the black holes at the hearts of AGNs to be supermassive (i.e., far more massive than normal, stellar mass black holes). They are, however, still just predictions. It would be nice to be able to actually measure the mass of these black holes directly. To do this we need to be able to measure the motion of "test particles" moving under the influence of the black hole's gravity.

4.3.1 Stellar orbits near the BH

One way to achieve this is by measuring the velocity of stars very close the black hole. However, the density of stars in the central regions of a galaxy is so high that the gravitational force of the central supermassive black hole only dominates (over the stars in the galaxy) within about 30 pc from the black hole. It took until the launch of the Hubble Space Telescope before these small scales could be anywhere near resolved in nearby galaxies.

Using the Hubble Space Telescope, astronomers in the mid-1990s measured the velocities of stars as a function of distance from the centres of two nearby galaxies: M84 and M87. What they found was that the stellar velocities showed a *sudden* turnover (i.e., from redshifted to blueshifted) within just a few parsecs of the central region. This demonstrates that whatever is governing the orbits of these stars must be incredibly compact (otherwise it wouldn't be such a sudden turnover). Indeed, based on these measurements, it was estimated that the central few parsecs of M87 contained a "dark" mass of 3.2×10^9 $M_\odot$ (i.e., it did not "shine" as much as you'd expect from a billion stars).

Later, in the early 2000s, aided with adaptive optics on ground-based near-infrared telescopes, astronomers measured the motions of stars around the central point of the Milky Way (known as Sagittarius A^* and pinpointed via long baseline radio interferometry). In doing so, they discovered that the stars close to Sagittarius A^* are orbiting a very massive, compact object (i.e., 4.3×10^6 $M_\odot$ that emits virtually no optical light. This object is so compact and massive that the only viable explanation is that it is a supermassive black hole at the centre of the Milky Way.

4.3.2 Reverberation mapping

The other main means of measuring the mass of a supermassive black hole involves using one of the main observable features of Type 1 AGNs as a test particle – the broad line region (BLR).

[1] At least, it's nice thing for this calculation, less so if you're the material falling in.

Under the assumption that the gas in the BLR moves under the influence of the black hole's gravity, we can use simple Newtonian dynamics to derive the black hole's mass:

$$M = \frac{v^2 r}{G} \tag{3}$$

where M is the mass of the black hole, G is the gravitational constant, v is the velocity of the clouds in the BLR and r is the radius of the BLR (i.e., equating $F = GMm/r^2$ and $F = mv^2/r$). We can easily measure v from the optical spectra of Type 1 AGNs, since the width of the emission lines gives a direct measurement of the circular velocity of the clouds via their doppler broadening.

By far the more difficult measurement to make, however, is the radius, r, of the BLR, since it is far too small to be resolved in even the most nearby AGNs. For this, astronomers exploit the variability of AGNs. When an AGN varies in brightness, this change propagates out from the central engine at the speed of light. After a certain amount of time (corresponding to the light travel time between the BLR and the central engine), we see the BLR respond to the change in luminosity of the nucleus (it is said to *reverberate* in luminosity in response to the change in luminosity of the central engine). By monitoring the luminosity of the central engine and BLR, and measuring how long it takes for the BLR to respond to a change in luminosity of the central engine, astronomers can determine the radius of the BLR. Typically it takes a BLR roughly a day or two to respond to a change in luminosity of the central engine, meaning that radius, r, of the BLR is roughly a couple of light-days. With a measurement of r, it is comparatively easy for astronomers to now calculate M.

Of course, since BLRs are only seen in Type 1 AGNs, reverberation mapping can only be used to measure the masses of black holes in Type 1 AGN (i.e., not non-AGN galaxies).

4.4 The Black Hole – Bulge relationship

Today, astronomers have measured the masses of hundreds of central supermassive black hole by one way or another. One remarkable outcome of this has been the realisation that the masses of the central supermassive black holes are remarkably well-correlated with the stellar mass of the bulge or spheroid in which they reside. This correlation spans over four orders of magnitude in both black hole mass and bulge mass.

The reason why this quite so remarkable is that, as we have seen, the gravitational influence of a supermassive black hole on the stars in its host bulge is extremely small. The black hole's gravity only dominates over that of the stars in the bulge for about 30pc, whereas a typical bulge will be many kiloparsecs across. Indeed, in terms of gravity, only about 1% of all the stars in a bulge "know" about the central supermassive black hole. A such, if it were based on purely gravity alone, we shouldn't expect there to be any relationship between the black hole and bulge, and yet we do see a tight correlation.

As we shall see later in the course, this tight correlation between the black hole mass and key properties of its host bulge has led many extragalactic astronomers to think there must be some interaction between black holes and their galaxies beyond gravity alone. This interaction is the "AGN feedback" that was referred-to in our discussions of the theory of galaxy evolution. We will cover this important aspect of galaxy evolution over the remaining few lectures.

5 Lecture 12 learning objectives

In this lecture we continued to look at the phenomena of AGNs, focussing mainly on the black hole and accretion disk that form the so-called "central engine". Here are the main learning objectives from this lecture:

- Have knowledge of the general properties of AGN: high nuclear luminosities; compact energy generation regions; long lifetimes; jets, peculiar spectra containing broad emission lines
- Understanding of why accretion of material by black holes is the most plausible energy generation mechanism for AGN
- Knowledge of the evidence for supermassive black holes in the nuclei of nearby galaxies (i.e., material on M84, M87, and the Milky Way) and AGNs.
- Awareness of the black hole – bulge relationship.

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:

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

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: $M_{\rm BH} \gtrsim 1.5 \times 10^9\ \mathrm{M}_\odot$. 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 $2 \times 10^9\ \mathrm{M}_\odot$. 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 $10^9\ \mathrm{M}_\odot$ 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:

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

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:

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

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

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

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

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

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

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

or,

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

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

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

where here Δt is in years.[1] For $\Delta t = 744$ million years, this gives $M_{\rm F}/M_{\rm S} \approx 1.2 \times 10^7$. Since, in the case of ULAS, $M_{\rm F} = 2 \times 10^9\ M_\odot$, this implies that $M_{\rm S} = 170\ {\rm M}_\odot$. 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.

Lecture 14:
The triggering of AGN

Dr. James Mullaney

March 26, 2017

1 Introduction

In the previous few lectures, we've seen that two ingredients are needed to produce an AGN: a supermassive black hole and supply of gas and dust to accrete onto it. In the last lecture, we considered current theories on where the supermassive black holes originate from. Now, we will consider the other major challenge in AGN astronomy: what causes the gas and dust to accrete onto a supermassive black hole to produce an AGN.

2 Dormant and active black holes

It has already been highlighted in previous lectures that it is thought that most, if not all, massive (i.e., $\gtrsim 10^9\ M_\odot$) galaxies contain a supermassive black hole at their centres (see the Solan Argument of Lecture 12). Further support of this is the clear evidence of the supermassive black hole at the centre of the Milky Way (known as Sagittarius A$*$), which clearly demonstrates that normal (i.e., non-AGNs) galaxies contain dormant supermassive black holes. These dormant black holes may "flare up" once in a while as they consume small amounts of gas and dust, but never normally enough to warrant being labelled as an AGN.[1]

In order to become an AGN, the supermassive black hole needs to accrete at a rate of at least a few percent of a solar mass per year. This material forms an accretion disk which is of the order 0.01 pc in size, but to reach these small scales the gas must lose $>> 99\%$ of its angular momentum, which is no mean feat. How this is achieved is known as the "AGN triggering problem", since the transport of gas to the nucleus is needed to "trigger" an AGN, and is the focus of much research in AGN astronomy.

3 Suggested AGN triggering mechanisms

Since their discovery, a number of different mechanisms have been suggested as possible means of transporting gas from galaxy-scales (i.e., $\sim$kpc) to accretion disk-scales (i.e., sub-pc) in order to trigger an AGN. The most popular of these are:

[1] I once attended a conference presentation in which it was said that Sagittarius A$*$'s flares correspond to it accreting roughly a mountain's worth of gas within a few hours. Pretty impressive, but far from the $\sim$solar mass per year needed to power a quasar.

- **Galaxy mergers and interactions**: Because of the disruption of a galaxy's internal dynamics caused by a merger (i.e., from rotational support to a much more chaotic system) or even an "interaction" (i.e., a close fly-by leading to tidal streams), they are an extremely effective means of removing angular momentum from internal gas. As such, mergers have long been suggested as a possible means of triggering an AGN.
- **Secular processes**: The opposite of galaxy mergers/interactions are called "secular processes", and refer to when a galaxy is just going about its usual business in isolation. Suggested secular accretion mechanisms include: spiral arms or bars channelling gas to the nuclear regions, the accretion of small satellite galaxies (i.e., those that form the tidal streams around the Milky Way and are too minor to be considered true mergers), winds or ejecta from processes associated with star-formation (including supernovae). Basically anything *internal* that could channel cold gas in the galaxy toward the nuclear regions.
- **Accretion of hot halo gas**: All galaxies sit within a dark matter halo. These halos contain large amounts of very diffuse gas (indeed, the majority of gas in the Universe is in this state) believed to be kept hot ($\sim 10^6$ K) by feedback mechanisms (i.e., energy injected from the galaxy, not least by AGN) and shock heating. It has been suggested that some of this gas can penetrate to the centre of a galaxy and by accreted in a "hot mode" to via Bondi accretion (in which the gas does not form an accretion disk; feel free to look it up, but we won't go into more details here). Alternatively, some of this hot gas may cool (to $\sim 10^4$ to 10^5 K) to form what are known as "cooling flows" which stream onto the galaxy (toward the centre of mass) from the halo.
- **Cold accretion from large-scale filaments**: Some galaxies (especially massive ellipticals) live at the nodes of large-scale (i.e., many tens of Mpc) filamentary gaseous structures. These structures are very effective at transporting gas toward the nodes where the galaxies sit, and it has been suggested that they may even penetrate the galaxies right to their nuclei. If that is the case, they may provide a direct channel to feed an AGN.

For the rest of the lecture, we'll consider if there is any evidence for these various potential mechanisms (we'll bunch the final two together, as they are essentially the same thing - accreting material from scales far larger than the galaxy).

4 Galaxy mergers and interactions

Perhaps the easiest triggering mechanism to test for is mergers or interactions. If we find a significantly higher proportion of AGNs in galaxies that are undergoing mergers compared to non-AGNs, then we can infer that the merger process is, indeed, an effective means of funnelling gas toward the galaxy nucleus to trigger an AGN. Note, however, the importance of a comparison sample in that statement: it is crucial that we compare like-for-like AGN and non-AGN galaxies when trying to identify AGN triggering mechanisms. This typically involves identifying a mass-matched sample of non-AGNs which, thankfully, is usually comparatively straightforward since non-AGNs outnumber AGNs by many tens-to-one.

To date, a number of studies have explored the question of whether AGNs preferentially reside in merging systems. Because of the difficulty in spanning very broad ranges of AGN luminosity in our samples (recall the lecture on extragalactic surveys, in which we saw that different depths and areas of surveys were used to identify different luminosity systems), these studies typically focus on a comparatively narrow luminosity range. Results from deep-field surveys in particular show no evidence of a higher fraction of mergers among moderate luminosity AGNs (i.e., $L_{\rm Bol} \lesssim 5 \times 10^{44}$ erg s^{-1}) compared to non-AGNs in the same fields. This suggests that secular processes of the type highlighted in the previous section are triggering these lower luminosity AGNs.

The triggering mechanism for more luminous quasars may, however, differ from more moderate, less luminous AGNs. This may well be because more violent processes are needed to channel the greater amounts of gas needed to trigger quasars than can be achieved with sedate, secular processes. Indeed, computer simulations predict that the peak of AGN activity (which we would observe as a quasar) take place during the final stages of a major galaxy merger. To test this, however, requires a sample of powerful quasars with sufficient quality observations to see signs of recent merger activity.

4.1 The 2Jy sample of radio galaxies

Among the most well-studied samples of nearby (i.e., $z < 0.7$) luminous AGNs is the 2Jy sample of southern radio galaxies (a radio galaxy is another name for a radio loud AGN). Although originally selected because of their high radio luminosities (they all have radio fluxes above 2 Jy, which makes them some of the brightest radio sources in the whole sky), it turns out that almost 80% of the AGNs in the 2Jy sample are also optically-luminous quasars (although in some cases, the nucleus is obscured from view; see the lecture 11 on AGN unification).

Using very deep (i.e., sensitive) optical imaging for the 2Jy sample, astronomers have found that around 15% show evidence of undergoing current major mergers. Furthermore, a further 70% of them show clear evidence of tidal features, which are a tell-tale sign of recent galaxy interactions. This is a much higher fraction than found in matched comparison samples of non-AGNs in the local Universe. This is consistent with the idea that powerful radio galaxies (and possibly most quasars) are triggered in galaxy interactions. However, contrary to what is suggested by simulations, it seems that the triggering isn't associated with a particular stage of a merger, simply that a merger has taken place in the recent past (within a few 100 Myr; i.e., a late-stage merger).

5 The role of star-formation

One of the main problems with trying to figure out what triggers AGN is that, invariably, multiple *potential* triggering mechanisms are present in a galaxy *at the same time*. In particular, major gas rich mergers also induce high levels of star formation within galaxies (due to the compression of cold gas clouds within the colliding galaxies). It can, therefore, be difficult to assess whether the AGN is, indeed, triggered by the merger, or whether it's really the star-formation that induces the AGN and the merger's role is simply to enhance the levels of star-formation. While this may be a moot point for mergers (after all, it's ultimately the merger which triggers AGN), but it's important for AGN triggering in general to know whether star-formation (which is common) is sufficient, or whether some kind of interaction (which is rare) is critical to trigger an AGN.

Trying to untangle the role of star-formation vs. merger is further complicated by the difficulties in measuring accurate rates of star formation in galaxies hosting powerful AGNs. This is because a powerful AGN can contribute to all of the wavebands traditionally used to measure star formation rates (SFRs). For example, a bright Type 1 AGN will dominate over any star formation at UV wavelengths, and even a Type 2 AGN can contribute to the UV bands via reflected light (reflected from clouds of gas within the galaxy). Similarly, while AGNs are typically quite weak at far-infrared wavelengths, even here they can dominate over low levels of star-formation (to add a further complication: the intrinsic SED of AGNs remains poorly constrained at infrared wavelengths).

Perhaps the most reliable means of measuring the SFRs and star-forming histories of *powerful* AGNs is via sensitive spectroscopy observations of the host galaxies. Stellar absorption features in the spectra can be modelled (via spectral synthesis) to give precise ages of the stellar populations in the host galaxy, and by placing the slit off-nucleus, the emission from the AGN itself can be mitigated. When this is done for the 2Jy sample of nearby bright quasars, young stellar populations are only detected in about 20-35% of the sample. This has been interpreted as evidence that, while interactions may play a key role in triggering AGNs, in most cases they are not triggered at the *peak* of major, gas rich mergers when most of the star-formation takes place. This has led some to speculate that there is a *delay* between the closest approach of a merger and the triggering of an AGN.

6 Triggering via cool gas accretion

The final suggested means of AGN triggering we will consider is via direct accretion of cold gas from intergalactic space (via cooling flows or channelled along large-scale filaments). In order to power a quasar with a bolometric luminosity of $L_{\mathrm{Bol}} > 10^{45}$ erg s^{-1}, a black hole must accrete at a rate of roughly 0.2 M$_\odot$ yr^{-1}. If a typical quasar lifetime is of the order 10^6 to 10^8 years (the former based on the size of the largest radio jets, the latter based on the fraction of massive galaxies hosting quasars within a given redshift range [i.e., within a given time interval]), then this means that the black hole will accrete roughly $2 \times 10^5 - 2 \times 10^7$ M$_\odot$ of gas during a typical quasar episode.

However, that only represents the gas that falls into the black hole, yet the black hole-to-bulge mass relationship tells us that for every one solar mass of gas/dust that falls into the black hole, there must be 500 M$_\odot$ that forms stars. As such, to fuel a quasar for about 10^6 to 10^8 years requires a *total* gas reservoir of $10^8 - 10^{10}$ M$_\odot$ (and that's assuming a 100% efficiency in converting gas into stars, which is far from the case in reality). So, the key question is: "Is there any evidence that such a large gas reservoir is even available to fuel a quasar?".

In astronomy, measuring the mass of gas contained within a given region is notoriously difficult. If the gas is ionised, we can use the strength of ionisation lines as a proxy-measure. However, the ionised phase only represents a small fraction of gas in a galaxy available to form stars or accrete onto a black hole. Instead, the dominant gas supply is either in the neutral or molecular phase, which doesn't emit at optical wavelengths. It is possible to use the Hydrogen 21 cm line in the radio bands, but this is weak and so only detectable in the most nearby galaxies. So, to measure the amount of neutral gas available, astronomers use the fact it is often accompanied by large amounts of dust, and so use the dust mass as a proxy for gas mass (a ratio of $M_{\mathrm{Gas}}/M_{\mathrm{Dust}} \sim 100$ is typically assumed). Since dust emits as a black body, if we know its temperature (which we can calculate using infrared colours), then we can calculate its mass from its (infrared) luminosity.

If quasars are being fuelled by cold gas within their host galaxies, then we should measure dust masses of around $10^6 - 10^8$ M$_\odot$ (i.e., around 1% of the required gas mass) in galaxies hosting powerful quasars. Using the *Herschel* infrared telescope, which was launched in 2009, astronomers have measured the infrared temperatures and luminosities (and consequently, masses) of the dust around powerful quasars, including the 2Jy sample. As predicted, they do indeed contain typical dust masses of around 10^7 M$_\odot$, confirming that there is sufficient gas to fuel their resident quasars for $\gtrsim 10^6$ years.

Finally, to give you some sense of how large 10^8 $M_\odot$ of gas is, the Large Magellanic Cloud (LMC) contains roughly this amount of gas. So it is feasible that, should the Milky Way eventually merge with the LMC (which would be classed as a minor merger), there would be sufficient gas supplied by the interaction to trigger Sagittarius A$*$ into becoming a quasar.

7 A summary of AGN triggering mechanisms

Over the course of this lecture, we have seen how various different mechanisms have been suggested as possible means of triggering an AGN. As you have probably already noted, there is no "single -fix" to this issue, with mergers/interactions, secular processes and cool accretion all possible mechanisms (we'll also consider evidence that star-formation may also be linked to AGN in Lecture 16). However, this probably shouldn't be too much of a surprise; all it takes to power an AGN is material falling onto a black hole. The black hole doesn't care how the material is funnelled onto it, so it's probably to be expected that different triggering processes can all play a role.

I feel the key thing to take away, however, is that there seems to be evidence that AGNs of different luminosities seem to be triggered by different processes. For more moderate luminosity AGNs, it seems that secular processes are sufficient (including, as we'll see in L16, non-merger-induced star-formation). By contrast, there is increasing evidence that the most luminous AGNs are, indeed, triggered (or at least, helped) by a galaxy merger/interaction. One thing I can tell you for *for certain* is that this is a highly active area of current research (so nothing is really certain!), so our understanding of AGN triggering may change considerably over the coming years.

8 Learning objectives for Lecture 14

In this lecture we've considered the possible mechanisms of driving gas and dust from the outskirts of a galaxy toward its nuclear regions in order to trigger an AGN. This is a highly active area of research and, consequently, remained surrounded in uncertainties. Having said that, there are some key things you should take from this lecture(!):

- It appears that moderate luminosity AGNs are triggered by "secular" processes.
- Local radio galaxies (i.e., the 2Jy sample) are diverse in terms of their detailed morphologies, star formation properties, and cool ISM contents.
- A small but significant minority ($\sim$15%) are triggered in major, gas-rich mergers in which both the super-massive black holes and stellar masses of the host galaxies are growing rapidly.
- But the majority of local radio galaxies represent much later stages of galaxy interaction, possibly indicating a late-time re-triggering of AGN activity via galaxy interactions and/or minor mergers (2xLMC gas mass)

Lecture 15:
AGN feedback and outflows

Dr. James Mullaney

April 24, 2017

1 Introduction

Over the past few lectures we've looked extensively at supermassive black holes (SMBH) and the AGN that are produced when these black holes accrete matter. However, while you may have (hopefully) found AGNs interesting in their own right, you may be asking yourself what they've got to do with galaxy evolution. In this lecture, we will be addressing that very question.

2 Indirect evidence of links between AGN and galaxy evolution

As you should probably be aware by now, no AGN is in complete isolation: as the name suggests, they're all found in the centres of galaxies. Furthermore, we saw in the previous lecture that the trigerring of AGNs is intrinsically linked to their host galaxies since all accreted material ultimately comes from the host. Indeed, we saw that it's likely that some AGNs are triggered by galaxy interactions. As such, there is clearly at least some level of co-evolution between AGNs and their host galaxies.

Until the mid-1990's, this "co-existence" picture of AGNs and galaxies formed pretty much our whole understanding of the relationship between AGNs and their host galaxies. However, toward the end of that decade and into the mid 2000's, a number of major discoveries were made in AGN astronomy that suggest that the relationship between AGNs and their host galaxies is much closer than previously thought.

2.1 Evidence from observations

The first of these discoveries was that the masses of supermassive black holes are tightly correlated with the masses of their host galaxy bulges. As we saw in Lecture 12 (§4.4), this tight correlation is remarkable because the gravitational influence of the black hole only dominates within a few tens of parsecs, meaning that – gravitationally at least – the vast majority of stars within the galactic bulge "knows" nothing of the black hole in the centre.[1] The tight correlation between black hole mass and bulge mass has been interpreted as evidence of a level of interaction between the two *beyond* that expected from gravity alone.

The second discovery that was interpreted as signalling a strong evolutionary link between SMBHs/AGNs and galaxies is that the star formation rate density and the black hole accretion rate

[1] It should go without saying that, in terms of gravity, the black hole known nothing of the bulge. Gauss's theorem tells us that.

density of the Universe have evolved "in-sync" over the entire observable history of the Universe. As such, by one way or another, the total amount of material accreting onto a SMBH (during an AGN episode) at any epoch is closely tied to the total amount of material that forms stars. This is seen as clear evidence that black holes and galaxies have co-evolved over the past 13 billion years.

2.2 Evidence from simulations

The final major discovery linking AGNs to galaxy evolution that we will consider came not from observations, but from computer simulations (or, more correctly, from comparing the predictions made by computer simulations against observations). One of the most important – and reliable – outcomes of cosmological-scale, dark-matter dominated semi-analytic models is the mass function of dark-matter halos.[2] It is "reliable" because dark matter only interacts with itself and other matter via gravity, which we understand *very* well on cosmological scales. What semi-analytic models predict is that the dark-matter halo mass function can be described very well as a simple power law (i.e., a straight line on a log-log plot). Of course, we can't actually measure the dark matter mass function to compare against this prediction, so instead we to use analytic prescriptions to populate the dark matter halos with galaxies which we can then compare to the (measured) *galaxy* mass function.

The prescriptions used in the first truly cosmological-scale semi-analytic models (in the mid-2000s) predicted that the galaxy mass function was very different to that observed in the real Universe. In particular, the first simulated mass functions hugely overpedicted the numbers of very low mass and very high mass galaxies in today's Universe (by many orders of magnitude). In fact, the earliest simulated galaxy mass functions were also well-described by a single power-law, whereas as we have seen throughout the course, the observed mass function is better described by a broken power law (or, more specifically, a Schechter function). It became clear that what was needed was a way to prevent too many very low and very high mass galaxies from forming within the dark matter halos.

The way to prevent a galaxy from forming or getting bigger is to prevent gas from cooling and collapsing to form stars. This is done by injecting energy into the gas to either heat it up or drive it out of the gravitational potential well that is pulling it together. At the time of the first simulations, it was already known that a major source of energy input is from stars themselves, in the form of stellar winds, supernovae and simply radiation from the photosphere. However, when this "feedback" from stars was included in the semi-analytic models, it was found they only solve one part of the problem: they only prevent the formation of too many small galaxies. In order to prevent too many *massive* galaxies from forming, much more energy needs to be injected. It was found that by including (reasonable levels of) energy from AGNs in the semi-analytic models prevented gas from collapsing and forming too many massive galaxies, and closely reproducing the observed galaxy mass function. As such, the comparison between the galaxy mass function predicted by semi-analytic models and that observed in the real Universe provides indirect evidence of "feedback" processes (both stellar and AGN in origin) regulating galaxy growth.

Further indirect evidence of "AGN feedback" came from smaller-scale simulations of merging galaxies. Around the same time that cosmological-scale semi-analytic models were predicting the need for AGN feedback (i.e., mid-2000s), simulations of galaxy interactions were also providing evidence of a link between AGNs and their host. In these smaller-scale simulations, it was found

[2] In the same way that there is a galaxy mass function – simply the histogram showing the numbers of galaxies in a given mass bin per unit volume – there is a dark matter halo mass function.

that, unless energy from an AGN was used to "sweep-out" the gas from a galaxy post-major merger, then the resulting galaxies were far too compact (compared to real galaxies) and did not reproduce the observed black-hole to bulge mass correlation.

3 How might AGNs affect galaxy evolution

The indirect evidence of a close link between AGNs and galaxies highlighted above caused a significant shift in our consideration of AGNs. In the space of a decade, AGNs went from being studied as a mere "curiosity" to being regarded as having a major impact on how galaxies have evolved.

Despite their perceived importance, it remains unclear precisely *how* AGNs affect galaxy evolution in the *real* Universe. Although we know that AGNs produce a lot of energy, it's not clear how (or even whether) this energy is effectively transferred into their host galaxies to affect a change. However, simulations do predict two key mechanisms:

- Heating gas on large (i.e., dark matter halo) scales which prevents it from cooling and collapsing toward the centre of the halo (where the galaxy lives) and forming stars;
- Heating-up or driving-out the cool gas already within the galaxy which *would* have formed stars.

Of course, it is also possible that once gas driven out by the second mechanism reaches large enough scales it is prevented from re-cooling by the first mechanism. In the following sections, we'll look in more detail at the observational evidence supporting these two mechanisms.

4 Heating inter-galactic gas

To heat inter-galactic gas (i.e., the gas *between* galaxies), an AGN would be required to have an influence on the environment well outside its host galaxy. In other words, on scales of tens to thousands of kiloparsecs (remember, a galaxy is a few tens of kpc across, while the distances between galaxies are typically measured in megaparsecs). These are the kinds of scales that are reached by the jets of radio-powerful AGNs, so these provide an obvious potential means for an AGN to influence the inter-galactic material.

With clear observational evidence of kpc to Mpc scale radio jets being produced by AGNs, the next key question is: do these jets actually heat the inter-galactic material and prevent it from cooling onto the host galaxy where it can form stars? One way this could be achieved is via *shock heating*. Shocks occur when discontinuities in gas pressure and density move though a medium (in this case the inter-galactic gas) at speeds faster than the local sound speed. When this happens, the atoms of gas are accelerated at the interface of the discontinuity, which heats up the gas.

The plasma from which AGN jets are formed are travelling at close to the speed of light, so much, much faster than the local sounds speed of the inter-galactic material (which is typically around 300 km s^{-1}. As such, these jets generate shockwaves that expand outward from the jet as it advances through the intergalactic material. Importantly, the shock waves expand spherically around the progressing jet, so are not just confined to the small opening angle of the jet and, instead, will propagate throughout the surrounding intergalactic material. These advancing shocks are predicted to heat the intergalactic material to $\sim 10^6$ K, which is easily high enough to prevent it from cooling onto the galaxy.

Do we actually see any evidence of AGN-launched jets actually transmitting shockfronts into the intergalactic gas? At temperatures of 10^6 K, gas emits X-rays as an extremely hot black body. By observing inter-cluster gas around radio AGNs, astronomers have indeed found evidence of extremely hot, shocked gas forming "bow-shocks" around the the collimated radio jets. Crucially (and as expected) these shocks have a much larger opening angle compared to the highly collimated jets, so display clear evidence of radio jets transmitting large amounts of energy into large volumes of the inter-cluster material. This energy is, indeed, heating the gas up to millions of Kelvin, preventing it from cooling to form stars.

Another way that AGN-launched radio jets can affect the inter-cluster gas is by also physically moving the gas away from the galaxy. In this scenario, the jets excavate cavities of gas which is similar, in principle, to blowing up a bubble. The energy contained within these bubbles is sufficient to prevent the cooler gas outside the bubble from cooling and falling onto the host galaxy. Such bubbles of hot gas are, indeed, seen in X-rays behind the shock waves caused by the propagating jet.

5 Affects on the inter-stellar material

As well as preventing gas *outside* the galaxy from cooling and collapsing to form stars, it is thought that AGNs may also affect the gas *within* the host galaxy; the so-called inter-stellar material. After all, a typical AGN produces many times more energy than is required to disrupt (i.e., ionise, heat, eject) the *entire* gas content of a galaxy. In order to do so, however, a significant fraction of this energy must be captured by the interstellar gas. It remains to be determined whether the coupling between the energy released by AGNs and the interstellar gas is sufficient to prevent the latter from forming stars.

There are two main mechanisms by which astronomers think that AGNs are affecting the interstellar gas in their host galaxies: one is via radio jets (again) and the other is via non-relativistic "winds". Both, ultimately, lead to the same result: the evacuation of interstellar gas from the host galaxy.

5.1 Jet-ISM interaction

As well as the huge, Mpc-scale radio jets displayed by some AGN (in fact, a minority of about 10%), many more show evidence of compact radio emission. On closer inspection, these compact radio AGNs do display evidence of radio jets, but they are typically extended on sub-kpc (rather than Mpc) scales. As such, these jets are extended on galaxy-scales.

What is particularly interesting about these galaxy-scale jets in terms of AGN feedback is that optical spectroscopy has revealed that they are often associated with high-velocity gas in the interstellar material.[3] It is not unusual for this gas to have speeds of 1000 km s^{-1} or more, but they are *not* relativistic. In other words, this is *not* the plasma that is forming the jets, but is instead interpreted as interstellar material that is being accelerated to many hundreds of km s^{-1} *by* the jet. Based on reasonable (but highly uncertain) assumptions about its density, this jet-accelerated interstellar material is thought to be outflowing at a rate of many hundreds (and in some cases up to thousands) of solar masses per year. If this is maintained over a the expected lifetime of an AGN of a few tens of millions of years, this could, in principle result in around 10^{10} $M_\odot$ of gas

[3]Recall that we can extract the kinematics of the gas from the profiles of its emission lines.

being evacuated from the host galaxy – enough to have a significant impact on the future growth of the galaxy.

5.2 AGN "winds"

So far, we've seen that radio powerful AGNs are prime candidates for "AGN feedback". However, only a minority of AGNs are radio loud, so what about the dominant population of radio-quiet AGNs? Are they thought to induce "AGN feedback"? If so, has this been confirmed by observations?

In previous lectures, we've seen how an AGN (radio quiet or otherwise) can produce the same amount of light as an entire galaxy in a region of space only a few times bigger than our solar system. Not only that, but a lot of this light is emitted in the UV. This concentration of energy production is truly staggering, and leads to some intense physical processes. One of the most interesting of these in terms of AGN feedback are "AGN winds".

It is thought that one of the consequences of the intense heat and radiation of the accretion disk is that it "boils off" large quantities of gas (possibly as much as half of the mass of the accretion disk may evaporate in this way). This is not dissimilar to stellar winds being driven-off massive stars. As soon as the gas rises out of the accretion disk, it is subject to the full intensity of the radiation from the disk and is accelerated from the accretion disk via photon pressure. This gas can reach speeds of tens of thousands of km s^{-1} very close to the central engine, but as they travel outwards they lose velocity due to gravity (from the stars in the galaxy) and because they crash into the ambient interstellar gas. However, despite their loss of speed, it is still predicted that they are sufficiently powerful to eject a substantial amount of the interstellar gas from the host galaxy.

But do we *see* any evidence of such AGN winds among radio-quiet AGNs? The answer is an emphatic Yes. Indeed, roughly a fifth of all quasars show clear evidence of broad, blushifted absorption lines in their rest-frame UV spectra. It is not unusual for these absorption lines to be shifted by many tens of thousands of km s^{-1}, meaning that the absorbing gas is moving *toward* us (relative to the host galaxy) at an incredibly high speed. This has been interpreted as clear evidence of gas *outflowing* along our line of sight toward the central engine. Due to the levels of variability displayed by these lines, it is thought these they trace outflows with roughly 10 pc of the central engine.

At even smaller scales (and higher energies), there is evidence of even faster winds in X-ray observations. These show absorption features that are blueshifted by many tens of thousands of km s^{-1} (the fastest are approaching relativistic speeds). Again, these are interpreted as providing evidence of extremely fast moving outflows on scales of less than 1 pc from the central engine. Further, the amount of mass contained within these "ultra fast outflows" is an appreciable fraction of the Eddington luminosity of the SMBH.

So, there is clear evidence that radio-quiet AGNs do, indeed, drive powerful outflows. However, until fairly recently, these discoveries were limited to small scales, whereas for an AGN to influence galaxy evolution these outflows must interact with the interstellar material on kpc-scales. Recently, however, integral field observations of AGNs both in the local Universe and in the high-redshift Universe show evidence of strongly blueshifted *emission* lines on kpc scales. Again, this has been interpreted as evidence of AGN-driven outflows, but with velocities somewhat lower than those observed in UV and X-ray absorption lines. Despite this, they still show evidence of outflowing gas travelling at hundreds or, in some cases, thousands of km s^{-1} on *galactic* scales. It is thought that these outflows are in the processes of sweeping-out the interstellar material within the host galaxy, and thereby quenching future star-formation. However, finding galaxy-scale outflows is one thing, but actually *proving* that *quenching* is actually taking place is proving to be far more of a challenge and is currently one of the hottest topics in extragalactic astronomy research.

6 Learning objectives from Lecture 15

In this lecture, we've considered how, rather than a simple astronomical curiosity, AGNs are now thought to have had a major influence of galaxy evolution. The key learning objectives are:

- Have an understanding of the indirect evidence of AGN feedback, including that derived from simulations.
- Have knowledge of the ways in which AGN might affect the evolution of the host galaxies;
- Have an understanding of the differences between jet mode and quasar mode feedback
- Have Knowledge of the direct observational evidence for radio and quasar mode feedback, and the impacts on the host galaxies.

Lecture 16:
The link between star formation and AGN activity

Dr. James Mullaney

April 25, 2017

1 Introduction

In the previous lecture, we saw how over the past two decades our perception of AGNs moved from them being regarded as an "astronomical curiosity" to playing a key role in regulating the growth of galaxies. The main way in which this regulation is achieved is through AGNs affecting star-formation in their host galaxies. In this lecture, we'll take a closer look at the connection between star formation and AGNs.

2 Star formation in AGN hosts

As we have seen previously, AGNs are the result of interstellar material (i.e., gas and dust) falling into the supermassive black holes that reside at the centers of galaxies. As well as producing large amounts of energy, the other direct effect of this accretion is that the black holes get more massive. As such, the mass of a supermassive black hole is the sum of all its accretion events to date.

We have also already seen that the mass of a supermassive black hole is tightly correlated with the (stellar) mass of its host bulge. However, we still don't fully understand what has caused this relationship (aside from the somewhat ambiguous catch-all term of "AGN feedback"). To try to address this, astronomers have spent a lot of time considering the star-forming properties of AGN hosts. The justification being that since galaxy bulges are formed from stars, which are produced in episodes of star formation, and black hole mass is built-up during episodes of AGN activity, then there should be some kind of link between star-formation and AGN. By studying the star-forming properties of AGN hosts, we are measuring the concurrent build-up of black hole and stellar mass.

2.1 Measuring the build-up of black hole mass

As we saw in Lecture 12, it's fairly straightforward to measure the growth rates of black holes during an accretion phase (i.e., AGN). This is because the energy radiated by the accretion process is directly proportional to the accretion rate of the black hole:

$$L_{\rm AGN} = \eta \dot{M}_{\rm BH} c^2 \tag{1}$$

so all we need to do is measure the luminosity of an AGN which can then be converted into a black hole growth rate. This is, of course, complicated by the fact we can't measure the total *bolometric* luminosity of an AGN; instead, we usually measure it in one or two bands (e.g., X-rays, optical, UV). To overcome this problem, however, astronomers measure the luminosity measured in one part of the electromagnetic spectrum and multiply it by a *bolometric conversion factors* to give an approximate bolometric luminosity.

2.2 Measuring the build-up of stellar mass in AGNs

As we saw in Lecture 4, there are a number of different ways to measure the star formation rates of galaxies. All these techniques rely on measuring the numbers of hot, young stars in a galaxy (i.e., UV continuum, Hα, infrared emission). However, in the case of measuring the star-forming properties of AGNs, it's difficult to use either the UV or Hα since AGNs contribute significantly to this type of emission. Instead, the far-infrared is widely used to measure the star-forming properties of AGNs, since AGNs are not thought to emit strongly in this part of the spectrum.

3 Non-AGNs: Main Sequence and Starbursting galaxies

Prior to considering the star-forming properties of AGN host galaxies, we should first consider what our baseline is. In other words, we need to know what the star-forming properties of non-AGNs are before we can assess whether AGNs show any systematic differences in terms of their star-forming properties.

It turns out that, at a given stellar mass, the galaxy population is bi-modal in terms of its star-forming properties: there are so-called "star-forming" galaxies and "quiescent" galaxies. These form two very distinct populations in terms of their optical colours, with star-forming galaxies being blue, and quiescent galaxies being red (due to the dominating population of old stars, not due to dust). Curiously, there are relatively few galaxies with intermediate green colours. As such, when we plot galaxy colours as a histogram, the population forms two peaks – one blue peak, and one red peak – with a "green valley" in between (yes, it is actually known as the green valley).

An important feature of the star-forming galaxy population is that the *rate* at which they form stars is tightly correlated with the mass of the host galaxy (at least for $> 95\%$ of the star-forming population). This correlation has become known as the galaxy "Main Sequence" (MS). The remaining $\sim 5\%$ of *star-forming* galaxies have star-formation rates that are *above* that of MS galaxies and are consequently known as "Starburst" (SB) galaxies. Starbursts typically have SFRs three or more times higher than MS galaxies of the same mass.

When astronomers measured the star-forming properties of MS galaxies out to higher and higher redshifts, they found that the SFR of a galaxy of a given stellar mass (which can also be express as the SFR per unit stellar mass, or specific SFR [sSFR]) *increases* with redshift. So, a typical MS galaxy at redshift 2 has a sSFR about 10 times higher than a typical Main Sequence galaxy today. At first, it was suspected that this rising SFR of MS galaxies was due to an increase in the occurance of gas-rich major mergers at earlier times, since mergers are a key means of enhancing star-formation in today's galaxies.

On more detailed morphological inspection, however, it turned out that MS galaxies at high redshifts were *not* dominated by merging systems. Instead, SBs tends to be associated with mergers at all probed redshifts, whereas MS are typically undergoing "secular" (i.e., isolated) evolution. Since MS galaxies *dominate* the numbers of star-forming galaxies at all redshifts, it therefore seems that the dominant mode of star-formation in the Universe is *not* triggered by major galaxies mergers.

If not major mergers, what *is* causing the rapid rise in the sSFRs of MS galaxies? It transpires that MS galaxies at high redshifts have significantly higher gas contents compared to their low redshift counterparts. Because of this, it is thought that the reason they are forming stars so rapidly at high redshifts is simply due to a far more abundant supply of gas in the early Universe compared to today.

3.1 Key points to remember about MS and SBs

Main-sequence and starbursting galaxies are a key feature of our current understanding of galaxy evolution, so I wanted to provide a quick summary of their properties:

- At a given redshift, galaxies on the Main Sequence (MS) have star-formation rates (SFRs) that are proportional to their stellar mass.
- So, at a given redshift, their specific SFRs (sSFR = SFR/stellar mass) is constant (but with some scatter).
- Galaxies with sSFRs $\sim 3\times$ above the MS are known as Starbursts (SBs).
- Star-formation in MS galaxies is thought to be triggered in isolation by "secular" processes, whereas in SBs it is triggered by major mergers.
- But, SBs are comparatively rare, so MS galaxies dominate the star-formation budget.
- The average (or typical) sSFR of MS galaxies increases strongly with redshift.
- This redshift evolution is thought to be due to the greater availability of cold gas in the early Universe from which to form stars.

4 AGN and the Main Sequence

Now that we have characterised the star-forming properties of normal (i.e., non-AGN) galaxies, we can consider where AGNs fit within this picture. Do AGNs predominantly live in starburst galaxies, which would suggest they are also triggered by major mergers, or quiescent galaxies (which may suggest they are "switching-off" star-formation via AGN feedback). Actually, it turns out that most AGNs reside in Main Sequence galaxies, suggesting that most AGNs are also triggered via so-called "secular processes".

You may feel that the finding that AGN preferentially reside in MS galaxies is somewhat contradictory to what we saw in Lecture 14, in which AGNs were linked to merger events. It should be noted, however, that the vast majority of AGNs in the Universe have relatively modest luminosities, whereas the merger-triggered AGNs we considered in Lecture 14 are among the most luminous AGNs in the local Universe. As such, this result reinforces thet idea that the dominant population of moderate luminosity AGNs are triggered by secular processes, but the most luminous AGNs are triggered by major mergers.

Since AGNs seem to prefer star-forming galaxies, then it makes sense to ask: "Is the luminosity of an AGN (i.e., its BH growth rate) in any way related to the star formation rate of its host galaxy?". In other words, is there a correlation between galaxy BH growth rates and star-formation rate? To investigate this, AGN astronomers have measured the average SFRs of AGN host galaxies binned in terms of the luminosity of the AGN. However, this experiment revealed little or no correlation between a galaxy's SFR and the current luminosity of its AGN. Interestingly, when we instead calculate the average AGN luminosity of galaxies binned in terms of their SFR (i.e, averaging the other way round), however, then a strong correlated between SFR and AGN luminosity *is* uncovered. In other words, when we average one way, we find no correlation, but when we average the other way, we do reveal a correlation. What's going on??

It is thought that the answer may lie in a key property of AGNs: that they vary (stochastically) on timescales that are much shorter than typical episodes of star formation. What this means is that an AGN will vary in luminosity by many orders of magnitude whilst the SFR of its host galaxy stays relatively constant. By grouping galaxies in terms of their AGN luminosity, we're selecting galaxies based on a highly stochastoc process. The effect of this is that it "dilutes" any underlying connection between the AGN and the host galaxy. By contrast, selecting galaxies based on the far more stable property of star-formation then *averaging over* the stochastic AGN variability, uncovers the true underlying links between AGNs and star-formation.

5 The probability of AGNs

As with any stochastic process, it is becoming increasingly common to think in terms of what processes affect the *probability* a galaxy hosting an AGN. For example, are AGNs more *likely* to reside in star-forming galaxies or – even better – how does the SFR of a galaxy affect the likelihood of it hosting an AGN of a given luminosity?

By thinking in such terms, AGN astronomers have begun to identify some important features of AGNs. Notably, it seems that the mass of a galaxy has no effect on whether it hosts an AGN of a given accretion rate. By contrast, recent studies have found that the likelihood of *rapid* black hole growth is enhanced in galaxies with high SFRs. As such, it seems that there may well be an underlying correlation between black hole growth and star-formation, but that uncovering this link in the face of AGN variability is going to take a lot more effort.

6 Learning objectives from Lecture 16

We've covered some quite conceptuallly-advanced ideas in this lecture, so there isn't as much reading to do as usual. I wanted to make sure there was enough time during the lecture to really explain some of the key concepts. So, don't worry too much if some of the ideas in these notes are tricky to grasp, there'll be plenty of opportunity for further explanations in the lecture. The key objectives you should take from this lecture are:

- Understand the evidence to support a connection between AGN and star-formation, i.e,;
  - BH-bulge relationship.
- Understand how we measure the AGN-SF connection, i.e.,
  - $L_{\rm AGN}$ (X-rays, etc), SFR (Optical, UV, and particularly IR etc.)
- Know what the star-forming Main Sequence is:
  - sSFR rises with redshift due to the increased availability of gas in the early Universe.
- Understand the importance of AGN variability in hampering our ability to connect AGN luminosities to other galaxy properties:

  - Washes out the underlying connections

- Be aware of our current understanding of the AGN-SF connection:
  - it seems rapid BH growth is more prevalent in star-forming galaxies.