GALAXIES, COSMOLOGY & THE HIGH ENERGY UNIVERSE

PHY2003 Part 3. Full Notes.

Dr. Peter Keys

QUB | School of Mathematics and Physics

Lecture 1 – The Interstellar Medium

Learning Objectives

The material that we will cover in this lecture includes:

1. What the Interstellar medium consists of
2. H I and H II regions
3. Molecular clouds
4. Interstellar dust and its effect on observations

The Interstellar Medium - Gas

In our galaxy the Milky Way, 90% of the observed mass is in the stars, 90% of the observed volume is in the Interstellar Medium (ISM). You can see this as in the vicinity of the Sun there is about a star in every 10 pc^3 of volume. The diameter of a solar-type star is about 10^{-7} pc so most of the ISM is empty of stars and is filled with gas and dust.

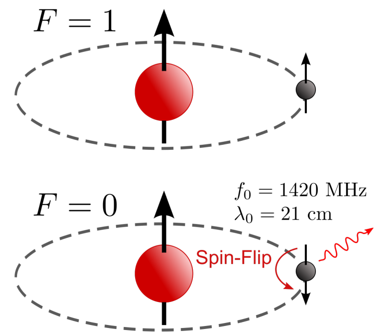
You can see this by looking at the arms of the Milky Way in images at optical wavelengths. You can see dark patches from absorption, and you can see radiation from gas that has been ionised, neutral atoms and molecules in the form of emission lines and continuum emission.

(a) Neutral hydrogen (H I)

The most abundant element in the Universe is hydrogen. Between the stars, any hydrogen gas will be very cold and therefore in the ground state. We call this neutral hydrogen (or H I).

In the cold neutral hydrogen atom, we can get a transition. This is a result of the fact that the electron and proton spins may be either aligned or anti-parallel (which is referred to as hyperfine splitting). Both the proton and electron in the neutral hydrogen have intrinsic angular momentum, called spin. This spin can have two orientations: 'up' or 'down'. The relative orientations of the spins affect the magnetic force between the proton and electron. When the spins are parallel there is slightly more energy.

The transition between these two states releases a low-energy photon at $\nu = 1.420\ \mathrm{GHz}$ and $\lambda = 21\mathrm{cm}$. Although this transition is rare for a single atom (once every few 10^6 years), enough H-atoms exist that this radiation appears very strong. Due to the wavelength, this is often referred to as the 21 cm line and is one of the most important diagnostic tools for the ISM and galactic structure.



Hyperfine splitting in the H-atom

(b) Ionised H II regions

These are found near luminous stars (O-type and B-type). The UV photons ionize neutral Hydrogen atoms into H II, producing a hot gas with a kinetic temperature T ~ 10, 000K. This happens in a roughly spherical region around the ionizing source known as a Stromgren Sphere.
The boundary of the sphere between the HII and H I regions is fairly thin (i.e., it is a sharp transition) and is determined by the density of hydrogen and the ionisation cross section of neutral hydrogen. The size of the Stromgren sphere is dependent on the rate of release of ionising photons and the density of the gas. If the density of the gas is high, photons won't make it far before coming across the quota of atoms that can be ionised. Therefore, the spectral type of the star is important as this determines the rate of emission of ionising photons. Some typical values are in the table below.

The Stromgren Sphere.

Spectral Type	Photons s^{-1} ($\times 10^{48}$)
O5	51
O6	17.4
O7	7.2
O8	3.9
O9	2.1
B0	0.43
B1	0.0033

The table shows that H II regions around the star are can be very different depending on the spectral type. In fact, where there are small groupings of O and B type stars we can have overlapping H II regions. There are two conditions under which the boundary of the H II region exists:

1. Bound by the ionising photons (*Ionisation bounded*)
2. The cloud ends prior to the absorption of all ionising photons (*Density bounded*)

The second case is an important source of ionising radiation in the general ISM outside of H II regions.

Within the H II region, the protons and electrons can recombine to produce excited H I. This gives off recombination radiation within H II regions.

$$\mathrm{H\ II} + e^{-} \rightarrow \mathrm{H\ I}^{*}$$

The atom quickly cascades down to the ground state through the Balmer series, with photons emitted in the jump between energy levels as it returns to the ground state. At higher energy levels, the levels are close enough for emission in radio part of the spectrum, while in lower energy levels the radiation goes from infrared to visible. The excited atoms can make it all the way to the ground state before being re-ionised by the UV photons from the star again.

Spectra of visible gas clouds show strong emission lines due to Hydrogen, particularly Hα at $\lambda = 6563\text{Å}$. This corresponds to transitions between the n = 3 and n = 2 energy levels in the Bohr model. Due to this, these regions tend to have a reddish glow. This allows us to distinguish between a H II region and reflection nebulae.

Emission lines due to other elements allow the relative abundances in the gas to be deduced.

H II regions can expand over time. With that expansion the surrounding cloud of neutral hydrogen can be compressed, and this compression may lead to star formation.

(c) Molecular clouds

Remember that hydrogen is the most abundant element in the universe. So, we define the ISM regions based on how this hydrogen forms.

- H I region: Neutral hydrogen dominates
- H II region: Ionised hydrogen dominates
- Molecular cloud: H_2 molecule dominates

Radio surveys show the existence of interstellar clouds containing many detected molecular species. The discovery of CH, CH^+ and CN raised the possibility that molecules may be an important constituent of ISM, though the low densities would imply that chemical reactions are unlikely to occur. This is essentially shown through the presence of CH, CH^+ and CN in molecular clouds, as these molecules are unstable under normal conditions.

Abundances of these molecules was found to be higher than was expected by theory. Molecular clouds contain $\sim 10 - 10^5 M_\odot$.

In the late 1960's, CO was found in molecular clouds with $\lambda = 2.6\text{mm}$. This line comes close to the 21cm line in terms of importance for studying the ISM. As it is in the millimetre wavelength band, the angular resolution will be good even with modest sized telescopes such as ALMA.

When observing a molecular cloud, we are not observing transitions in which an electron transitions between one energy level to another. This is unlikely in the cool ISM. Rather these are due to vibrational transitions in the molecules, which can occur at lower energies. For these molecules, the energy levels are given by the rotational inertia and generally occur at mm wavelengths. They can tell us quite a bit about the physical conditions in the molecular cloud. For example, the energy of the 2.6mm photon implies and equivalent temperature of 5.5K.

H II regions are generally associated with these clouds - they are the site of most star formation in our galaxy today. As the H II region expands into a molecular cloud, it will heat it up through compression. If the temperature of the region can cool enough prior to the gas in the molecular cloud expanding due to the increase in temperature and pressure, then star formation can take place. Molecular clouds containing molecules such as CO are crucial in this process, as they are more efficient in cooling the gas to allow fusion to start and stars to form.

(d) Planetary Nebula and Supernova Remnants

PN's and SNR's are easily visible in optical telescopes. These ejected shells of gas expand continuously outwards until they merge with the general ISM. PN's and SNR's will inject slightly different elements into the ISM based on the star size and evolutionary processes involved.

The ISM – Dust

Dust in the ISM appears to exist in tandem with the gas. About 1% of mass of the ISM is dust. It is mainly in the form of silicates and forms of carbon. Dust is found as solid particles of size 0.02 → 1.0μm with an average particle density of ρ_d = 1 per $10^6 m^3$. It may be observed in several different ways.

(a) Absorption

For distant stars, dust absorbs or scatters an increasing amount of light coming from them.When dust is spread uniformly, light loses an equal fraction of its power for every parsec that it travels through the dusty gas. The intensity of light observed is given by

$$I_{obs} = I_{true} e^{-\tau}$$

Where τ is a quantity referred to as the optical depth (this is a concept that is discussed more fully in PHY3003). The coefficient, K, is related to optical depth as $\mathrm{K} \equiv e^{-\tau}$.

So, this can basically be expressed as

$$\frac{I_{obs}}{I_{true}} = \mathrm{K}$$

We can quantify the interstellar extinction as the number of magnitudes that a cloud will dim starlight passing through it. Therefore, if a star has an apparent magnitude, m, without any interstellar extinction, but light form this star passes through a cloud with A magnitudes of extinction, then the star will be observed to have apparent magnitude, m', such that, $m' = m + A$. More specifically,

$$-2.5 \, log \, I_\lambda \, (\text{obs}) = -2.5 \, log \, I_\lambda \, (\text{true}) - 2.5 \, log \, \mathrm{K}$$

$$M_\lambda(\text{obs}) = M_\lambda(\text{true}) + A_\lambda$$

Where A_λ is termed the interstellar absorption and is wavelength dependent. In visual wavelengths in the galactic disk $A_V \sim 1$ mag/kpc.

If we know a star's spectral type and distance, we can work out the extinction between our line of sight and the star. The spectral type of the star gives us the absolute magnitude, M_λ, and we can measure the apparent magnitude, m_λ. In the presence of A_λ magnitudes of extinction, the star will be A_λ magnitudes fainter. Therefore, the equation for the distance modulus becomes

$$(m - M)_\lambda = 5 \, log \, d - 5 + A_\lambda$$

If you take an image of a region where you think you have some interstellar extinction, you should observe fewer stars as the light will be dimmed by the extinction due to dust and some will be too dim to detect. To work out the extinction in a line of sight, you basically need to do some star counting by looking at the number of stars per unit area in two nearby patches of sky, one where you think there is extinction and the other where you think there isn't any. They should have similar distributions, with any differences coming from the extinction along the line of sight.

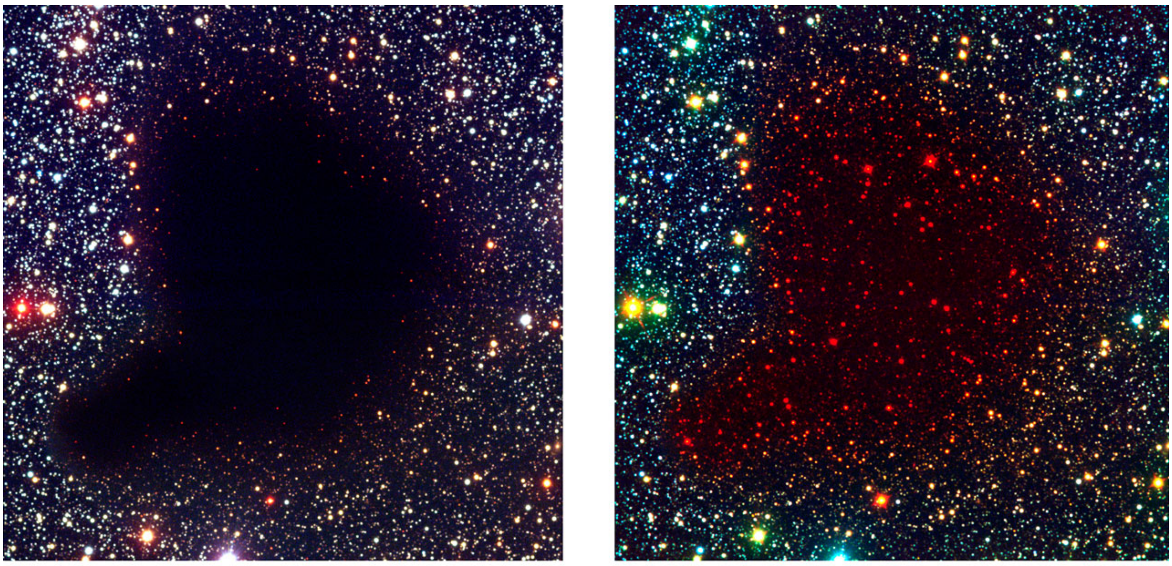
Interstellar extinction is not the same at all wavelengths. In general, shorter wavelengths will have a higher level of extinction and so blue light will be more efficiently blocked. As such, stars will appear redder in colour. For distant stars, the observed spectral energy distributions do not match those of the same spectral type near the sun. This is known as the colour excess of a star and is given by

$$E(B-V) = (B-V)_{\text{observed}} - (B-V)_{\text{true}}$$

By measuring $E(B-V)$ for an object, the visual absorption can be calculated, and the true distance can be derived. Throughout the galaxy, measurements show that the colour excess and the visual absorption are given by

$$A_V = 3.1E(B-V)$$

We can plot interstellar extinction curves and we see that they are proportional to $1/\lambda$. In the UV part of the spectrum we see a hump in the extinction curve at about 2200Å due to carbon bonds. At long wavelengths, is very low and so we can use infrared images to observe behind dust clouds. Infrared images can also give us an indication on the dust grain composition as well.



B, V, I B, I, K

Barnard 68. A molecular cloud 400 light years away with mass ~ $2M_{\odot}$ showing the effect of extinction and interstellar reddening (left) and how to observe through such a cloud with infrared observations.

(b) Thermal Emission

The dust grains are heated due to the absorption of radiation. Their radiation spectrum is thermal, which implies that Wien's law holds:

$$\lambda_{\max}T = 2.898 \times 10^{-3}\text{mK}$$

Where $\lambda_{\max}$ is the wavelength at which the peak occurs for the blackbody function and T, as you probably expected, is the temperature.

At:

- $T = 300\ \mathrm{K}$, $\lambda_{\max} \approx 10\ \mu\mathrm{m}$ (which corresponds to silicates)
- $T = 30\ \mathrm{K}$, $\lambda_{\max} \approx 100\ \mu\mathrm{m}$

Thus, infrared observations can allow observations of both obscured sources and interstellar dust itself. The $10\ \mu\mathrm{m}$ and $12\ \mu\mathrm{m}$ observations correspond to vibrational bonds in silicates and water ice. Silicates are a component of dirt on Earth, so often the dust grains are referred to as 'dirty ice'.

(c) Reflection nebulae

If dust lies near a bright star, the starlight may be reflected towards us if the energy is not sufficient to ionise the material and produce an emission nebula. We still get scattered light due to the dust. The dust then appears as a luminous area whose spectrum matches that of the nearby star. The scattering is due to carbon compounds and other elements such as iron and nickel in the dust. Elements such as iron and nickel can be aligned with the galactic magnetic field, so light scattered by these elements can be slightly polarised.

(d) Optical absorption

Spectra of heavily obscured stars show absorption bands in optical wavelengths due to the dust. There is a reasonable correlation with dust extinction, but levels off at high A_V, so they are diffuse clouds only. These are called *Diffuse Interstellar Bands* (DIB's).

The most prominent DIBs are broader than interstellar atomic lines. Several other characteristics such as the lack of emission and constancy with wavelength suggest they are not quite the same as dust. Their general origin remains unknown after 60 years, but some appear to be due to complex organic (carbon based) molecules such as C_{60}^{+}.

Lecture 2 – Galaxies

Learning Objectives

The material that we will cover in this lecture includes:

5. What a galaxy is
6. What the terms ‘Stellar populations’ refers to within the context of galaxies
7. The different types of galaxies (spiral, elliptical, irregular) and their properties
8. What galaxy clusters and super clusters are
9. How these clusters effect galaxy evolution in events such as mergers

Introduction to Galaxies

Galaxies appear in the sky as huge clouds of light thousands of light-years across. Each contains anywhere from a million to 9 million million stars 10^{12}. Gravity binds the stars together so that they do not wander off freely from each other. As well as stars, galaxies contain inter stellar material bound together by mutual self-gravity. That is,

$$\mathrm{v_{esc}} < \left(\frac{2\mathrm{GM(Galaxy)}}{\mathrm{R}}\right)^{\frac{1}{2}}$$

Almost all light from a galaxy comes from its stars. Galaxies appear to exist in the larger universe and appear to be a fundamental structure of the universe.
That is,

- Universe contain galaxies.
- Galaxies contain stars.
- Stars possess planets.

It wasn't clear until the 1920's what a galaxy was. It was thought that galaxies were just nebulae and they were referred to as such. From their appearance they may be confused for other nearby objects and phenomena such as H II regions in our own galaxy.

In 1924 Edwin Hubble settled the issue observationally. To do so what he did was look at Cepheids in 3 spiral nebulae (one of which was the Andromeda galaxy). He established that their distance was large in comparison to the size of the Milky Way. Therefore, it must be a different object. He concluded that Andromeda was 300 kpc away and so must be a galaxy in its own right. He was slightly off in his calculation as Andromeda is about 800 kpc away. He came up with a rough classification system for galaxies which is still in use today.

From the apparent brightness of giant and supergiant stars, the closest galaxies are the Magellanic Clouds, visible by eye from the Southern hemisphere, at a distance of only $55\text{x}10^3$ pc (55 kpc). Only galaxies within 20×10^6 pc (20 Mpc) can have individual stars resolved by telescopes; beyond this distance stars are normally unresolved and merged together in images. By extrapolating the numbers seen in surveys, there are about 10^{11} galaxies potentially observable in the Universe.

Stellar Populations

Spectroscopy of stars in the Milky way and nearby galaxies shows that they can be split into roughly two compositional classes.

- Population I: These are stars with chemical compositions similar to the Sun, where the abundance of elements except H and He is ≥ 1% of the mass of the star.
- Population II: Stars with metal abundances 1/10th to 1/100th that of the Sun.

You've probably seen this concept when talking about the HR diagram. In older stars you would have seen how you can use the HR diagram to work out the age of the cluster whether that is open / globular cluster. Globular clusters have stars with lower metallicity (Z) in comparison to younger stars in the open clusters. Because of the increasing metallicity of stars due to ISM recycling, Population II stars are believed to be older than Population I stars. This is supported by the observation that all O and B type stars are classed as Population I.

The first stars born in the Universe would have only contained Hydrogen and Helium and are called Population III stars. However, no Population III star has been discovered in our galaxy so far.

The difference between open and globular clusters help us think about old and new material in the galaxy. Globular dusters are older and form a spherical distribution. Open (or Galactic) clusters are newer and confined to the galactic disk. This suggests that star formation used to occur in a large spherical volume but now it occurs within the disk (in our galaxy at least). This is supported by the fact that Globular clusters are free of interstellar gas and dust, which is the material that forms new stars. Open clusters are frequently associated with gas and dust. The concept of stellar populations is important for our understanding of evolution of our galaxy.

Galaxy Types

In his studies, Hubble realised that not all galaxies had the same shapes. He realised that some never have a spiral shape. He set out his scheme for classifying stars based on appearance in his 1936 book "The realm of the nebulae." with later additions and modifications.

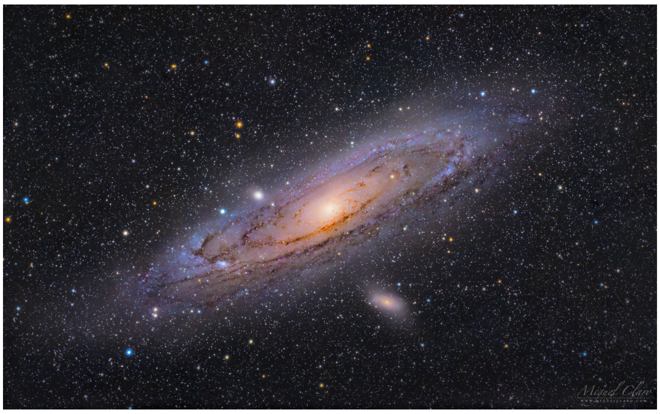
He had the main class of galaxies as: elliptical, lenticular, spirals and then a fourth class called irregulars for those that didn't fall into his class system. Today we've simplified it to basically 3 classes. (as we usually denote lenticular as S0). The 3 main types are:

1. S - Spiral
2. E - Elliptical
3. Irr - Irregular

Spiral Galaxies

<u>Examples</u> - Whirlpool Galaxy, Andromeda Galaxy.

Whirlpool Galaxy (M51)



Andromeda (M31)

Spiral galaxies have a disk of stars and ISM orbiting a central hub. The disk normally shows well developed spiral arms, containing many regions of star formation (Population I stars, Open clusters, H II regions). The degree of winding of the arms is indicated by a letter system running from a to d. That is, we use a system of letters to denote the shape of the spiral arms for this type of galaxy.

The two important features of the classification system are:

1. The openness or tightness of the winding of the spiral pattern
2. The relative importance of the central bulge and the disk of the galaxy.

Sa galaxies have the largest bulges and the most lightly wound arms.
Sd galaxies have the smallest bulges and most open arms.

The cartel bulge becomes less important with a more rapidly rotating disk, while the spiral arms become more open as the fraction of gas and young stars in the disk increases. The significant amount of ISM means that interstellar dust is also present in large quantities.

The central bulge is composed of Population II stars and possesses little, if any, star formation. The arms are outlined by clumps of bright and hot 0 and B type stars, and the compressed dusty gas from which these stars form.

Sometimes spiral galaxies have a bar of stars through the central bulge. These are called barred spiral galaxies - SB. About a half of all spiral galaxies have a central linear bar. In general, barred spirals can be subclassified based a to d as well for example, SBa, SBb etc. Where the subdivisions have the same meaning as before.

In general, the spiral pattern in barred spirals is well defined.

Hence Spiral Galaxy types can be Sa, SBc etc. Total luminosities of 10^8 - 10^{11} $L_{\odot}$

Luminosity can be important for trying to work out the distance to galaxies that are so far away that we cannot distinguish individual stars. Some people have taken to adding a luminosity class to spiral galaxy classifications. We can work out the distance to a galaxy with the absolute and apparent magnitude measurements for the galaxy.

An important aspect of spiral galaxies is the presence of an ISM containing both gas and dust. You can still tell it is spiral when looking at it edge on because of this. That is, we have a lane of obscuring dust in the disk of the galaxy.

Light from spiral galaxies has important contributions from small blue stars suggesting that star formation is still taking place in spiral galaxies.

Typical radii for the luminous part of the disk in spirals is about 10 to 30 kpc
Stellar masses of the spirals we can see range from 10^7 to 10^{11} $M_{\odot}$

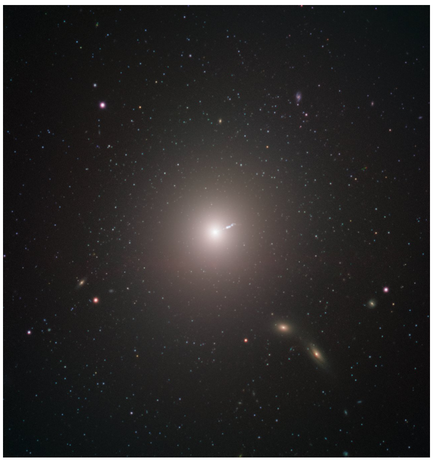
For spirals we can get a good idea of the luminosity as a function of the radius.

$$L(r) = L_0 e^{-\frac{r}{D}}$$

Where L_0 is the luminosity at the centre and r is the radius. D here is known as the luminosity scale length. It gives an indication of the characteristic radius of the galaxy in visible light. Typical valves for D are about 5 kpc.

Elliptical galaxies.

Examples - M87, M32.



M87

M32 (in cut-out with Andromeda behind)

These are symmetric elliptical systems, composed of low-mass stars. The degree of ellipticity is given by numbers 0 to 7, so E0 is spherical, E7 is highly flattened.

Most common type of elliptical galaxies are called dwarf ellipticals as they are also the smallest. Typically, the size is around a few kilo parsecs. They normally have masses of around a few million solar masses. Giant ellipticals have extents up to 100 kpc and masses up to $10^{12}\ M_{\odot}$.

They have very little ISM, so very little star formation is occurring. Luminosities of normal elliptical galaxies span 10^5 to $10^{11}\ L_{\odot}$.

Studies of the 21cm line of H I regions indicates that the mass of the ISM may be about 1% of the mass of the stars that we see. Low gas content rules out the possibility that elliptical eventually flatten out to form spirals. Continuing formation of stars depletes the ISM of matter. This essentially shows us how elliptical can't be evolved spirals (as was previously thought) as the spirals have more gas and dust (which they shouldn't under this assumption as they'd have less gas and dust after star formation has occurred.)

Ellipticals generally contain more evolved stars. That is there are generally no 0 or B type stars in stellar populations have metal abundance's that are not low. Giant ellipticals have metallicities twice that of solar values.

Some elliptical galaxies rotate very slowly. Their slow rotation means they could collapse without much flattening. Also, for these slow rotating ellipticals, they tend to be more luminous and the stars in this case will orbit the centre of the galaxy with random orbits.

In between E and S galaxies are the S0 galaxies - these have a disc and bulge, but without any spiral arms. These are what were previously referred to as lenticular galaxies. These form a transition between spiral and elliptical galaxies. They resemble elliptical as they lack extensive gas and dust and prefer regions of space that are more densely populated with other galaxies (similar to how ellipticals are generally found).

Irregular galaxies

<u>Examples</u> - Magellanic clouds, M82.

Large Magellanic Cloud

M82

Irregular galaxies show two main types:

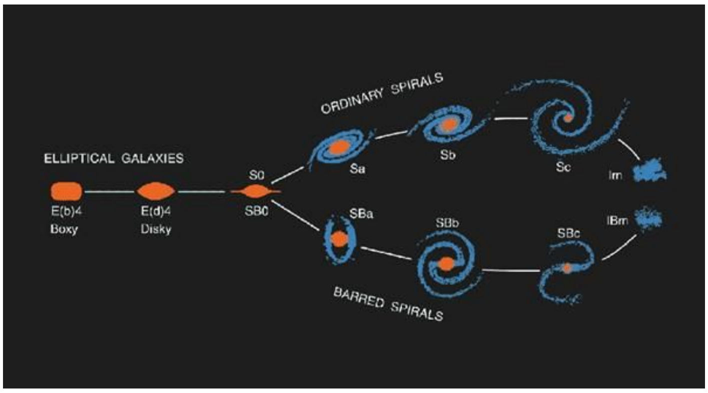
- Irr I are resolved as stars and nebulae
- Irr II just have a general amorphous appearance.

Several other active galaxy classifications but in general they can be added as sub-divisions of those already mentioned.

In the local universe, the number of galaxies are:

- Spirals ~ 35%
- Ellipticals ~ 15%
- Irregulars ~ 50%

The different morphologies may be remembered via the Hubble Tuning Fork diagram.



Hubble Tuning Fork Diagram

Galaxy clusters and superclusters.

Note: these are not to be confused with ‘Galactic (Open) clusters’, which are actually clusters of stars!

Galaxies appear to lie within gravitationally bound groups where the individual velocities v < v(esc). The groups are called galaxy clusters.

Our own galaxy cluster is called the Local Group, approximately 1 Mpc (10^6pc) across. It is dominated by the Milky Way and M31 (Andromeda galaxy). There are > 30 galaxies in the local group - this makes it a small cluster. Most galaxies in the local group are small elliptical or irregular galaxies, sometimes called dwarf galaxies. It is likely that not all have been identified.

The nearest large cluster is the Virgo cluster, with ~1000 members. The name for observed galaxy clusters generally comes from the constellation in which they are found (e.g., Virgo, Hercules, Hydra etc.). The Virgo cluster, the Local Group and other nearby small clusters together form the Virgo supercluster.

The Virgo Supercluster

Galaxy clusters are useful for further study as they can give us an indication of how galaxies themselves form and evolve. One theory suggests that cluster sized objects formed first before breaking apart to form the galaxy sized objects we can currently see in the night sky. Conversely, the other theory is that galaxies formed first, then rearranged into super clusters.

Galaxy evolution.

Within galaxy clusters, collisions between galaxies are observed to occur. The Andromeda Galaxy is moving towards us at –300 km/sec. At a distance of 780 kpc, it will reach the Milky Way in approximately 2.5×10^{9} years. We'll see now, why you shouldn't worry about getting insurance for that eventuality...

In most galaxies the separation between stars is ≈1 pc with an average radius for a star given as $\approx R_{\odot}$. The average cross section for one star to collide with another can be found with:

$$\sigma = \pi D_{\odot}^{2} = 4\pi R_{\odot}^{2}$$

Now, if we consider a star from Andromeda moving with relative velocity, v, to the stars in the Milky Way, in time, t, the star will travel a distance $d = vt$. This sweeps out a volume, V, given by:

$$\begin{aligned} V &= \pi D_{\odot}^{2} d \\ &= \pi D_{\odot}^{2} vt \\ &= \sigma v\, t \end{aligned}$$

If we define the number density, n_p, then the number of stars that the star from Andromeda will end up encountering will be given by:

$$n = n_p \sigma v\, t$$

We can then establish the mean time between physical collisions, $t(\text{coll})$, by setting $n = 1$. This gives us:

$$t(\text{coll}) = \frac{1}{n_p \sigma v}$$

Therefore, we should be able to see, that in a collision between galaxies not a single star to star collision may actually occur (your insurance premiums should be fine). However, what will end up happening is that the ISM in the merging galaxies will get compressed. With this compression we end up getting gravitational collapse of the gas which then results in a large burst of star formation in the region of the merger.

Mergers of the Antennae Galaxies (NGC 4038 and NGC 4039)

If a single galaxy is left after the merger, it will often take an elliptical form. The new form will have little ISM left over. Studies of merging galaxies now make it clear that the larger giant elliptical types are formed from the collision of large galaxies.
Collisions between galaxies sometimes results in a “tidal tail” due to a large number of stars being placed on slightly different large orbits around the galaxy. As the orbital velocity of each star will be slightly different, the stars will spread out to form a stream. The observed motion of galaxies and the stellar tails allows us to reconstruct part of the group’s history. As galaxies merge, or have near misses, bright stellar trails can be ripped from the galaxies resulting in a loss of gas. Similarly, there is evidence that H I gas from the Magellanic clouds has been pulled out into a stream around the Milky Way.

Studies of the Milky Way and the Andromeda galaxies show evidence for star-streams above and below their galactic disks. These halo streams are left over from the galaxies swallowing smaller galaxies.

Lecture 3 – The Milky Way

Learning Objectives

The material that we will cover in this lecture includes:

10. The structure of the Milky Way
11. How we can establish structure of the galaxy we are in using H I regions, open clusters and globular clusters
12. Determining the properties of the inner region
13. Our supermassive black hole

Introduction to the Milky Way

The study of the large-scale structure of the Milky Way is difficult from our perspective. We are in the plane of the galaxy, so we see is the band of light from the spiral arms. The dust acts to observe our view and so we cannot look very far into the galaxy. As such, we have to make use of different wavelengths besides optical wavelengths to study the centre of the Milky Way.

The structure of our Milky Way may be traced by studying the location and distribution of a number of components that make up the content of the Milky Way. These can also be used to set the scale of the Milky Way. The first evidence of our place in the galaxy came with the work of Harlow Shapley. He studied the distribution of globular clusters by looking at Cepheids and RR Lyrae stars. With this he showed that globular clusters form a spherical distribution. The centre of this distribution he found as about 10 kpc from the Sun. It was presumed that the centre of this globular cluster was the centre of galaxy. The distribution of Population I and II stars can give us an indication of how the galaxy evolved.

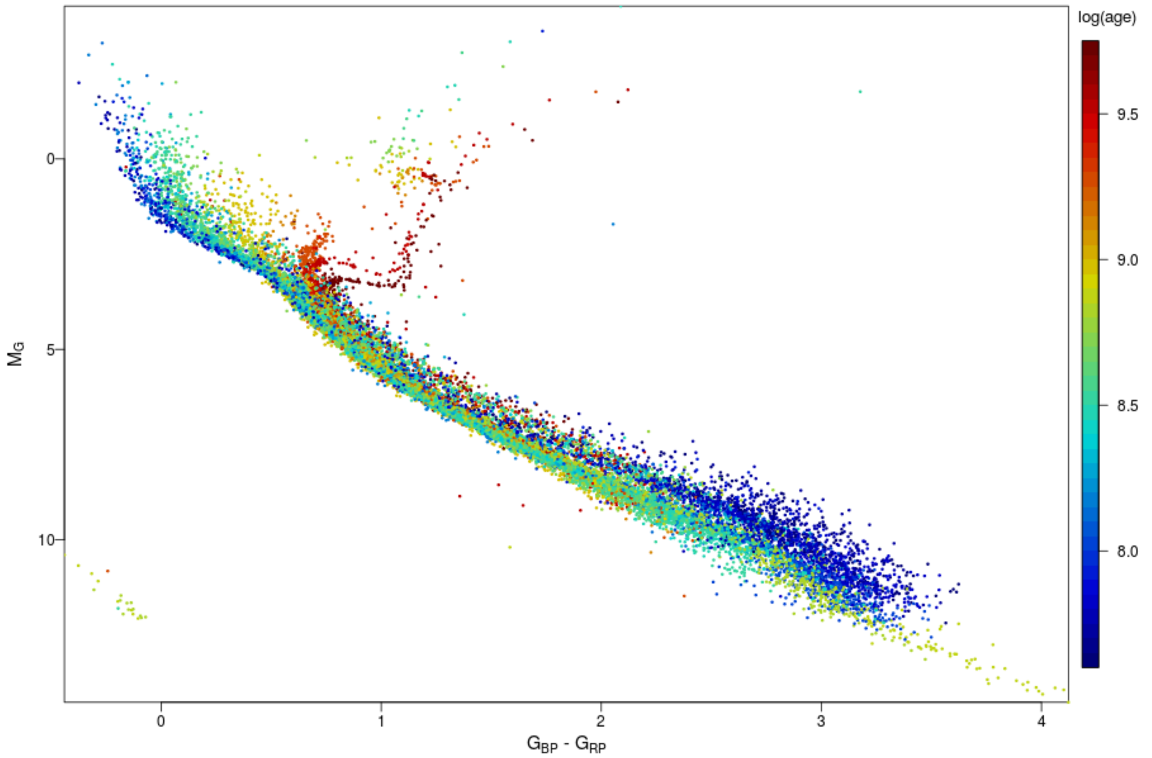
The structure of our Milky Way

(a) Open (or Galactic) Star Clusters

These contain between 50 - 1000 Population I stars in a random pattern. Remember that these stars are considered as forming from younger material and, therefore, have a higher metallicity. Some are associated with regions of interstellar gas and dust. They are confined to the galactic plane.

All the stars in a star cluster formed at roughly the same time, so we can establish the distance by fitting the cluster main sequence. Remember, with the Hertzsprung-Russell (HR) diagram of a individual cluster, we are studying stars at a common distance. If we don't know the distance to the cluster we can plot the apparent magnitudes on the HR diagram. We then work out how many magnitudes up or down we need to shift the HR diagram to get the right absolute magnitudes for each spectral type within the cluster. The amount of shift gives us the distance modulus to the cluster and therefore the distance.

This is known as main sequence fitting. It gives you similar information as you would get from establishing the spectroscopic parallax measurement, but in this case you are using information from all main sequence stars in the cluster, so it is more accurate than the study of a single star.



A combined HRD for 32 open clusters from Gaia, where the effects of cluster age can clearly be distinguished. Blue is younger, red is older.

<u>Example</u>: Praesepe is in the Beehive Cluster, which is an open cluster in the constellation of cancer. It is one of the nearest Open Clusters to Earth. It is one of the first objects that Galileo studied with the telescope. The main sequence of Praesepe is observed to lie 5.85 magnitudes fainter than the absolute main sequence. What is its distance?

The main sequence generally breaks at a certain magnitude. This is known as the *turnoff point*. Stars above the turn off point now lie above the main sequence, so they are now more luminous and therefore larger than main sequence stars of the same special type.

Each cluster has its own turnoff point depending on its age. Stars brighter than this have already evolved off the main-sequence and into giant stars. This marks the time elapsed since the formation of the cluster. The loci of stars in the H-R diagram form the isochrone for that particular lifetime.

Open clusters are found to be anywhere between 10^6 years and 10^9 years old.

The shape of our galaxy may be found from plotting the positions of young ($< 10^8$ years old) open clusters and HII regions.

With this information we know that the sun lies in the arm of a spiral galaxy.

(b) HI observations.

All material in the galaxy orbits the centre of the galaxy. If the galaxy was a rigid body, then all of the material would orbit at the same period. However, in reality material close to the centre orbits with a shorter period than that further out. We've seen this sort of behaviour before though when considering planets. The Earth takes 365.25 days to orbit the Sun while Saturn takes 10,759.22 days.

In this scenario when we say that the orbital period depends on distance from the centre, we say that the material exhibits *differential rotation*. The orbital period of a particle will depend on the mass about which it is orbiting and the radius of the orbit.

We use orbital periods at different distances from the galactic centre to tell us about the distribution of mass in the galaxy.

Galactic rotation.

Radio surveys along the Milky Way reveal large amounts of H I exist at particular radial velocities, which depend on the point being observed.

Differential galactic rotation causes Doppler shifts in spectral lines for gas observed at different distances from the centre of the galaxy. The Doppler shift depends on the relative velocity between the observed position and the Sun (at our viewing position and radial velocity). Material closer to the centre of the galaxy along our line of sight will be red shifted with a higher degree of red shift the closer you get to the centre. For material further out than the Sun it will be blue shifted due to differential rotation.

Assuming that the H I with the highest radial velocity is travelling directly along our line of sight, then the H I spectrum can be converted from radial velocity into distance. This can be done using Doppler shifts from our old friend the 21cm line.

This allows the H I distribution to be mapped and we see that it primarily exists within the spiral arms. The gas and dust of the disk lie in a very thin layer. Near the sun's position most of the neutral hydrogen is within 100 pc of the midplane.

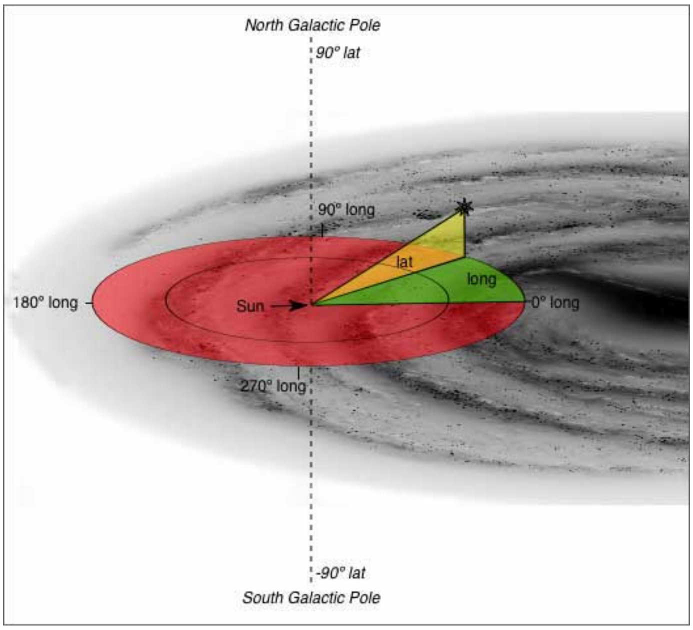
How accurate is this idea that it is a thin disk? Well 95% of the disk stars are found in this disk which includes all of the young and massive (Population I) stars. The density of stars, n, in the disk drops by a factor of e as we move outwards in R by one scale length, h_R, such that

$$n(R) \propto e^{-\frac{R}{h_R}}$$

For the thickness of the disk, we can say that the scale height, the distance perpendicular to the disk to see density drop by a factor e, is about 300 - 400 Pc. The central bulge has a scale height of about 1000 pc and is sometimes referred to as the *thick disk*, we will come back to the central bulge later in this lecture though.

To determine the differential rotation curve for the galaxy it is useful to introduce a set of coordinates. In this case we call this the *galactic coordinate system* which is different to the system I showed you in your very first lecture for PHY1002 (for those that took PHY1002).

In this case we use the Sun as our viewing point. Then we use the galactic latitude, b, as a measure of the angle of an object above or below the galactic disk. The galactic longitude, l, then gives the angle around the plane starting from the direction of the galactic centre.



The Galactic Coordinate System

We also use a reference frame for calculating the velocities which is called the *Local Standard of Rest* (LSR). If the only motion that the Sun had in its orbital motion was about the galactic centre then the local standard of rest would coincide with the Sun's and we would just measure the motions of other objects with respect to that of the Sun (heliocentric motions). However, as a result of gravitational interactions with its nearest neighbours, the Sun has a small motion superimposed on its circular orbit motion so we can't use that as a reference point for velocities. This random additional motion is of the order of a few tens of kilometres per second and is common for all stars in the disk.

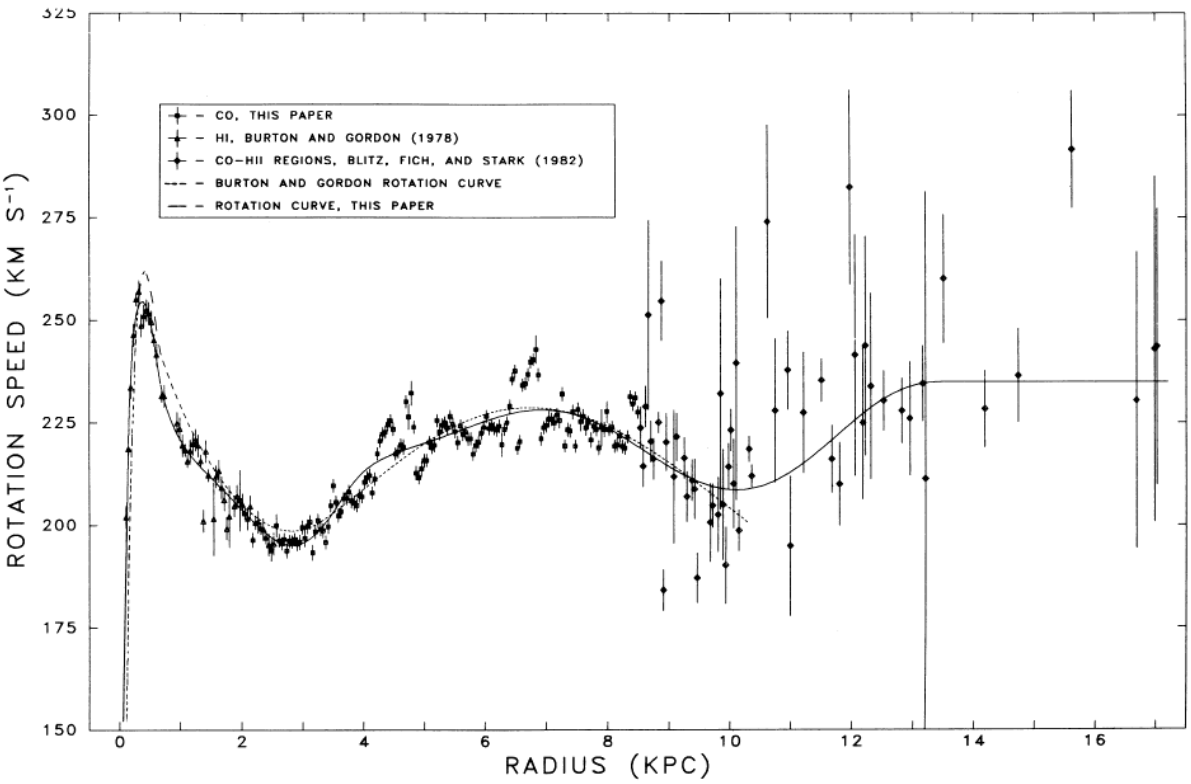
There are 2 ways to define LSR.

1. *Dynamical*. The origin of the coordinate system orbits at a distance of R_0 from the galactic centre with R_0 representing the distance from the galactic centre to the Sun.
2. *Kinematic*. The origin of the coordinate system moves with the average velocity of all the stars in the vicinity of the sun. This averages out the effects of the random motions of the stars (which will affect all stars in the disk to some extent).

The 2 systems will result in the same velocity system essentially. The small difference between the two methods can tell us some interesting things on the dynamical properties of the galaxy. The Sun is moving at about 20 km/s with respect to the LSR. Once we know the motion of the sun though we can use it to correct Doppler shift measurements to give us the radial velocity of un object with respect to the LSR

The relative radial velocity.

Often the galaxy is divided into quadrants based on l to work out the radial velocity v_r



Clemens (1985), ApJ, 295, 422

The quantity R_0 is the distance from the Sun to the galactic centre and is established by looking at globular clusters that we will talk about soon. In the 80's this was revised using masers (stimulated spectral line emission from sources in the galaxy) to work out a value of 8.5 kpc for R_0. An agreed upon value is essential for studies that look at various aspects of the galactic structure between research institutes as these values are pivotal for studies to be repeatable by different researchers and facilities. Likewise, a value for the orbital velocity (v_0) of 220km/s was adopted at this time by the IAU as well.

By looking at the Doppler shift (maximum Doppler) at each longitude we can work out the angular speed for that particular line of sight as,

$$\Omega(R_0 \, sin \, l) = \frac{v_{max}}{R_0 \, sin \, l} + \Omega_0$$

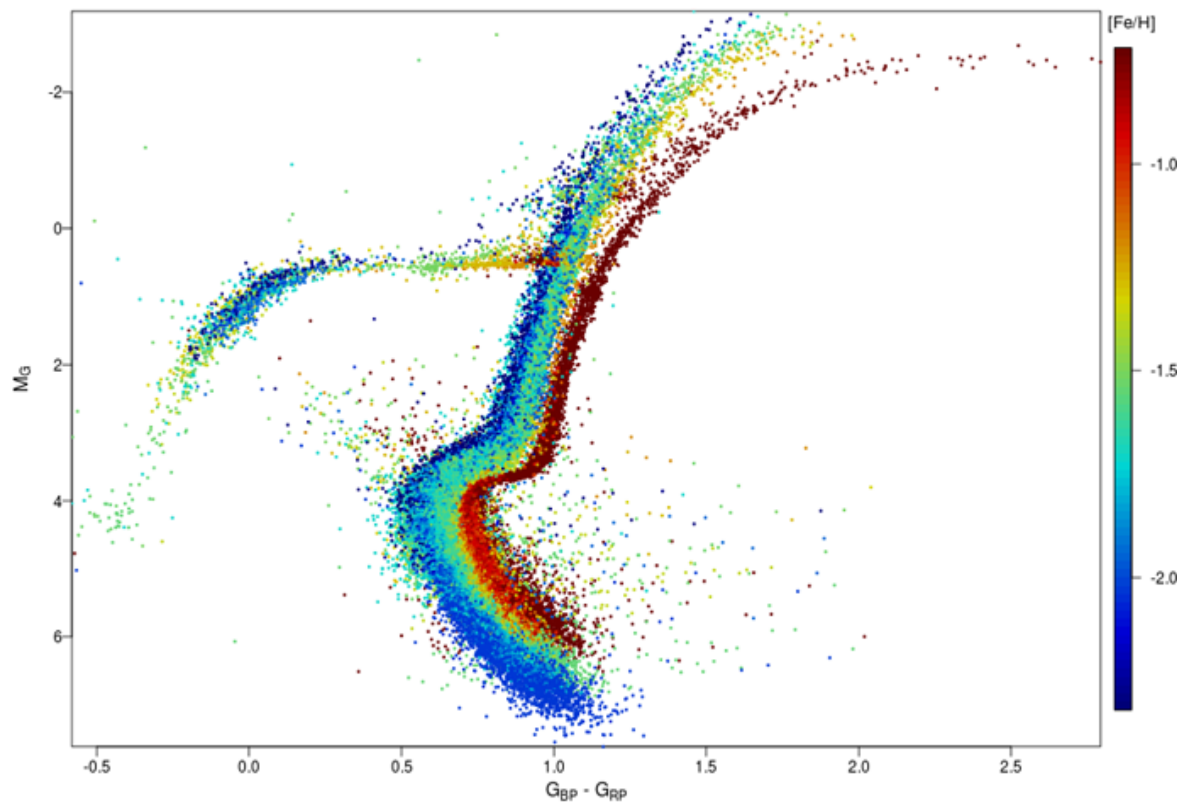
Where v_{max} is the maximum radial velocity along the line of sight and Ω_0, as you might expect, is the angular speed for the Sun and our part of the galaxy.

Using H I to determine the rotation curve outside the Sun's orbit is a little more difficult as there is no maximum Doppler shift along any line of sight. To get the rotation curve passed the Sun we have to use molecular clouds associated with H II regions. This was done using the CO emission line from stars exciting the H II region surrounding it to get a good distance measurement. This gives us a fairly reliable rotation curve out to about 20 kpc. The implication of these measurements was that it meant that there is actually more mass outside the sun's rotation curve than previous thought. This can be seen in the rotation curve above and we'll talk about this fact a bit more later.

(c) Globular clusters

These contain between 10^4 and 10^6 stars in a spherical system. They are not constrained to the milky way in the sky.

Main sequence fitting shows that
i. Globular clusters are generally very distant (many kpc).
ii. Globular clusters are generally very old, around 10^{10} years old.



H-R diagram of a globular cluster. This is produced from Gaia data of a combined 14 globular clusters. Red shading contains more heavy elements, blue is fewer.

Old globular clusters appear to form a spherically symmetric distribution around the centre of our galaxy. This places us at $d_\odot$ approximately 8,500 pc from the centre of the galaxy. Combined with the Sun's measured velocity relative to this point of 220 km/sec, this implies a mass within the Sun's orbit of $10^{11}M_\odot$. Hence the Milky Way contains ~10^{11} stars.

Population II stars in these globular clusters are thought of as the old component of the galaxy and are metal poor. They form a spherical distribution as opposed to a disk. By spherical distribution here we mean that they form a spherically symmetric distribution whose density falls off with increasing distance from the galactic centre. We call this spherical distribution the *halo*.

The stars in the metal-poor halo do not have any organised rotation about the centre of the galaxy. Their motions resemble comets in the solar system with their orbits following random directions with a frequently eccentric path. They spend most of their time in the outer regions of the galaxy but plunge deeply inward at pericentre (the point of nearest approach in its orbit about the galaxy – the equivalent of perihelion). They also seen to have a larger velocity spread than Population I stars.

As mentioned previously, a large amount of mass is seen outside the solar rotation curve, this implies significant mass from material we can't identify as stars. We call this *dark matter* and we will come back to it at some point. We usually assume, without any real compelling reasons, that this falls in a roughly spherical *dark halo*.

A rough sketch of the Sun's position in the Milky Way

The Central Regions of our Milky Way

The central bulge can be studied at radio and near-infrared wavelengths, where absorption due to interstellar dust is minimised.

Photometric and radial velocity surveys of stars in the 1990's showed the probable presence of a central bar, confirmed during the past 20 years.

In 2019 the central bar was directly mapped using distances to red giant stars measured by GAIA and surveys performing infrared photometry of stars towards the galactic core.

The very centre of the galaxy is invisible at optical wavelengths due to high interstellar absorption (A_V of approximately 25 magnitudes) but can be observed at infra-red and longer wavelengths. When we refer to the central regions, we normally refer to regions within about the inner 500 pc of the galaxy.

In radio we can see continuum emission from ionised gas and line emission from molecular clouds which are remarkably warm and dense in comparison to molecular clouds in other parts of the galaxy. In the infrared, continuum observations tell us about the dust temperature and opacity while spectral line observations tell us about the neutral and ionised gas. Near infrared can be used to study star clusters near the centre. Millimetre observations tell us about the cold dense component.

Due to the distance to the centre, many studies are limited by poor angular resolution.

Observations show that the central parsec is rich in star formation. The star formation rate is estimated to be as high as $1 M_{\odot}\ \mathrm{year}^{-1}$. The central cluster shows evidence of 10^4 new stars superimposed on 10^6 older cluster stars.

At the very centre of the Milky Way is a bright compact radio source, named Sagittarius A* (SGR A*). It is faintly visible at infra-red wavelengths, but stars are observed in orbit about it with orbital periods ≥ 11 years.

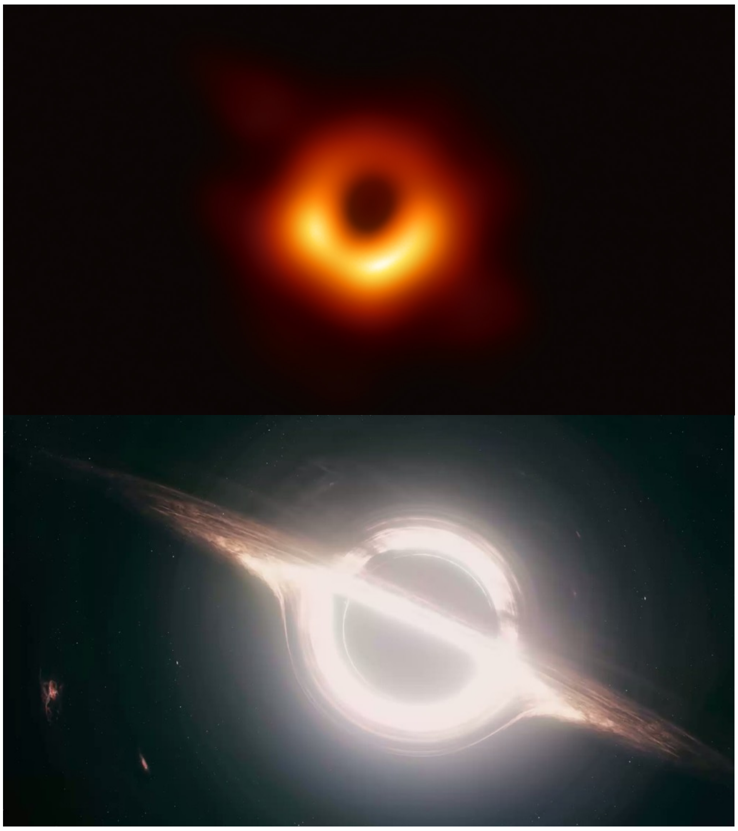
The stellar orbits imply a central mass of $4 \times 10^6 M_{\odot}$ but has a size ≤ 6 times 10^{10} m as measured by radio interferometry. It is believed to be a supermassive black hole at the centre of our Galaxy.

The Schwarzschild radius for a 10^6 $M_{\odot}$ black hole is 3×10^6 km. Remember the Schwarzchild radius is given by:

$$R_s = \frac{2GM}{c^2}$$

And the Tolman-Oppenheimer-Volkoff limit for neutron stars (similar to the Chandrasekar limit for white dwarfs) is around $M_{TOV} \sim 2.3 M_{\odot}$ so the mass and radius of the source for SGR A* is well within the limits we'd expect to have a black hole. Therefore, it was thought that it was highly likely that SGR A* is a black hole in the centre of the Milky Way. Specifically, it would be classed as a *supermassive black hole*, which are black holes with mass greater than 10^5 to 10^6 $M_{\odot}$. Note, there are 3 types of black hole.

Sub-mm observations by the Event Horizons Telescope at a wavelength of 1.3mm were obtained in 2017. These have finished processing resulting in a resolved image of the event horizon shadow of M87*. They are hoping to do the same for SGR A* at some point. In 2020, the Nobel Prize in Physics was awarded partially (half) for the discovery of this compact object (the other half going to the theory that black holes are a robust prediction of general relativity).



Top: M87* image from the Event Horizon Telescope
Bottom: Black hole Gargantua from the movie Interstellar

Lecture 4 - The Cosmological Distance Scale

Learning Objectives

In this lecture we will cover:

1. Different methods for calculating distances to objects in the universe
    a. Variable stars
    b. H II Regions
    c. Supernovae
    d. The Tully-Fisher relation
2. The concept of the extragalactic distance ladder

A fundamental measurement that we need for our studies is to establish the distance to external galaxies. There is no single physical body used within relatively nearby galaxies, but a variety of methods are used.

These various methods work at different scales (as we will soon see) and therefore we need to make use of several of them to get a full picture of the distances to different objects/galaxies. These various methods will ultimately help us in determining Hubble's constant.

(a) RR Lyrae Stars

By monitoring several stars for brightness variations, we see that some stars oscillate with time. These are referred to as variable stars. Eclipsing binaries appear as variable stars but there are many stars that have luminosity variations due to physical variations associated with the actual stars.

Typically, these stars that we use for standard candles are found within a region of the HR diagram referred to as the instability strip. Though other variable stars can be found in other parts of the HR diagram. Many stars (in fact, probably most) will have some sort of luminosity variation over time, for example, the Sun's luminosity varies slightly over the 11-year solar cycle

These are linked in a way to those stars explained in the next section (Cepheid Variables). They were discovered in the 1890's or so as a cluster type variable. In about the early 1900's (about 1915 - 1930) it became increasingly clear that RR Lyrae stars are different entities with different properties to the previously known about Cepheid variables. Differences include shorter periods, different locations and chemical abundance's to Cepheid variables.

RR Lyrae stars are unstable Horizontal Branch stars (A–F giants), undergoing regular intrinsic brightness changes. They are found in globular clusters (hence the name 'cluster variables'). Over 80% of variables known in globular clusters are RR Lyrae stars. They are found at all galactic latitudes as opposed to classical Cepheids which are strongly associated with the galactic plane. They are older, relatively low mass Population II stars.

They have quite short pulsation periods (2 – 24 hours) and a small magnitude range (1.0 – 0.5 mags). They all have the same average luminosity of approximately $49L_{\odot}$ and absolute magnitude of $M_V \approx +0.6$.

They obey a weak period-luminosity relationship. This allows them to be easily observed in local Globular Clusters and nearby galaxies

They can be used as a means to work out distances. The absolute magnitudes are established by using clusters at a known distance from other techniques (e.g. parallax). Once you have absolute magnitudes calibrated you can use the period-luminosity relationship to work out distances to other stars or galaxies within the local group.

One of the big benefits of using RR Lyrae stars for this purpose is that the period is short so less observation time is needed.

Example: For an accurate lightcurve, objects require an apparent visual magnitude of $m_V \leq 22.0$. What is the maximum distance to which RR Lyrae stars may be accurately identified?

(b) Cepheid Variables

Side note here on astronomical naming conventions. Bright stars are named in order of brightness in their constellation with a Greek letter in front. For example, the brightest star in Orion is α Orionis (or α Ori). Some will still have their ancient names (α Orionis is Betelgeuse). Variable stars are named in order of discovery with the first labelled R the second S etc. Once those letters are used up it moves on to RR, RS, and so on (hence the name RR Lyrae as this was the prototype star that it was discovered with). Eventually you get to a limit using this method when you run out of letter combinations, which corresponds to about 334 variable stars in a constellation. After this point you use V335 Ori, V336 Ori and so on.

Cepheids get their name from the prototype star that the variability was seen initially with in the 1780's, which was δ Cephei. In truth it was observed with η Aquilae 9 few months before but δ Cephei was the more widely known variable star.

These are rare, unstable F–G giant and supergiant stars with pulsation periods in the range 1 – 50 days, and a magnitude range 2.0 – 0.5 magnitudes. They have absolute magnitudes $M_V = +2.7$ to -1.9. They typically have masses of about 6 $M_{\odot}$ and a radius of about 25 $R_{\odot}$.

Note δ Celphei has a period of 5.4 days and a variation in magnitude from 3.6 to 4.3 mag. Polaris is also a Cepheid variable (and was mistakenly used as a reference star for years to measure magnitudes of other stars).

For Cepheids in the Magellanic clouds, the brighter the apparent magnitude, the longer the period. This gives the Period-Luminosity (P-L) relationship.

This relationship was first seen by Henrietta Swan Leavitt who worked as a 'computer' (an old technical term given to those that did mathematical calculations – which is where the term comes from for modern computers) at Harvard college observatory at ~1912. It was eventually known as Leavitt's law. By looking at photographic plates to measure and catalogue the brightness of stars, with Cepheid variables she showed this rule. Over the years this has had to be improved, with better instrumentation etc. The most recent calibration (2007) was done using HST observations of nearby Cepheids which were Population I stars. For stars with period, P, in days and absolute magnitude, M_V, we find:

$$M_V = -2.43(log_{10} P - 1) - 4.05$$

Then the distance to these variables can be calculated with

$$5\, log_{10}\, d = \mathrm{V} + 3.34 log_{10} P - 2.45(\mathrm{V} - \mathrm{I}) + 7.52$$

Where V and I here represent the mean apparent magnitude in the visible and infrared bands, respectively.

Cepheids may be observed in local-group galaxies, and out to 20 Mpc with the HST. RR Lyraes and Cepheids lie close to each other on the H-R diagram. This region is known as the *instability strip* and is due to the formation of doubly ionised helium within the star (He III).

When looking at spectral lines in Cepheids we see Doppler shifts that vary during their light cycle. This suggests that the surface of the star is moving. The size of the star changes with luminosity changes. The spectral type will also vary during this cycle. Luminosity changes will result in changes in the surface temperature and the radius.

Below the photosphere, the temperature rises and at some depth He is ionised

$$HeI \rightarrow HeII + e^-$$

When the star evolves off the main sequence, the rise in temperature allows the lowest layers of the stellar atmosphere to doubly ionise the HeII

$$HeII \rightarrow HeIII + e^-$$

HeIII has a high opacity - it absorbs significantly more photons. Hence the radiation coming from the core and H-burning layers of the star is absorbed, the star heats up and the atmosphere expands.

As it expands the temperature drops and the HeIII recombines into HeII

$$HeIII + e^- \rightarrow HeII$$

The extra opacity source is removed, the star contracts and the cycle begins once more...The oscillations are in the radial direction of the star.

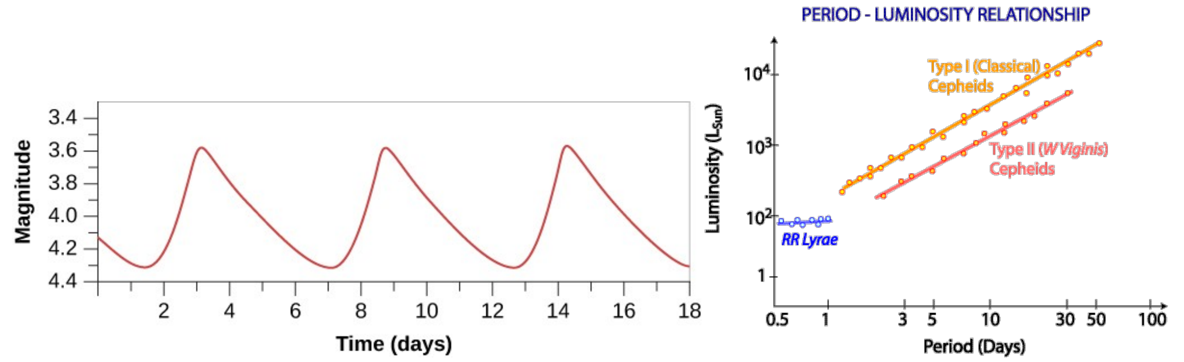
The Pendulum Star.

We distinguish between different types of variable stars by things such as their period and magnitude range. Work on Cepheids was crucial for Hubble to establish that Andromeda was not some nebula within the Milky Way but was a separate galaxy in its own right.

There are actually 2 different types of Cepheid variable this can be seen if we plot a histogram of Cepheid periods. We see a double peaked distribution, so there must be two different types. The shorter period Cepheids are due to Population I stars (which initially were observed in the LMC). In general, they are about 1.5 times the magnitude of Type II Cepheids. They are sometimes referred to as *Classical Cepheids* due to the fact they represent the originally studied Cepheid.

As you can probably guess, *Type II Cepheids* are due to Population II stars. They are sometimes referred to as W Virginis stars due to the star that was studied as the prototype.

There are slight differences in the period-luminosity relation between the 2 types. This fact caused some issues in the calibration. As Type II Cepheids were more readily studied in our galaxy for the calibration, the brighter Type I (Classical) Cepheid were studied in distant galaxies. As such the distance between distant galaxies and the Milky Way was underestimated. This was realised in the 1950's.



Left: The lightcurve for δ Cephei, with period 5.4 days
Right: The period-luminosity relations for Cepheids and RR Lyrae variable stars

(c) HII regions

For galaxies with known distances, the size of HII regions depends on the luminosity of the galaxy. Thus, the apparent magnitude of the galaxy together with the apparent size of the HII regions gives the distance.

Essentially, when we look at say a spiral galaxy, the red patches in the spiral arms tend to come from emission from H II regions. Studies of the Hβ line in extra galactic H II regions found that the luminosity of the region is proportional to the fourth power of the line width (σ^4) and the size of the region is proportional to the square of the line width (σ^2).

One of the main reasons that we use this method is that quite a few galaxies have H II regions within them. They do not require a SN to go off or the presence of a particular type of star. In general, the regions are large enough to be resolvable and essentially the uncertainties in the measurements tend to be smaller than those associated with SN 1a measurements.

(d) Supernovae

Supernovae can be as bright as a galaxy - so the obvious choice for a standard candle! From spectroscopy, there are TWO types of supernova.

Type II: Show mainly hydrogen. These come from massive star evolution (see previous lectures with Ernst). With Type II supernovae, the core of the star starts to convert to iron which will not undergo fusion. If the iron core remains smaller than the Chandrasekhar limit, then the core can be supported by degeneracy pressure. Past that point, however, it cannot be supported and will collapse. Energy that is liberated in the collapse is used by the iron and nickel in the core and they are eventually destroyed forming protons. Neutronisation occurs in the core forming a neutron core with lots of neutrinos. The large number of neutrinos can carry energy into the dense overlying material causing an explosion of material outwards (the supernova).

For Type II supernovae, the distance measurement is made by looking at the angular extent of the photosphere at two different time periods to get the angular velocity. Then we can use this angular velocity with the velocity of the supernova ejecta to get the distance. This method uses some assumptions and can be hindered by the fact that the ejecta is not spherically symmetric, and that interstellar extinction makes accurate measurements of the angular extent of the photosphere more challenging. Distance errors with Type II supernovae can be as high as 25%.

Type Ia: Show many emission lines, but no Hydrogen, and are also intrinsically fainter. They appear to be associated with older objects in our galaxy.

Type Ia appear to come from white dwarf stars, where somehow the mass increases to over $1.44M_{\odot}$, initiating collapse into a neutron star.

Two white dwarfs orbiting each other can spiral together due to gravitational radiation. Alternatively, they may come from binary systems where the matter transferred onto a white dwarf from the companion star increases the mass. The companion star can be another white dwarf or a larger star. So long as the mass accreted to the WD approaches the Chandrasekar limit, then it will undergo the initiation of fusion again, which leads to a runaway reaction that unbinds the star through a supernova.

What is believed to happen in the case of Type 1a supernovae is that the degeneracy pressure of the CO core of the star is able to support it, but at 99% the Chandrasekar mass, fusion is initiated at some point in the star in Carbon. This increases the temperature but due to the degeneracy in the WD it cannot regulate the temperature like a normal main sequence star (which would expand and cool) so the fusion goes into a runaway reaction. A substantial amount of C and O in the core of the star fuses to heavier elements in the order of seconds releasing enough energy through an explosion to release a shock wave of matter at significant speeds into the surrounding regions. There is an extreme increase in luminosity with the explosion.

Type Ia calibrations only became available in the 1990's: Type Ia Supernova have a peak brightness of M_V = −19.6 and are similar across different SN. This fairly consistent peak is a result of the fact that the explosion is a result of the WD going over (or at least close enough to) the Chandrasekar limit. Measuring the apparent magnitude at peak brightness allows us to find the distance to the supernova. Because of the precise calibration developed, this is now the preferred method of directly measuring distances to distant galaxies. The uncertainties in the distance measurements with Type 1a supernovae is quite small.
The problem is that you have to wait for a Type Ia supernova to explode!

SN 1994D, a Type Ia supernova, on the outskirts of the NGC 4526 galaxy.

(e) The Tully-Fisher Relation

If we want to know only the peak rotation speed ($v_{\max}$) in a galaxy, we can use a single-dish radio telescope with a large enough beam to include all the H I gas and measure how much there is at each velocity. This will give you the *global profile* for the galaxy.

In the late 1970's, Brent Tully and Richard Fisher found a relationship between the luminosity of a spiral galaxy and its spin rate. Brighter Galaxies rotate faster on average, which tells us that they are more massive. This might be expected, as the more mass a galaxy has in stars, the more luminous it will be.

At the same time, the more stars it has, the faster the orbital velocities will be in the disk. It is found that for spiral galaxies at a set distance from their centres:

$$L \propto v_{\max}^4$$

Originally, this was calculated using optical luminosities. However, the relationship is closer when using longer wavelengths such as infrared or microwave to establish the luminosities. Luminosity in the blue is more likely to fluctuate over time, as younger massive stars contribute to light in these wavelengths. When there is a burst of star formation in a galaxy, the luminosity in blue will be much brighter than it usually is, while the $v_{\max}$ remains the same, so the observed luminosities will scatter widely about their mean at any given rotation speed.

The distance to the galaxy is calculated by directly measuring the line width (for rotation velocity) and comparing the luminosity to the apparent brightness. This is usually calibrated with closer galaxies prior to being used to establish the distance to much further galaxies in the universe. The calibration for close-by galaxies is established with Cepheid variables initially.

Bizarrely, the Tully-Fisher relation implies that the amount of dark matter is somehow co-ordinated with the luminous mass. This implication comes from the fact that the rotation is dependent on the dark matter, while the luminosity comes from the disk. Without dark matter it would be an easier relation to understand. However, calibrations with other techniques on closer galaxies, show that the relation holds regardless.

The Tully-Fisher relation only works for spiral galaxies, as they have rotation, however, there is an analogue relation for elliptical galaxies (which we will not go in to in any depth).

(f) Brightest galaxies

Large clusters of galaxies are normally dominated by a single giant elliptical galaxy, called the centrally Dominant (cD) galaxy. They are also referred to as supergiant ellipticals. About 20% of rich clusters will contain a cD galaxy.

It is currently thought that they are formed through either galaxy mergers (due to cluster collapse in the galaxy), or through a ‘cannibalistic’ growth method due to tidal forces stripping contents of smaller galaxies. We’ve seen observations of mergers. For the tidal stripping, the second brightest galaxy in the cluster tends to be under-luminous in comparison to the cD due to the fact that it has been ‘eaten’ by the cD. This is often regarded as evidence for the mechanisms for their formation. Most suggest that the merger model is the more likely formation mechanism of the two.

The cD are generally found in the geometric and kinematical centre of the galaxy cluster.

They are some of the most luminous galaxies known. These all appear to have an absolute magnitude of $M_V \approx -21$. This then allows us to work out the distance to the cD galaxy.

The problem with this method comes with identifying the cluster and the cD galaxy. Usually, they are coincident with the peak in cluster X-ray emission, but again this is not an easy thing to identify.

The extragalactic distance ladder

These various techniques allow us to measure distances out to the far Universe and often we have to use a collection of them to get to different scale measurements. This is done by using one method to calibrate the next and so on. For example:

- The parallax method gives a direct measurement of distances to RR-Lyrae stars and Cepheid variables in the Milky Way.

- Observations of Cepheid variables gives distances to nearby galaxies.

- Observations of Type-Ia supernovae in nearby galaxies of known distance gives absolute magnitudes of Type-Ia Supernovae.

- Observations of Type-Ia in more distance galaxies gives their distances.

- Observations of spiral galaxies with Type-Ia distances gives the Tully-Fisher Relationship.

Assuming m_V ≈ 22.0 for ground-based telescopes:

Object	$\boldsymbol{M_V}$	$\boldsymbol{d}$ **(ground)**	**Distance**
RR Lyraes	$+0.5$	200 kpc	Globular Clusters
Cepheids	-6	4 Mpc	Local Group
HII regions	-9	16 Mpc	Virgo Cluster
Type I Supernovae	-19	2 Gpc	Distant superclusters
Brightest Galaxies/Tully Fisher	-21	4 Gpc	Distant superclusters

Lecture 5 - Cosmology

Learning Objectives

In this lecture we will cover:

1. The concept of cosmology
2. How we know the universe is expanding
3. Hubble's constant
4. The Big Bang Theory
5. The Cosmic Microwave Background radiation

Introduction

Cosmology is the study of the large-scale structure of the universe. Think of it like a study of gas in a room. In studying the gas in a room, you treat it as the motion of molecules within the gas in the room. You don't really care that the molecules are made up of atoms, the atoms are made up of protons, neutrons and electrons and the protons and neutrons are made up of quarks and so on. You only really care about how the molecules in the gas interact with each other within the room. Likewise, when we study the universe, we take a similar approach. We treat it as a gas. The equivalent of the interacting molecules are the galaxies. We are not really concerned about what the galaxies are composed of (like stars and planets - similar to the atoms, protons/neutrons etc). The smaller objects like stars can tell us stuff about the larger scale structures but they don't really affect the larger structure at the scale of the galaxies in the universe.

To study cosmology from the theoretical side of things, we treat the large-scale universe as a fluid and ignore the lumpy bits. The large-scale structure will only be affected by gravity, though the electro-magnetic force can carry useful information for us (though it won't affect the structure of the universe). In terms of gravity, we can describe it with general relativity, however, the Newtonian treatment of gravity is generally sufficient (don't worry, we won't worry too much about general relativity here!). When considering the large-scale structure of the universe with Newtonian gravity, we call this Newtonian Cosmology (which we'll discuss more in a future lecture).

As you may have seen from labs., there are some observational clues that we can utilise in understanding the large-scale structure of the universe (we'll come back to these in a bit). With all areas of astrophysics (and physics in general) we need to use a mix of theory and observations to try an understand the world (and universe) around us. As such, observational cosmology is a field that has seen a good deal of growth in recent years.

With theoretical models of cosmology, we will often start with the assumption that on the largest scales, the universe is both homogeneous and isotropic. This implies:

- Homogeneous: At any instant, the general properties of the universe (e.g., density, composition) are the same everywhere.

- Isotropic: The universe appears the same in any direction at any instant in time.

We will come back to these concepts when we talk about the Big Bang later. This is known as the cosmological principle. For superclusters of galaxies, the cosmological principle appears to hold.

The Expansion of the Universe

Another assumption that was often made in theories about the universe was that the universe is the same at all times in what were referred to as steady-state theories. A large amount of observational evidence has effectively reduced the importance of these steady-state theories. We'll come back to this evidence later on in this lecture. Effectively, steady-state theories were favoured for a while as they made the debate of beginning of the universe simpler. If the universe has a beginning, what came before it? If the universe exists in a steady-state, then you don't need to worry too much about that philosophical question. There are some problems with the idea of a steady-state universe besides accounting for the cosmic background radiation. One such issue is with the density. If the universe is expanding, then the density must be decreasing, so it can't be in a steady state. Theories such as the continuous creation of matter within the universe were suggested to counteract this and these theories can exist in tandem with the conservation of energy (they just happen at a low level, below what we can verify experimentally for the conservation of energy). The idea of a steady state universe has fallen away in more recent years.

We can make a very simple observation that effectively invalidates this steady-state idea and allows us to understand some concepts on the expansion of the universe. You can try this by going outside at night and looking at the sky. I hope that you would see that the sky is dark. The fact that it is, illustrates a problem known as Olbers's paradox.

The paradox basically states that, if you have an infinite universe any line of sight that you take should land upon a star. Therefore, the night sky (and the sky in daylight as well) should be covered completely in stars with no gaps. Now you could find a way out of this paradox and say, 'a line of sight might leave the galaxy before hitting a star', which is true. However, it will eventually hit a galaxy and you have now just readdressed the paradox in terms of galaxies (either your line of sight would hit a star in the foreground, or it would hit a galaxy in the background and the entire sky would still be covered in stars and/or galaxies). Now you may also try to explain it by saying 'well you taught us about extinction due to dust within our line of sight - couldn't that just obscure starlight', and you would be right. It will obscure light from behind through absorption or scattering. However, if the universe has existed for a very long time (as it would in the steady-state theories), the dust should have absorbed enough energy to have its temperature similar to that of the surface of a star, and so bright dust as well as stars and galaxies would fill up the night sky. Scattering by dust would make the night sky appear like a reflection nebula, and it would still be bright.

There are two ways out of this paradox. One is that the universe is of finite size, therefore, your line of sight will not cross a star in each direction. Another way out of this paradox is that the universe has finite age (t_0). If that is the case, then we will only see stars that are close enough to us that their light has had sufficient time to reach us. That is, we will have an artificial cut-off a distance ct_0 away, as light will not have had enough time to reach us yet (in truth this cut-off time is probably lower than that as we would have to account for star formation etc). A finite time puts a limit on what we can observe in the universe. The simple fact that the night sky is dark should convince us that the universe is of finite length and/or is of finite age.

Cosmological Redshifts

Edwin Hubble measured the radial velocity of bright galaxies via spectroscopy. He observed that all galaxies have redshifted spectral lines, which implies that they are all moving away from us. Coincidentally, this is another possible solution to Olbers's paradox. If the light is sufficiently redshifted, it will be outside the visible spectrum and we wouldn't see it in the night sky.

The redshift of an object is defined as

$$z = \frac{(\lambda - \lambda_0)}{\lambda_0} = \frac{\Delta\lambda}{\lambda_0}$$

If this change in wavelength is due to a velocity shift, then

$$\frac{\Delta\lambda}{\lambda_0} = \frac{v}{c}$$

$$v = cz$$

Notice that if an object is moving towards us, $\Delta\lambda$ is negative and hence z is negative. Hubble found a strong correlation between distance and redshift - almost all external galaxies apart from those very close to us are moving away from us.

$$v \propto d$$
$$cz \propto d$$
$$cz = H_0 d$$

H_0 is the Hubble constant, currently measured as H_0 = 71±1 km/sec/Mpc. The units are km/sec/Mpc, primarily so distance and velocity can be expressed in convenient units. The subscript zero here indicates that this is the current value of H_0. H is a constant in the sense that it is the same at every place, but can change with time. This relation between the speed of a galaxy and the distance is known as *Hubble's law*.

The units of H_0 are (km/sec)/Mpc (dimension T^{-1}).

This finding has two important consequences.

1. In the nearby Universe, the redshift can now supply the distance to an object.

$$d = \frac{cz}{H_0}$$

2. All galaxies (apart from the Local Group) appear to have redshifts. Therefore, the universe is expanding.

The Hubble redshift law is a sign of uniform expansion (but we'll come back to that in more detail later when we talk about Newtonian Cosmology).

Consider a grid of points on a rubber sheet, each separated by R. Stretching the sheet to double its size in time t will mean that each adjacent point is now $2R$ apart.

Points initially separated by $2R$ will now be separated by $4R$ in time t – as is observed in our universe. Also, each point will see all other points expanding away from it. The Scaling factor, R, represents the relative distances between points in a co-ordinate system.

The redshift measures not a Doppler shift, but the expansion of space itself. In an expanding Universe, the *wavelengths of photons emitted will depend on when they were emitted.*

$$\frac{\lambda(\text{observed})}{\lambda(\text{emitted})} = \frac{R(\text{now})}{R(\text{then})}$$

$$\frac{\lambda(\text{observed}) - \lambda(\text{emitted})}{\lambda(\text{emitted})} = \frac{R(\text{now})}{R(\text{then})} - 1 = z$$

Therefore $z + 1$ measures the fractional increase in the size of the universe since emission of the photons.

As stated above (point 1.), Hubble's law is useful in determining distances to distant objects. Normally, it is reserved for objects that are far enough away from us that their velocities relative to our vantage point is governed by the expansion of the universe. Normally, this is referred to as the *Hubble flow*. Local group galaxies are too close for the velocities to be solely determined by the expansion of the universe, and so an extra Δv component is required to calculate their actual radial velocity. This Δv term is independent of the distance, d, to the galaxy, so they are not in the Hubble flow. At large distances, Δv is significantly smaller than $H_0 d$ and so the Δv term only adds a small fractional error in our calculation of H_0.

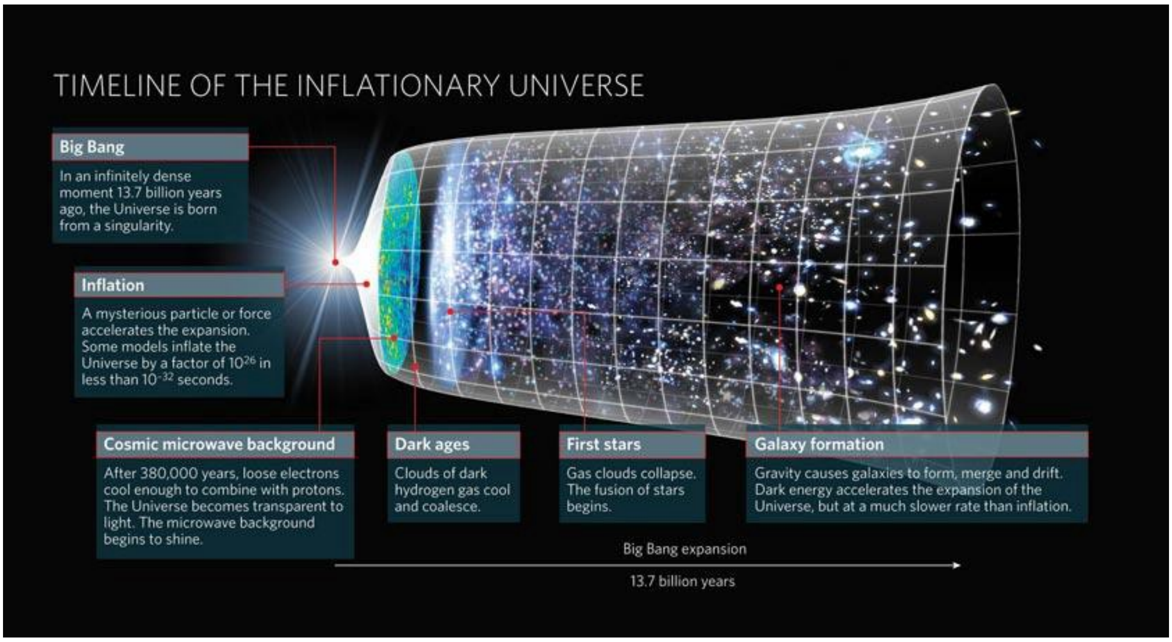
The Big Bang Cosmology

As stated previously, cosmology is the study of the formation and/or evolution of the Universe.

Any cosmology must then explain three observations:
(a) It is homogeneous - matter and radiation are spread out uniformly.
(b) It is isotropic - galaxies appear the same no matter where we look.
(c) The universe is expanding.

The Big Bang

The universe began in a very condensed and very hot state, then expanded and cooled over time. This hot dense era was named the big bang by Fred Hoyle, who actually came up with the term in a BBC radio interview as a means to try and ridicule the theory as he was an ardent steady-state cosmologist. The derisory name stuck and outlived many steady-state theories. Perhaps this is a reason why the TV show is named the same, or I'm reading too much into things…



Cartoon of the big bang theory of an inflationary universe.

The big-bang theory rests on four pieces of evidence.

(i) The Hubble expansion
This implies that the average density of the Universe has been decreasing, unless matter is somehow created out of nothing.

(ii) Galaxy Evolution
If the Universe is of infinite age, then galaxies should have remained the same over the observable Universe. Distant galaxies beyond $z \approx 0.5$ show significantly different morphologies and star formation rates than in the present Universe. This implies that the intrinsic structure of galaxies has changed over time, so the Universe cannot be infinitely old. We will come back to these at a later stage.

(iii) The Hydrogen-Helium Ratio

If all matter in the Universe was closer together in the past, as it must be for an adiabatic expansion, the Universe must have been hotter at that time.

In the very early universe ($T > 10^9$ K), protons had high enough energies to overcome their mutual repulsion and combined with neutrons to produce He nuclei.

As the universe continued to expand, a point was reached where the plasma was too cool and rarefied to allow further fusion. The Big Bang predicts that He should make $\approx$ 25% by mass of the universe, with very little in heavier elements. This is exactly what is observed in old stars of age $\approx 10^{10}$ years.

(iv) The Microwave Background Radiation

The basic premise (or implication) of the big bang theory is that we should see cosmic background radiation. This results from the fact that.

- At some stage in the past the temperature would have been above 3000K and the atoms would have been fully ionised. The matter would have been a plasma.
- The free electrons would have been effective at scattering and so would have resulted in continuum opacity for any radiation present.
- The radiation then would have remained in equilibrium with the matter and would have had a blackbody spectrum at the temperature of the matter.
- As the matter cooled, the radiation cooled.
- At some point the temperature drops below 3000K and electrons and nuclei can recombine (*the era of recombination*).
- The neutral atoms that formed are now inefficient at absorbing the radiation (as it only occurs at discrete wavelengths)
- Therefore, the universe ends up being transparent to radiation.
- The radiation and matter are no longer interacting, and we say they are *decoupled*.
- When we look at the universe in any direction then we should see the last photons emitted by the plasma on decoupling at the epoch of recombination.
- As the universe expands, this will be redshifted.
- It will still have a blackbody spectrum, but it will appear cooler.

To put it another way:

$$\lambda \propto R$$

$$\lambda T = \text{constant}$$

$$T \propto R^{-1}$$

Thus, the temperature of an expanding universe falls. In all directions, the isotropic radiation field will have a black body temperature of 2.7K. It is <u>precisely fit</u> by the black-body spectrum formula:

$$I(\nu)d\nu = \frac{2h\nu^3}{c^2}\frac{1}{e^{h\nu/kT} - 1}d\nu$$

From Wien's Law this appears strongest in the microwave region, and is known as the *3K microwave background* as the peak in spectral radiance is at 160.23GHz.

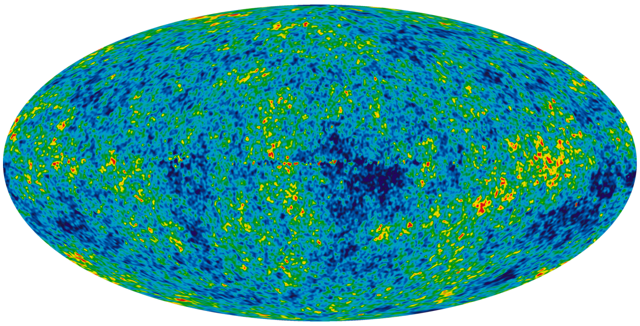
Remember that in the early hot universe, photons and electrons must have been strongly coupled by scattering. As the universe expanded it cooled, then we get a point at which the electrons combined with protons to form Hydrogen. Photons were free to travel and the universe became transparent. As stated previously, de-coupling occurs at T = 3000K. Therefore, we can show that

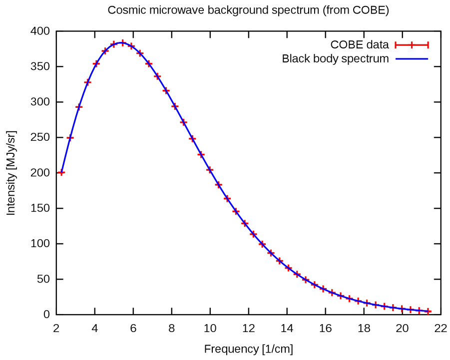
$$\frac{T_{\text{then}}}{T_{\text{now}}} = \frac{R_{\text{now}}}{R_{\text{then}}} = z + 1$$

This implies that the microwave background lies at a redshift of $z \approx 1000$, and that decoupling occurred when the universe was only 1/1000th of its present size.

Early theory on the temperature of the background temperature in the 1940's placed it at about 5K, however, this relied on a more inaccurate value for parameters such as Hubble's constant. Not much was done observationally with this information at the time, as the equipment needed to measure the signal probably wouldn't have been good enough and most people still believed in a steady-state universe so there was little appetite to look for it.

In the 1960's however, two physicists working for Bell Telephone laboratories (Penzias and Wilson) were using a telescope for both astronomy and communications and were trying to work out all sources of noise in their data. After taking large steps to account for instrumental and human error, they concluded that the radiation was coming from everywhere in the sky, although they only had measurements at one wavelength and couldn't confirm the shape of the spectrum. Meanwhile, in Princeton, the group that made the prediction in the 1940's were redrawing their prediction. Someone that knew both groups put them in contact with each other, the Princeton group made observations at a different wavelength and showed that the radiation was consistent with a blackbody spectrum. Thus, as Nobel prize was won… and the steady-state theories were no more (for the most part).





Left: The CMB temperature fluctuations over the celestial sphere.
Right: The measured CMB Spectrum (red) against the blackbody spectrum (blue)

Precise measurements of the cosmic microwave background (CMB) are crucial to cosmology. While looking at the CMB, it is nearly uniform in all directions. You will notice some anisotropies in the CMB which are expected if you have a fairly uniformly distributed hot gas that has expanded to the size of the universe that we observe, that is, the anisotropies match small thermal variations due to quantum fluctuations of matter in a small space prior to expanding out to what we now observe. Many models could explain the uniformity off the CMB, but the big bang based theories are currently the only theories that can explain these anisotropies.

Lecture 6 - The High-energy Universe

Learning Objectives

In this lecture, we are going to look at the high energy Universe and some interesting features, events, and galaxy types in the Universe. This is a basic introduction to the topic. By the end of this section, you will know:

- The different types of Active Galactic Nuclei
    - Radio galaxies
    - Seyfert galaxies
    - BL Lac objects
    - Quasi-stellar objects
- The unified model to explain these objects and associated features
- Gamma ray bursts and their possible sources

Introduction

The majority of normal matter lies in normal stars in normal galaxies. However, the detection of high-energy photons (X-rays and gamma-rays) indicate that extreme environments can exist.

On a small scale, these include neutron stars and stellar-mass black holes. We also see other phenomena such as Active Galactic Nuclei and Gamma ray bursts. For years it was though that these features were unique, but with further study it was realised that many of them have similar origins.

AGN and Quasars

Some galaxies appear to have extremely luminous centres. These are known as Active Galactic Nuclei (AGN), and typically show strong emission lines of Hydrogen and other elements. The activity seems to be within the nucleus of the galaxy or at least is driven by activity within the nucleus.

Radio-Loud AGN

Many strong radio emitting sources are not found within our galaxy. When we observe a strong source in radio it is often useful to look at other wavelengths (such as visible) to see if the source corresponds to some visible feature that we can identify. In the past this was difficult to achieve given the angular resolution of radio telescope. Interferometry has aided in the improvement in isolating the source of these Radio-loud regions.

The radio energy output by these galaxies is huge, often 10^6 times that of the total output of a normal galaxy. In these galaxies, radio imaging shows emission coming from large jets emitted from the centre of the galaxy. The source of this radio emission is caused by synchrotron emission. This is confirmed by looking at the radio spectrum of the galaxy which follows a power law consistent with synchrotron radiation.

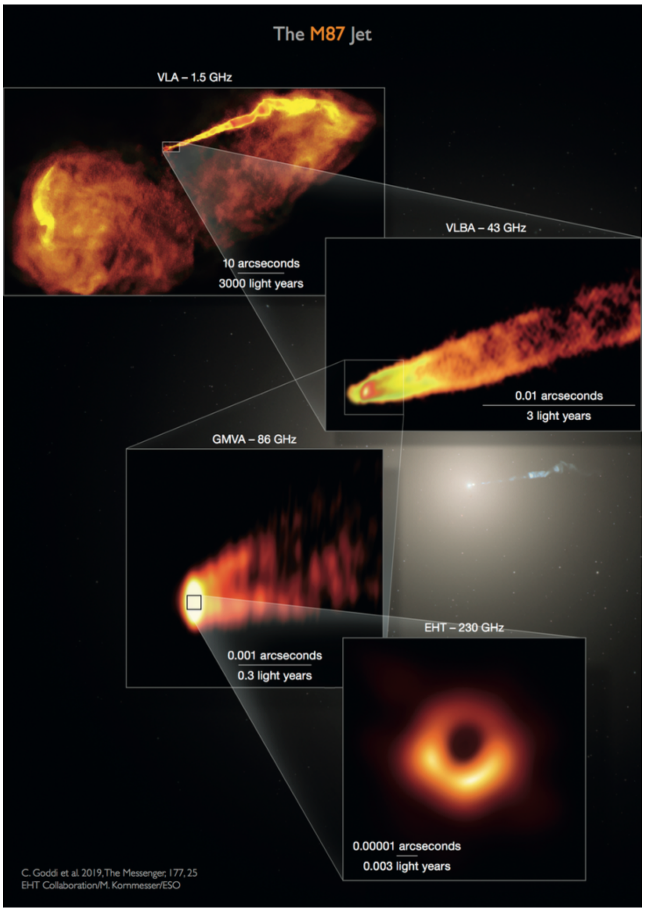
Furthermore, the radiation is polarised, which is another signature of synchrotron radiation. Synchrotron radiation is a result of high velocity electrons spiralling around magnetic fields. You may have heard of the term before due to the fact that it is emitted from synchrotrons, which are a type of particle accelerator.

The synchrotron spectrum for such stars depends on the energy distribution of the electrons. As such, as the electrons radiate, they lose energy, and the spectrum will evolve over time from these sources. By studying the synchrotron spectrum of a radio galaxy, we can estate how long it has been radiating.

Typically, radio galaxies have a small source near the centre of the galaxy and then two large sources (*radio lobes*) far beyond the extent of the galaxy in visible wavelengths. The radio lobes may be separated by 10Mpc and can be as wide as 1Mpc. Higher resolution observations will also show *radio jets* pointing from the centre of the galaxy towards the larger radio lobes.

The density of the radio lobes is on the order of $10^{-3}\mathrm{e}^{-}\ \mathrm{cm}^{-3}$. The low density ensures that collisions are infrequent and so the energy is lost due to the synchrotron radiation with the electrons. The magnetic field strength in the lobes is of the order of $10 - 100\mu\mathrm{G}$.

The radio jets are highly collimated and have estimated velocities in the range of 10^3 - 10^5 km/s. The density of electrons in the jets is about $10^{-2}\mathrm{e}^{-}\ \mathrm{cm}^{-3}$ and the magnetic field is at the same strength as the lobes. The mechanism suggested for the formation of the jets is that material is outflowing into a region with denser material surrounding it and confining it. The denser material has a hole in it and the material flowing out from the source takes the path of least resistance forming the jet. The material shoots out like the effect of a nozzle. The denser material surrounding the source needs to have material or something acting upon it to keep it from expanding with the outflowing material. Also, only one jet is likely to form at a time, with a jet coming from both sides to support the observations of the two lobes.



The M87 radio jet and lobes. The Lobes are seen in the top image, then the close up of the jet before looking at the strong radio source at the base of the jet. The final image is the well-known Event Horizons Telescope image of the black hole at the centre of the radio source powering the jet.

Seyfert galaxies

These appear as normal spiral galaxies but with very bright and compact cores. They form about 1% of known spirals. The nuclei strongly dominate the total light from the galaxy. On a short exposure the look like a star with a blurry patch around them while at longer exposure they look like a spiral galaxy. In comparison a regular spiral galaxy will not look like a star on long or short exposure. Therefore, we are more than likely not resolving the nuclei in most Seyferts. The blurry patch around the nucleus resembles a spiral disk, hence the belief that they are spiral galaxies with a bright nucleus.

There are two main types of Seyferts, classified based upon the emission lines in their spectra.

- Type II Seyferts only have narrow emission lines (both permitted and forbidden).
- Type I Seyferts also have broad hydrogen emission lines (e.g., H I) as well as forbidden lines (e.g. O III).

Forbidden lines: these are spectral lines due to electron transitions not traditionally allowed by quantum mechanics selection rules, but with a small probability of occurring.

The width of lines in Type I Seyfert galaxies is more than ten times that of normal galaxies. If this is due to thermal broadening it implies a temperature of 10^7K. The broad lines appear to originate above the accretion disk of a supermassive black hole, while the narrow lines occur beyond the broad line region of the accretion disk. Type I Seyferts are very bright in UV and X-ray as well as the visible light from their cores. Type II Seyferts appear bright when viewed at IR wavelengths.

BL Lac Objects

These appear as almost starlike, with very little structure. Spectra appear almost featureless, with featureless non thermal emission continuum over the whole EM range. This means that distances are very uncertain!

They are characterised by rapid and large amplitude flux variability and significant optical polarisation. Therefore, they were thought of as variable stars. The name for these objects comes from the prototype (BL Lacertae) which was thought to be an irregular variable star in the Milky Way (hence the variable star definition - see the origin of variable star names in the RR Lyrae and Cepheids part of the Cosmological Distance Scales Lecture).

QSO's (Quasi-Stellar-Objects)

These are often referred to as quasars and are the most famous form of AGN known. In an attempt to observe optical counterparts of radio sources, early attempts used lunar occultation (basically observe the moon passing over the object and noting the times that the object disappears and reappears) to observe the sources. Using this technique led to the sources being associated with faint stars, so originally, they were thought to be radio stars, which had never been observed. Of course, this wasn't quite the source of the radio sources after jets were observed associated to them and the spectrum was analysed from the source. This is where the name is derived, as they looked like stars with radio sources, Quasi-stellar radio sources (QSR or quasar). Subsequent studies showed that some do not have radio emission, so the designation shifted to QSOs.

Appearing like stars (unresolved point sources), they show large redshifts of $z \geq 0.16$ and luminosities up to that of a large galaxy. All spectra show narrow emission lines with doppler widths of ~ 1,000 km/s, but some also show broad emission lines with widths of ~ 10,000 km/s. With one of the original radio sources, the spectrum puzzled researchers for quite some time until they realised that they were observing the Balmer series in Hydrogen shifted by these large redshifts.

For redshifts $z > 0.3$, one must use the relativistic redshift formulae. In these cases, Hubble's law as presented in previous lectures is not a useful means of finding the distance to modulus to an object.

$$z = \frac{\Delta\lambda}{\lambda_0} = \left[\frac{1 + v/c}{1 - v/c}\right]^{\frac{1}{2}} - 1$$

$$\frac{v}{c} = \frac{(z+1)^2 - 1}{(z+1)^2 + 1}$$

$$d = \frac{cz}{H_0}\left(1 + \frac{z}{2}\right)$$

The spectra of quasars show a large UV excess, which implies that they are brighter in UV than you would expect form their visual brightness. This gives us a means to identify quasars. By using space-based observations we can look across a large field-of-view to look for excess in the UV. Then spectra can be acquired (as the spectrum takes time to observe as they are faint).

High resolution imaging shows galaxies surrounding QSOs. The variability associated with QSOs means that we can put an upper limit on the size of the emitting region. The size of the emitting region is constrained by the time scale, τ, over which the energy output varies. All AGN show variability on timescales, τ, of a few weeks. For a body of variable brightness, the *maximum* radius r is then given by:

$$r \sim c\,\tau$$

As quasars vary on the timescales of a few weeks to a few years, the central energy source must be roughly a parsec or smaller. The best source of energy is an accretion disk around a super-massive black hole, with masses $\geq 10^6 M_\odot$.

As material moves around and closer to the black hole, the gravitational potential energy is converted to thermal radiation.
The Schwartzchild radius of a $10^8 M_\odot$ black hole is:

$$R = \frac{2GM}{c^2} \approx 2\mathrm{AU}$$

The potential energy for masses brought to this distance is given by

$$PE = -\frac{GMm}{R} \approx 10^{16} \quad \mathrm{J/kg}$$

As $L_\odot = 3.8 \times 10^{26}$ J/s, to produce $10^{11} L_\odot$ requires an infall of 10^{29} kg/year, or $1 M_\odot$ every 10 years.

Unified model

When we look at the various examples of AGN, we see many common features:

- large energy output
- rapid variability
- jets

This then suggests that radio galaxies, Seyferts and QSOs are all different manifestations of the same underlying phenomena.

The different types of AGN are explained in the *unified model*. The structure of an AGN is a central black hole, surrounded by a thin hot accretion disk emitting radiation at high velocities. This is surrounded in turn by a thick disk of dust and gas. Slower-moving gas above and below black hole is illuminated by the accretion disk. Accretion powers fast particle jets along the spin axis of the black hole and disks.

If we look through the thick disk, all we see are the jets and the narrow-line region. If we see the accretion disk, we also see the broad-line region and the AGN is brighter. If we look along the spin-axis, our view is dominated by jets, and we see a BL Lac object.

The amount of material in the accretion disk governs the brightness of the AGN (Seyfert or QSO).

Observational evidence includes:

1. Detection of the host galaxies around QSOs.
2. Imaging of accretion disks in nearby AGN.
3. Measurement of spectroscopic velocities in nearby AGN.

For the AGN M87 (which we have seen several times in class), $v = 550$ km/s at 20pc from the centre. This gives the black hole mass as

$$m = \frac{v^2 R}{G} = 1.4 \times 10^9 M_\odot$$

In 2017, the Event Horizon Telescope was composed of radio telescopes across the world capable of observing radio waves at $\lambda = 1.3$mm. In 2019, the processing of these data revealed the first direct image of the ‘shadow’ of the Event Horizon of the M87 black hole. We have already seen this image in previous lectures (for example talking about the supermassive black hole at the centre of the Milky Way).

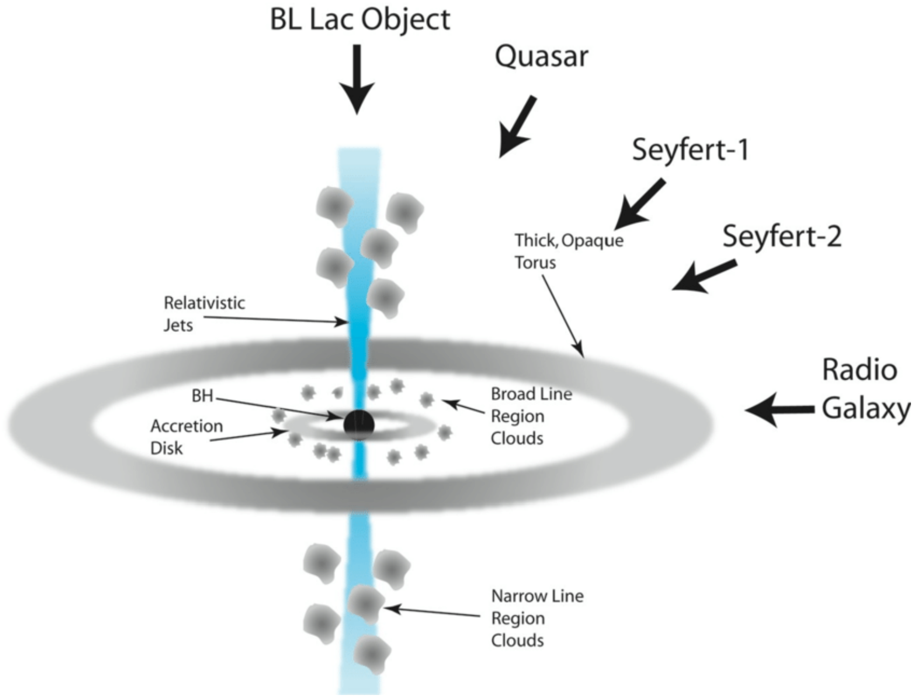


Image showing the relation between different types of observed AGN and the formation mechanism based on the unified model.

Gamma Ray Bursts

Gamma-ray bursts (GRBs) were discovered in the 1960s by satellites watching for nuclear bomb tests. They are short, intense bursts of gamma-rays with peak energies around 1 MeV.

These short bursts of high-energy photons are detected on average once per day by satellites. They appear at random positions in the sky, and last < 60 seconds down to < 1 second, but with extremely rapid variations. Hence the events occur within a scale of $c\tau \sim 3 \times 10^5$ km or smaller. Longer bursts (> 2s) are more typically found in star forming galaxies.

Finding the source of these GRBs is difficult, as gamma-ray telescopes cannot normally pinpoint sources more closely than about $1°$ on the sky. The first visible light counterparts to GRBs were found in 1997 as fading sources fainter than $m_V = 18$. They were observed by looking for a burst then pointing quickly to take an X-ray image of the region (as they have a low-energy tail of X-rays). The resolution acquired with the X-ray images is good enough to tell optical astronomers where to look for the sources. The first spectra revealed redshifted spectral lines of $z = 0.8 - 3.4$, showing these are very bright events.

On 23rd January 1999, GRB 990123 had a peak magnitude of $m_V = 8.9$, and a redshift of $z = 1.6$. Assuming isotropic emission over 4π steradians, this would imply this GRB was as luminous as the entire visible Universe! GRB080319 beat this in March 2008, reaching $m_V = 5.6$ while at a redshift of $z = 0.937$.

It was shown in 2003 that GRBs are associated with Supernovae, with spectra being obtained of the relatively close-by GRB 030329, showing a supernova spectrum.

The current model is that GRB's are caused by the most massive Type-II supernovae. In these events, the iron core immediately collapses to form a black hole of $> 4M_{\odot}$. The black hole immediately forms an accretion disk and jets within the star as the shockwave of the core collapse is still propagating through the star. When the jets burst through the surface of the star after ~ 10 seconds, if we are looking down the jet, we see a GRB. We see about one GRB for every 100, 000 supernovae, but this is likely a small fraction of the total as we fail to detect most bursts as the jet is not pointing in our direction.

GRBs may have been relatively more frequent in the early Universe as it was easier to make very massive stars as metals were absent from the gas. If the first generation of stars were all large and have now ended their lives, this explains the absence of GRBs in the Milky Way today. Gamma-rays are not scattered by dust or readily absorbed by the ISM so they can help us to probe the earliest massive stars in the Universe in the most distant galaxies.

Lecture 7 - Newtonian Cosmology

Learning Objectives

In this lecture we will cover:

1. Newtonian Cosmology as a means to estimate the age of the Universe
2. The implications of the curvature of the Universe (open, flat and closed Universes)
3. The implication for the future of the Universe
4. The density of the Universe as a measure of Universe geometry
5. Measuring the density of the Universe

We can learn a lot about the evolution and expansion of the universe by applying Newtonian gravitation. Of course, for a better indication, we need to account for general relativity, but that is outside the scope of this module.

The age of the Universe

If the universe has undergone constant expansion since the Big Bang, then the age of the universe is given simply by

$$t_0 = H_0^{-1}$$

Example: If H_0 = 71 km/s/Mpc, calculate the age of the Universe.

A constant rate of expansion, however, implies a massless universe. If deceleration $\neq 0$ then the universe is younger than we think.

Assume that there are no other forces acting in the Universe apart from gravity. The assumption of isotropy is the equivalent of saying that the Universe appears to be spherically symmetric from any point. Therefore, the spherical volume will evolve only under its own influence. The gravitational forces exerted on the volume by material outside the volume will sum vectorially to zero.

The equation of motion of a particle in the universe on the surface of this sphere then is,

$$a = \ddot{R} = -\frac{GM(R)}{R^2}$$

where $M(R)$ is the mass interior to R.

Multiplying both sides by $\dot{R}$ gives

$$\dot{R}\ddot{R} = -\frac{GM(R)}{R^2}\dot{R}dt = 0$$

Integrate with respect to t:

$$\int_0^t \dot{R}\,\ddot{R}dt + \int_0^t \frac{GM(R)}{R^2}\dot{R}dt = 0$$

To solve this, it is useful to note that

$$\frac{d}{dt}\frac{\dot{R}^2}{2} = 2\frac{\dot{R}}{2}\ddot{R} = \dot{R}\ddot{R}$$

Then the integration gives

$$\frac{\dot{R}^2}{2} - \frac{GM(R)}{R} = k$$

Where k is referred to as the curvature constant. If we assume $k = 0$, this gives us

$$\dot{R}R^{1/2} = (2GM)^{1/2}$$

Integrating this with respect to t gives

$$\frac{2}{3}R^{3/2} = (2GM)^{1/2}t$$

Rearrange with respect to R gives

$$R = \left(\frac{9GM}{2}\right)^{1/3} t^{2/3}$$

Differentiating R with respect to time we then get

$$\dot{R} = \left(\frac{9GM}{2}\right)^{1/3} \left(\frac{2}{3}\right) t^{-1/3}$$

Using the equation for R as above, we can rewrite this as

$$\dot{R} = Rt^{-2/3} \left(\frac{2}{3}\right) t^{-1/3}$$

$$= \left(\frac{2}{3}\right)\left(\frac{R}{t}\right)$$

Rearranging this expression with respect to t gives us

$$t = \left(\frac{2}{3}\right)\left(\frac{R}{\dot{R}}\right)$$

At the present time, t_0, the expansion rate now will be given by

$$\frac{\dot{R_0}^2}{R_0} = H_0$$

So

$$t_0 = \frac{2}{3}\frac{1}{H_0}$$

For $H_0 = 71$ km/sec/Mpc, $t_0 \approx 1.0 \times 10^{10}$ years. Note that this is the age of the Universe if the average density of the Universe is greater than some critical density (see next section). If the average density exceeds this critical density then the Universe is younger than this, if the density is less, then the Universe is older than this. We shall see that, this expression gives us basically the minimum age of the Universe.

The Density of the Universe (Derivation)

Remember that

$$\frac{\dot{R}^2}{2} - \frac{GM(R)}{R} = k$$

Assuming uniform density, ρ, then the mass can be represented by

$$M(R) = \frac{4}{3}\pi R^3 \rho$$

Therefore,

$$\frac{\dot{R}^2}{2} - G\frac{4}{3}\pi R^2\rho \;=\; \mathrm{k}$$

$$\frac{\dot{R}^2}{R^2} - \frac{8\pi G\rho}{3} = \frac{2k}{R^2}$$

This can be rearranged as

$$\frac{\dot{R}}{R} = \left(\frac{8}{3}\pi G\rho + \frac{2k}{R^2}\right)^{\frac{1}{2}}$$

This equation is the Newtonian, non-relativistic form of the Friedmann Equation. They do not specify the nature of the density, ρ, nor do they include any particle interactions apart from gravitational attraction. Inclusion of particle interactions will result in a pressure term being included in the above equation. As such this is sometimes referred to as a 'pressure-less universe'.

In the equation above, as $R \geq 0$, and the universe is expanding ($(dR/dt) > 0$), then the end state depends on ρ and k. Therefore, we should look to see how this varies with varying values of k as this will tell us whether the Universe is open or closed. Note that if ρ is not zero then $\ddot{R}$ cannot be zero. This implies that a universe with matter cannot be static. It must be expanding or contracting.

1. $k > 0$
   The right-hand side of the equation is always > 0, and the universe expands forever. There is not enough density for gravitational attraction to stop expansion.
2. $k < 0$
   The right-hand side eventually becomes < 0 as ρ becomes smaller, and the universe collapses (the "big crunch"). This is described as a *closed* Universe or a gravitationally bound Universe.
3. $k = 0$
   In this critical case, the expansion slows down but only stops after an infinite time has taken place. This $k = 0$ condition can be used to express the critical density in terms of the present value of the Hubble parameter.

When $k = 0$ we get:

$$\frac{\dot{R}}{R} = \left(\frac{8}{3}\pi G\rho\right)^{\frac{1}{2}}$$

$$H = \left(\frac{8}{3}\pi G\rho\right)^{\frac{1}{2}}$$

Therefore, for a critically open Newtonian (*flat*) Universe the critical density can be established with,

$$\rho_c = \frac{3H_0^2}{8\pi G} = 9.5 \times 10^{-27}\mathrm{kg/m^3}$$

Often it is useful to express the measured density of the Universe by the density parameter, Ω, as

$$\Omega = \frac{\rho_0}{\rho_c}$$

Where $\rho_0 = \rho(t_0)$, or the current density. From observing the total mass in stars and ISM in local galaxies, $\Omega \approx 0.03$. The density parameter determines the overall geometry of the Universe. When $\Omega = 1$ then $\rho_0 = \rho_c$ and the geometry of the Universe is flat (or Euclidean). Therefore, the density parameter, Ω, ultimately governs the curvature of the Universe:

- Negative ($\Omega < 1$)
- Positive ($\Omega > 1$)
- Flat ($\Omega = 1$)

The table below gives a summary of the relations between the curvature constant, k, and the density parameter, Ω, to the type of Universe, the curvature and the extent of the Universe.

Type	$\boldsymbol{k}$	$\boldsymbol{\Omega}$	Curvature	Extent
Boundary	0	1	Flat	Infinite
Closed	+1	>1	Positive	Finite
Open	-1	<1	Negative	Infinite

Density of the Universe in Context

There are two observations that imply that the density due to matter may not be the only story.

(a) The age of the Universe:

If $0 \le \Omega \le 1$ then the Universe is expanding forever, and the age of the Universe must lie in the range:

$$\frac{2}{3}\frac{1}{H_0} < t_0 < \frac{1}{H_0}$$

Specifically, this implies that the current age of the Universe is $t_0 = 10 \rightarrow 13 \times 10^9$ years. This is a similar issue if $\Omega > 1$, which implies that the age of the Universe is less than $\frac{2}{3}\frac{1}{H_0}$, although this is unlikely given our estimates of the density of the Universe.

However, the oldest stars seen in the Milky Way are ~ 13×10^9 years old, which we can establish by isochrone fitting of clusters. Furthermore, these already contain some heavy elements, implying that they are not the first generation of stars (which would only consist of Hydrogen and Helium). This could lead to an “age paradox”.

Take HD 140283 (often dubbed the ‘Methuselah star’) which is ~ 62 pc from earth as an example. It is one of the oldest observed stars in the Universe at an estimated age of 14.46 ± 0.8 billion years. Now this is larger than the age of the Universe estimates we have, although the uncertainties in the age estimate mean that it may have been one of the first stars to form after the Big Bang. However, the metallicity of this star implies that it is not a firs generation star and so formed after some first-generation star evolved and underwent a supernova to produce the elements besides H and He observed in HD 140283.

(b) The microwave background:

The microwave background is not completely smooth but shows small temperature fluctuations of size $\Delta T/T \approx 10^{-5}$.

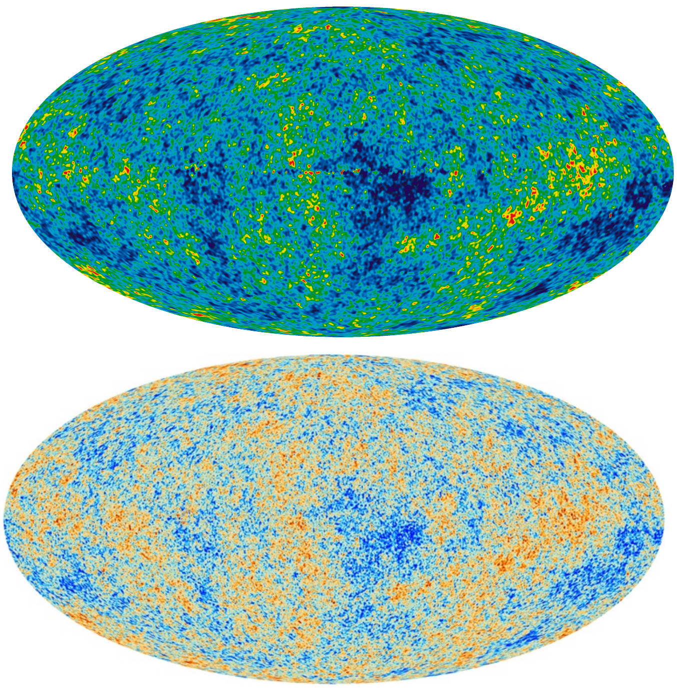
These small fluctuations were caused by sound waves in the early Universe in the pre recombination plasma and are the seeds from which galaxy superclusters formed.

The apparent size of the fluctuations gives the geometry of the Universe, as an open Universe will result in diverging photons paths and hence a smaller angular size for the fluctuations.

The Wilkinson Microwave Anisotropy Probe (WMAP) measured this in 2001 – 2006, and the ESA Planck mission measured this in 2009 – 2013:

$$\Omega = 1.00 \pm 0.02$$

Remeber that estimates of Ω from the circular motion of spiral galaxies is $\Omega \sim 0.03$, and elliptical galaxies give the same estimate for Ω. The contribution to Ω from relativistic particles (photons and neutrinos) is basically negligible at $\Omega \approx 9.24 \times 10^{-5}$, so observed value of $\Omega = 1.00$ has to have contributions from something else. We will discuss what this contribution is in the final lecture.



Anisotropies in the cosmic microwave background. These anisotropies are due to very small temperature fluctuations and can give an indication of the early Universe and how that resulted in the observed Universe we see. Measurements of the density of the Universe can be made to give an indication of the likely shape and possible future of the Universe. The top plot is from the WMAP probe while the lower plot is from the Planck mission. Planck had a higher resolution and sensitivity than WMAP.

Lecture 8 – The Dark Universe

Learning Outcomes

In this lecture we will cover:

4. Three pieces of evidence for dark matter.
5. Theories on what dark matter actually is.
6. The concept of dark energy and where that arises.
7. Summary of our current estimates on their contributions to density in the Universe.
8. Brief discussion on the implications of this with regards to the fate of the Universe.

Dark Matter

When we look at a galaxy, we see the most luminous objects. However, not all of the galaxy may be luminous and there may be other objects within the galaxy that are hard to detect. For example, extrasolar planets will be difficult to detect (though there are several methods to detect their effects on host stars), and blackholes will be difficult to detect (though we may be able to see their effects on nearby stars or we may be able to detect accretion disks about the blackhole, e.g., in the case of AGNs). The only way to trace out the total mass, whether that is a luminous object or a non-luminous (dark) object, in a galaxy will be to study the gravitational effects of the luminous and non-luminous matter. The easiest way to do this in a galaxy is by tracing out the rotation curves of the galaxy. Of course, we will have to look at certain galaxies types to analyse this. Luckily, we've already encounter rotation curves when looking at determining the structure of the Milky Way.

(a) Galactic Rotation curves

If you remember back to the content on the Milky Way (Lecture 3) we discussed how to measure rotation within the galaxy. We are able to determine the rotational at specific radial distances from the centre by observing either open clusters of stars in the disk or by looking at the H I regions in the interstellar medium with the 21cm line. For other spiral galaxies, we can use a similar approach and search for Doppler shifts in spectral lines. When we do that, we can produce a *galactic rotation curve*, which is a plot of the orbital velocity (y-axis) as a function of radial distance (x-axis) from the galactic centre.

Spiral galaxies appear to be centrally concentrated. Hence, far from the centre, stars should follow Kepler's law and we will find that

$$v(r) \propto \frac{1}{\sqrt{r}}$$

Remember that for optical and radio spectroscopy of Spiral galaxies the Doppler shift measures the component $v(r)\sin i$, where i here refers to the inclination (which you should have seen earlier in this module). Assuming the disk is circular, $\sin i$ can be determined from the inspection of the major and minor axes of the disk as projected

onto the plane of the sky. Thus, $v(r)$ can be determined. It appears that at large r, $v(r) \simeq$ constant! It remains as such as far out as we see luminous material. This immediately tells us that the mass does not fall off as fast as the luminosity. In many galaxies, the rotation curves are still flat at radii where the interstellar medium can no longer be detected.
Galactic rotation curves, then were seen as evidence that there must be some form of non-luminous matter in galaxies. Normally, this non-luminous matter *dark matter*. (Note the original term came from the Swiss astrophysicist Fritz Zwicky while doing work on the Coma cluster. He gave the unseen mass the term 'dunkle Materie' – which I think is much better. Coincidentally, the idea of 'dark bodies' in the Milky Way was postulated by our own Lord Kelvin back in the 1880's as well).

Now, we may ask the question, where exactly does this dark matter reside in the galaxy? In a spiral galaxy, we may say that exists within the disk. However, theoretical models and simulations show that if the matter did exist solely in the disk, then the mass would gather the disk into a large bar which would be unstable. The galaxy will be more stable if the matter is arranged in a spherical distribution about the disk. This would place the dark matter in the halo around the galaxy. Remember the halo itself is not a ring but a spherically symmetric mass distribution (in which globular clusters can be found).

Using the rotation curve, we can establish the mass distribution in the halo. If we take the mass within radius, r, as $M(r)$, the acceleration due to gravity must provide the acceleration due to the circular orbit. Therefore, the velocity, $v(r)$, is related to the mass as

$$\frac{GM(r)}{r^2} = \frac{v^2(r)}{r} = \frac{v_0^2}{r}$$

If we assume $v(r)$ is some constant value, v_0. Solving for $M(r)$ gives us

$$M(r) = \frac{rv_0^2}{G} \propto r$$

The derivative of this with respect to r simply gives

$$\frac{dM}{dr} = \frac{v_0^2}{G}$$

If you remember back to stellar structure (which will be covered more fully by yours truly in PHY3003), the mass of a thin spherical shell is given

$$dM(r) = 4\pi r^2 \rho(r) dr$$

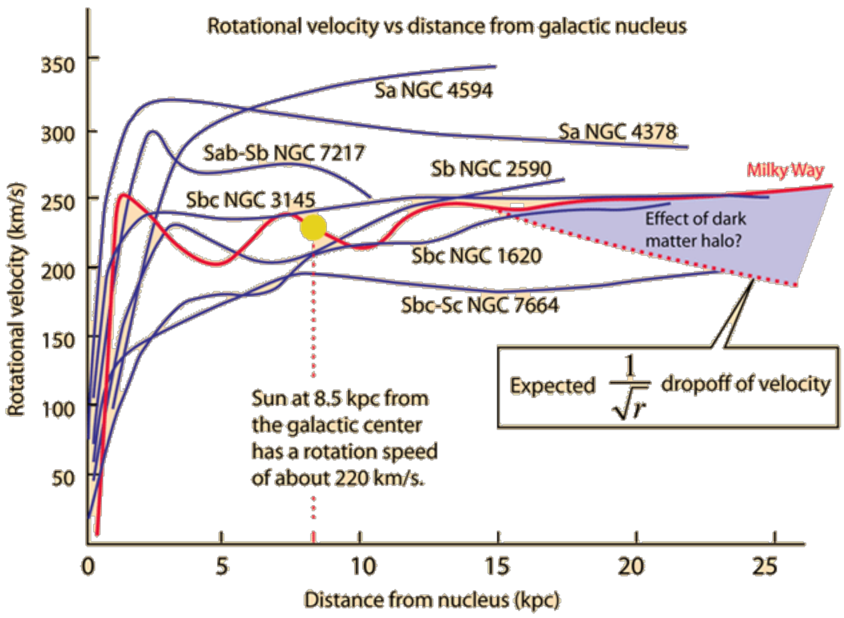
Now if we equate the two expressions, we get

$$4\pi r^2 \rho(r) = \frac{v_0^2}{G}$$

$$\rho(r) = \frac{v_0^2}{4\pi G r^2}$$

The density of the halo falls off as $1/r^2$ which is nowhere near as fast as the exponential fall off in light within the disk. This might seem that there is not a lot of matter further out. However, if you divide up the galaxy into spherical shells with thickness dr, then the volume of each shell is given by $4\pi r^2 dr$. This implies that the mass of each shell is constant. Further out in the rotation curve, significant amounts of matter are added to the galaxy. This would follow the observed pattern that we see with the rotation curves remaining basically flat to the edge of the galaxy.

So, the takeaway here is that a constant velocity means significant amounts of matter is found out to the very edge of galaxies, much more than is visible in optical imaging. This implies the existence of non-luminous matter (i.e., dark matter).



Examples of rotation curves for several galaxies. The effect of dark matter on the rotation curve of the Milky Way is shown with a dashed red line.

(b) Galaxy Clusters

On a larger scale, measuring the speeds of galaxies within a cluster can give us a value for the total mass of the cluster. A galaxy of mass m_g moves within a cluster of galaxies of total mass M_c a distance r from the centre. The kinetic energy and potential energy of the galaxy is given by,

$$KE = \frac{1}{2} m_g v^2$$

$$PE = -\frac{GM_c m_g}{r}$$

For a stable system, again we can state that the acceleration due to gravity must provide the acceleration due to the circular orbit, so:

$$\frac{v^2}{r} = \frac{GM_c}{r^2}$$

If we multiply through by r and multiple both sides by m_g, we get

$$m_g v^2 = \frac{GM_c m_g}{r}$$

Which we can see relates the kinetic energy and potential energy of the galaxy cluster as

$$2KE = |PE|$$

This is a form of the *virial theorem* (which relates the time averaged total kinetic energy of a stable system bound by potential forces – in this case gravity). For all clusters observed, we find that $KE > \frac{1}{2}|PE|$. As clusters must be stable to have survived > 10^{10} years, this implies that dark matter is found in clusters as well.

By measuring the speeds of galaxies in clusters, we can estimate the total mass of the cluster. The inferred mass appears much larger than the implied mass from the visible constituents. It can be useful to look at the mass-to-light ratio. This is just the ratio of the mass (expressed in solar masses) to the luminosity (again in terms of solar luminosity units). You would expect the mass-to-light ratio of solar-like stars then to be one. For a spiral galaxy the mass-to-light ratio will be 1:3 near the centre and 20:1 to at the edge of the disk. Therefore, the mass-to-light ratio increases radially. If M is the mass inferred from gravity, and L is the mass inferred from the luminosity (i.e., from stars), then in clusters the mass-to-light ratio is

$$\frac{M}{L} = 3 \rightarrow 50$$

(c) Gravitational lensing

Einstein showed that gravitational fields can deflect light in the general theory of relativity. Proof that his general theory was correct was obtained in 1919 by imaging the small shift in the position of stars near the Sun during a total solar eclipse.

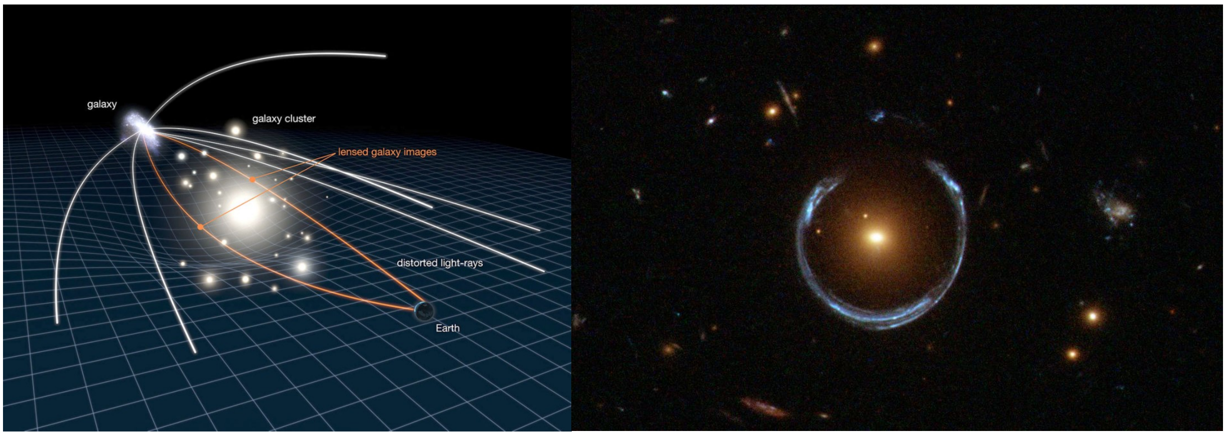
If a photon passes a distance d from a body of mass, M, Einstein showed it will be deflected by an angle α given by

$$\alpha = \frac{4GM}{c^2 d}$$

When we observe distant galaxies behind closer clusters of galaxies, their light is distorted and refracted by the galaxy cluster gravitational fields. The angle α and distance d can be found by measuring the redshift of the background and foreground galaxies and the apparent angular separation of the images. With this method we can measure the mass of galactic clusters and compare that to the total observed mass of the cluster (in the form of luminous matter). Again, this method yields $M/L > 1$, and generally agrees with virial theorem calculations performed on the same galaxies.

Overall, there appears to be ~ 10 times as much dark matter as normal (baryonic) matter. Hence best estimates give:

$$\Omega_{\text{matter}}(\text{dark} + \text{light}) \simeq 0.25$$



The image on the left shows gravitational lensing works, with light from the galaxy behind some foreground galaxy bent by the effects of gravity on the way to an observer on Earth. The image on the right shows an 'Einstein ring'. Light from a galaxy is bent almost completely into a ring shape by the gravity of the large galaxy in front of it.

(d) What is dark matter?

There have been a few theories postulated as to what dark matter actually is. The short answer is, we still do not know exactly what it is. It is somewhat difficult to verify what it is based solely on the gravitational effects it has with other objects. There have been some efforts to try and detect the dark matter directly as well as some experiments that have tried to find indirect evidence of the nature of dark matter (aside from the effects of gravity).

In essence dark matter is dark as it does not appear to interact with the electromagnetic field through absorption, emission or scattering. Initial theories suggested that galactic halos may be filled with faint, old red stars that would have masses less than $1M_{\odot}$. There is even some direct evidence of these bodies in the halos of nearby galaxies with a mass-to-light ratio of 20:1, so, in theory, they could explain the mass-to-light ratio at the edge of spiral galaxies. However, the mass-to-light ratio actually increases past the edge of spiral galaxies and old, faint red stars will not be able to account for the dark matter observed in galaxy clusters. Anyway, this theory lost traction when studies showed that nuclear burning material does not make up a significant portion of the halo.

Similar ideas suggested that the mass could come from non-luminous Jupiter sized objects, neutron stars and other faint objects (often referred to as massive compact halo objects or MACHOs). However, evidence suggests that dark matter is not baryonic in nature. For example, if there were more baryons present, we would expect more heavier elements generated during the big bang (which goes against the ratios of elements we see in older stars) and also gravitational microlensing in the Milky Way suggests that dark matter is not in the form of MACHOs.

Given that baryonic matter has been basically ruled out of the running for dark matter, current theories try to categorise dark matter as various (possibly hypothetical) non-baryonic particles such as sterile neutrinos and weakly interacting massive particles (WIMPs – I don't make the names of these things...). Various studies have tried to look into the validity of these sources for dark matter by looking for weak lensing or in-direct evidence such as annihilation between interacting particles. Some instruments have limited the possible particles that are likely to make up dark matter though work continues on what they actually are. Particle colliders are also being employed as a means to shed light on this problem. There is a Nobel prize waiting for the any one of you that solves this question, so get working on it.

Dark Energy

The most direct method of measuring the evolution of the universe is by measuring H as a function of lookback time. Remember that H_0 is referred to as Hubble's constant, but it is really just a constant at a given time in the Universe. However, that does not necessarily mean that this value is static, and it is often good to define this value generally as H at some given time. Often the use of "Hubble's constant" as a term is a bit of an unfortunate misnomer which has allowed to propagate in popular lexicon (I myself have fallen victim of the usage). The best terminology to avoid confusion is to represent H as the "Hubble parameter". If we look at far away objects (those with a high redshift), we are looking back to an earlier time period of the Universe, possibly with a different value for the Hubble parameter. The best method of measuring H as a function of time is by using Type Ia supernovae.

If we take our equation for the distance modulus, $(m - M)_{\lambda}$, and rearrange it in terms of the apparent magnitude, m_{λ}, we have:

$$m_{\lambda} = 5\, log\, d - 5 + M_{\lambda}$$

Then if we take the distance, d, from Hubble's law (remember to be careful with high redshifts!), that is,

$$d = \frac{cz}{H}$$

We can combine the two equations to get the apparent magnitude in terms of the Hubble parameter, H, and the cosmological redshift, z, which gives us

$$m_\lambda = 5\, log\left(\frac{cz}{H}\right) - 5 + M_\lambda$$

$$m_\lambda = 5\, log\, z + 5\, log\frac{c}{H} - 5\ + M_\lambda$$

If H is roughly constant, then plotting m_λ verses $\log z$ should result in a straight line of slope 5. If H was larger in the past, then it should curve downwards.

What we see is that the universe was travelling slower in the past than it is now, which suggests that the expansion of the Universe is accelerating. Subsequently, this implies the existence of a long-range repulsive force that can overcome gravity. Without the inclusion of a new form of energy we cannot explain how the Universe expansion is accelerating. This discovery was seen in the late 1990's with supernova observations leading to yet more Nobel prizes.

The strength of this force is governed by the *cosmological constant*, Λ, also called *dark energy*. This acts as an energy field throughout space of constant strength; as space expands there is an ever-increasing amount of energy, which then accelerates the expansion.

As an interesting aside, the constant, Λ, was initially used as a term that Einstein added to his field equation of general relativity to try and explain a steady-state or static Universe. This was essentially a "fine tuning" term (which I like to call a means of fudging your model/theory to fit the observations... a fudge factor if you will). With Hubble's work on the expansion of the Universe, Einstein eventually realised that the Universe did not actually need to remain static and could dynamically evolve. Therefore, he got rid of the term and referred to this whole escapade as his 'greatest blunder'. He famously met Hubble at the Mount Wilson Observatory and talked to him on his work on the expansion of the Universe. Though it is something of a myth that he decided to get rid of the constant straight after the meeting. It took a few months yet before he fully renounced the constant. It is interesting that the term actually ended being reinstated from the 1990's to try and account for the effects of dark energy, so maybe he did not make a blunder after all. Supernovae observations imply that the constant is a positive non-zero value.

Studies have shown that about 72% of the mass-energy density in the Universe is attributed to dark energy. The energy density of dark energy from the Supernova studies is enough then to give:

$$\Omega_m + \Omega_\Lambda = 1$$

Which you may remember is the value estimated from missions such as WMAP and Planck.

The actual nature of what dark energy is, is another thing entirely and is even more speculative than dark matter. It is not known to act through any other fundamental forces aside from gravity. That coupled with the fact that it is quite rarefied, implies that is unlikely to be detectable in laboratory experiments.

Hence, at the end of 2020, what we know about the Universe.:

- Age of the Universe: $t_0 = 13.8 \times 10^9$ years.
- Density of the Universe (Matter + Energy): $\rho = \rho_c$
- Fractional density of normal (baryonic) matter: $\Omega_b = 0.05$
- Fractional density of dark (non-baryonic) matter: $\Omega_d = 0.23$
- Fractional density of dark energy: $\Omega_\Lambda = 0.72$

But currently we do not know what dark energy and dark matter are.