Stars and **Galaxies**

Tom Theuns
Office 307
Institute for Computational Cosmology
Ogden Centre for Fundamental Physics
Durham University
and
University of Antwerp
Campus Drie Eiken
tom.theuns@durham.ac.uk

2006-2007

Prologue

Before the beginning

This course on galaxies runs over 14 lectures. The aim is to give you an overview of galaxy properties and of some of the physical processes that we think shape galaxies. In contrast to the beauty of stellar physics, you'll see that galactic physics is often much more murky and complex – a bit like weather predictions: we (think we) know the physics, but even so it still is difficult to make accurate calculations. And often, we don't know the physics very well – or even at all. On the bright side, galaxies are beautiful. And we can in principle probe physics on scales which are impossible to reach within the laboratory and so learn about really fundamental physics – just recall the recent excitement about the possible non-zero value for the cosmological constant. Anyway, I hope you'll enjoy the lectures.

As to the organisation of the course, I will basically follow the structure of Simon Morris, i.e. there will be hand-outs after each lecture, but more up-to-date versions will appear on the web in DUO. Look there as well for (most of) the pictures that I show during the lectures. Suggestions on material included in this course, its presentation, and lists of errors in the notes, will be most appreciated.

And for those that are only interested in passing their exams: there will be two questions on galaxies. The first consist of 5 very short questions where you're asked to describe *very briefly* some concepts that we have discussed. For example, what is the interstellar medium, and how do we observe it. Given five differences between football and basketball. The second type of question usually starts from a derivation that we've done. The end of the question is something you have not seen, but is very similar to what we have done.

My source of knowledge

Although this field is continually evolving, much of the basics is of course text-book stuff. The web provides a very valuable source of information, not just for pictures. Often, satellite mission web-pages have extremely well designed pages. Some links you might want to look at (and where I got many of the pictures) are:

- http://www.seds.org/
- http://www.aao.gov.au/images.html/
- http://hubblesite.org/newscenter/archive/
- http://teacherlink.ed.usu.edu/tlnasa/pictures/picture.html
- http://astro.estec.esa.nl
- http://space.gsfc.nasa.gov/astro/cobe/cobe_home.html
- http://astro.estec.esa.nl/Hipparcos/
- http://astro.estec.esa.nl/GAIA/

and the books I've used are

1 Carroll & Ostlie, *Modern Astrophysics*, Addison-Wesley, 1996

2 Zeilik & Gregory, *Astronomy & Astrophysics*, Saunders College Publishing, 1998

3 Binney & Merrifield, *Galactic Astronomy*, Princeton, 1998

4 Binney & Tremaine, *Galactic Dynamics*, Princeton, 1987

(I'll refer to these as CO, ZG, BM, and BT.) The first two are quite more general, [3] provides a nice overview on galaxies in much more detail than I'll do, and [4] of course concentrates on the dynamics of galaxies. [1] contains everything you *need* to know, in [2] you can often find the basics formulated in a very clear way, and [3] and [4] are where you could start looking if you want to know more.

A truly excellent web site where much of what I'll discuss is explained really well with many pictures, and by someone who knows what he's talking about, is by **Chris Mihos**, who is at the department of astronomy, Case Western University. You'll find it at:
http://burro.astr.cwru.edu/Academics/Astr222/index.html, and you'll recognise some of the picture in my course as originating from there.

Chapter 1

Introduction

1.1 Historical perspective

Galaxies were discovered in the 20th century. This is surprising, since you can see many of them on a clear night with the naked eye. They differ from stars in that they are *extended*. For example, the nearby *Andromeda galaxy* has an extent of several degrees (diameter of full moon is 0.5 degrees). And the *Large* and *Small Magellanic clouds* in the Southern hemisphere, are even bigger. And the nearest galaxy (which was only discovered in the last years!) extends over a significant fraction of the whole sky! The reason that they don't stand-out very clearly is not because they are *faint*, it is because they are so *large*. Since they are large, their light is distributed over a wide area, and our eyes do not pick-out the small increase in surface brightness over the night sky.

They were called *nebulae* to distinguish them from stars. However, there are other extra-solar objects which are also extended on the sky: Planetary Nebulae, star clusters, super nova (SN for short, and SNe for the plural) remnants, for example. Since before the 1920s we didn't know the distance to those objects, it wasn't clear what were the relations between them. And so many of the catalogues of nebulae, for example Messier's catalogue (objects denoted by an 'M' , like M31, which is Andromeda), or NGC (New General Catalogue, Dreyer 1888), contain a baffling variety of unrelated objects.

Immanual Kant, the German philosopher, suggested that the reason we see a faint band of stars across the night sky – usually called the *Milky Way* (MW for short) – is because most MW stars, as well as the Sun, lie in a disk. And so, when you look out at night in the plane of the disk, you see a faint band of light. Galileo already had resolved this band of light into stars. Kant claimed that most of the other nebulae were just other galaxies. It took until the 1920s before this hypothesis was proved correct. The historical introduction in BM is well worth reading.

As a final point: why study galaxies? Are they in some sense the most important building blocks in the Universe? I like BMs reasoning: stars are apparently huddled together in galaxies. And since (not surprisingly if you think about it) humans can *see* stars, we are naturally inclined to study them. But, for example, although a *cluster of galaxies* contains thousands of galaxies, they are by no means its main constituent. Most of the *mass* is in dark matter – which we cannot see at all. But most of the baryons are in hot gas, which we can *observe* with X-ray satellites, but we cannot *see*. And so the choice to study galaxies is a bit human centred: our eyes can see them.

1.2 Bringing order to the zoo of galaxies

Before we start doing any calculations, it might be a good idea to look around us and make an inventory of what types of galaxies are out there. There are many good web pages to start, for example I like http://www.taas.org/gnto/astropic.html, because it shows pictures made with a small telescope – so those are galaxies as you might have seen them yourself, if you are an amateur astronomer. http://www.aao.gov.au/images.html/ shows professional images. Such professional pages (for example the *Hubble Space Telescope* (HST) archive) tend to contain a large fraction of unusual galaxies – simply because those are the ones that astronomers tend to study.

1.2.1 Vocabulary

Colours. These images are taken through *broad-band filters*, typically[1] V, Band R, and so each individual exposure might as well be in black and white, the relative blackness being a measure of how much, or how little, of the light in that wavelength band fell on that part of the photograph/CCD. Images taken trough different filters are then combined to produce a *colour image*, scaling the different bands trying to mimic the colour your eye would see.

[1]The names give an indication of the corresponding wavelength region, B for blue, V for visual, R for red, U for UV (in the direction of the UV-that is), I for IR.

Sometimes though, one of the filters used is a *narrow-band filter*, for example Hα or O[III]. These filters block all light which is not in a narrow range around the hydrogen Hα transition, or a transition in doubly ionised oxygen. Combining such a narrow-band image with images made with others filters produces a *false colour image*, i.e. the colour of the object is *not* how your eye would see it, but the colour coding has been chosen to bring out a particular feature – for example Hα emission. The really pretty colour pictures of Planetary Nebulae are false colour images.

Luminosity, flux and surface brightness

A star emits a certain amount of energy per unit time. For example the Sun has a *luminosity* $L = 1L\odot = 3.4 \times 10^{26}$W. This quantity is not observable however: what we can measure is how much of this energy we receive, per unit time, per unit surface area (assuming you put the surface perpendicular to the radiation!), at a given distance from the source. The observable quantity, energy received per unit time per unit of surface area, is called flux. Clearly the flux will depend both on the luminosity of the star, and on its distance. How?

Consider a sphere of radius r, with the star at the centre. Since the star will distribute all of its energy equally over the surface of the sphere, the flux $F = L/(4\pi r^2)$. (As an exercise, find the unit of flux).

It is the *flux* of the star that determines how bright it appears to be. Indeed, star A may be brighter than star B, either because it is more luminous, $L_A \gg L_B$, or because it is nearer to us, $r_A \ll r_B$.

Although F is in principle observable, in practise astronomers express brightness in terms of magnitudes m, where the apparent magnitude $m = -2.5\log_{10}(F)+C$, where C is some constant that depends on the magnitude system used. Since the F depends on distance, so does m, it is not an intrinsic quantity of the star – hence 'apparent magnitude'. The absolute magnitude, M, is the apparent magnitude when $r = 10$pc.

Since a galaxy contains many stars, we can also talk about its luminosity, and the flux we receive from it. However, because galaxies are spatially extended[2], we can also try to measure what fraction of the flux comes from the centre of the galaxy and what fraction comes from the outer parts, say. This gives rise to the term *surface brightness.*

Consider a small patch of galaxy, namely that part contained in a (small)

[2] *Very* nearby stars can be resolved with special techniques.

solid angle $d\Omega$. Observationally we can measure the flux, dF, of light, coming from the part of the galaxy contained within $d\Omega$. The quantity $dF/d\Omega$ is called *surface brightness* (SB), and so it has dimensions energy/time/area/steradian.[3]

The surface brightness of a galaxy does not depend on its distance. To see this, assume that the patch of galaxy contained within $d\Omega$ contains N stars, with mean luminosity L, when it is at a distance r. The surface brightness SB=$N\,L/(4\pi r^2)/d\Omega$. Now increase the distance to the galaxy by a factor of 2. The flux from each individual star will decrease by a factor 2^2, since $F \propto 1/r^2$. However the number N of stars within $d\Omega$ will increase by a factor 2^2, since the physical size of the patch enclosed by $d\Omega$ will double when r doubles. Hence SB is distance independent[4], therefore whereas with brightness there was an intrinsic quantity (luminosity), and an observable one (flux), there is only one surface brightness.

A galaxy's SB depends on its distribution of stars. Assume a face-on galaxy at distance r has a surface density Σ of stars (in stars/pc^2, say), all with the same luminosity L. The surface area of galaxy dS, contained within a solid angle $d\Omega$, is $dS = d\Omega\, r^2$. Therefore the number, dN, of stars within $d\Omega$ is $dN = \Sigma\, dS = \Sigma\, d\Omega\, r^2$. The flux received from a single star is $L/(4\pi r^2$, and therefore the flux received from all N stars within $d\Omega$ is $dF = (L/(4\pi r^2)) \times (\Sigma\, d\Omega\, r^2) = \Sigma\, L\, d\Omega/(4\pi)$. The surface brightness is the flux per unit solid angle, is $df/d\Omega = \Sigma\, L/(4\pi)$ is distance independent, as we saw before.

1.2.2 Galaxy properties

Looking at the different images of galaxies, it becomes immediately clear that there are two basic types, elliptical and spiral galaxies.

Ellipticals (also called early type galaxies, E galaxies for short)

- have smooth light distributions, where the isophotes (lines of constant SB) have (nearly) elliptical shapes. They are denoted as En, where $n = 10(1 - b/a)$, and a and b are the semi-major and semi-minor axis of the ellipse. So, an E0 is round, and an E7 is flattened, $n = 7$ is about how flattened E's get. A 3D ellipsoidal shape, with semi-axes a, b and

[3] A solid angle is usually expressed in steradians, or arc sec^2, see http://whatis.techtarget.com/definition/0,,sid9_gci528813,00.html.

[4] This is only true for relatively nearby galaxies. On cosmological distance the SB decreases with distance because of the redshift.

c (ordered as $a \geq b \geq c$), is called oblate when $a = b > c$ (a pancake), prolate when $a > b = c$ (a cigar), and tri-axial when $a > b > c$ (a squashed American football). E's are probably tri-axial.

- have no current star formation
- have old stars
- have little gas and dust
- occur preferentially in groups and clusters of many Es
- have little evidence for rotation, so the stars must have large random motions to keep the system from collapsing (also called a hot stellar system)

Spirals (or late type galaxies, S galaxies for short)

- have most of the light coming from a thin disk. Seen face on, it is nearly circular, seen edge on, it is nearly a line. In their centre, they often have a nearly elliptical stellar system, similar in shape to a small elliptical, called a bulge. In addition, some have a bar as well.
- have light profiles with a characteristic spiral pattern
- have star formation going on, especially in their spiral arms
- have both old and young stars
- have gas and dust
- seem to avoid dense groups, and are very rare in clusters
- have large rotational velocities, enough to keep their stars on circular orbits.

We inferred the *stellar* properties from the fact that E galaxies tend to be redder than S. This suggested that they lack the young massive stars that produces most of the blue light. Hence our inference that the stellar population is likely to be older, with the massive stars that have short life times already dead. We will see later that the SB distribution, $SB(r)$, are quite different for these two types.

These are the archetypal properties of such systems, of course there are exceptions. For example, some ellipticals do have dust, as well as evidence for recent star formation. Some even have a small disk. Given that some spirals have a very large bulge, and some ellipticals have a disk, the distinction between the two types is not always as clear as you might think. Clearly it would be interesting to know what processes are responsible for forming an elliptical vs a spiral in the first place.

Both types are seen to have hundreds and sometimes thousands of *globular clusters* (GCs for short). These are very dense, spherical, gravitationally bound systems of typically 10^5 to 10^6 low mass stars, and they are usually found in a spherical distribution around their parent galaxy. The Milky Way has about 150 of them.

There are many schemes for classifying galaxies. A good scheme should be unambiguous, and ideally have some physics behind it. Classification along the *Hubble sequence* (Fig. 1.1) is still very popular. It divides galaxies in ellipticals and spirals. Spirals are divided in two strands, according to whether they are barred (SB), or not (S). The sequence has the shape of a tuning fork – the Hubble tuning fork. Along the handle of the tuning fork, the range goes from E0 to E7. The sequence then forks, toward the top are unbarred spirals (S), along the bottom are barred spirals (SB). Both strands are further divided along the sequence (away from the handle), as Sa toward Sc, and analogously SBa to SBc, depending on the relative strength of bulge to disk, and, correlated with this, how tightly/loosely wound the arms are. Sa's and SBa's have a large bulge and tightly wound arms, Sc's and SBc's have a much smaller bulge, and more loosely wound arms. Finally, any galaxy that does not fit into the sequence, is called irregular (Irr).

The physical scale of galaxies is huge: E's can have masses from as little as $10^7 M_\odot$ to as much as $10^{13} M_\odot$, with linear sizes ranging from a few tenths of a kpc to hundreds of kpc. In contrast, spirals tend to be more homogeneous, with masses 10^9–$10^{12} M_\odot$, and disk diameters from 5 to 100kpc or so.

Hubble thought (incorrectly) that this sequence was due to evolution, where galaxies evolve from being E0s, toward Sa or Sba, and later evolving in the more loosely wound Sc's or SBc's. Therefore these types are also called early type and late type. So E's are also called early type, and Ss are called late type. And an Sc is a late-type spiral, and Sa an early type spiral. We know now that this is not the way galaxies evolve, yet these names are still widely used.

There are other reasons why this is not a particularly good scheme. For

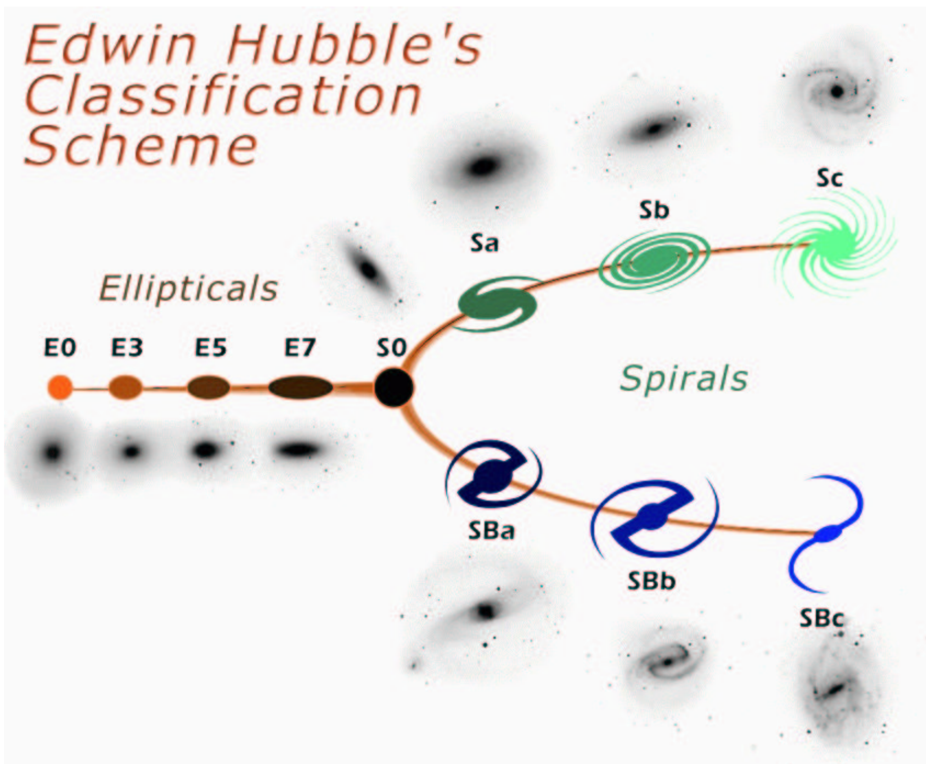

Figure 1.1: The Hubble sequence.

example, galaxies may look quite different depending on the filter you use to observe them in.

Another type you'll often come across are cD galaxies, which stands for **c**entral **D**ominant: these are very large ellipticals, found in the centres of clusters, with a faint but very large outer halo of stars.

Yet another type are called S0 (or SB0 when barred) or lenticulars: they are the divide in Hubble's sequence between E and S. They have disks without gas or dust, and no recent star formation.

Galaxies can fail to fit in the sequence, for example because they recently had a tidal encounter with another galaxy. This typically leads to spectacular tidal arms, like the case of e.g. the Antennae.

Most smaller galaxies are of the Irregular type: no disk, no nice spherical distribution of stars either, but a loose aggregate of stars, like for example the Magellanic Clouds. Since there are far more small than large galaxies, most galaxies are in fact of type Irr. The aim of these lectures is to understand which types of observation can be used to describe these different types a bit better (the 'biology bit of the lectures'), and also try to understand the physical processes behind them. May be you'll be surprised to learn that, for many of the most basic questions you might think of, such as – why are some galaxies of type E, and others of type S?, or, why do Ss have spiral arms? – we are only just beginning to understand the underlying physics, and we have by no means a well defined, generally accepted textbook answer. May be something for you to work on, when you start doing astronomy research yourself?

1.3 Summary

After having studied this lecture, you should be able to

- Describe the main galaxy types, Es and Ss, and list five characteristics of each
- Describe some properties of Globular Clusters.
- Define surface brightness, and show it to be independent of distance.
- Derive the relation between the surface density of stars in a galaxy, and its surface brightness.
- Know where to find web-pages with images of galaxies.

Chapter 2

The discovery of the Milky Way and of other galaxies

Around the 1920s a great debate was raging about the nature of the nebulae and the size of the Milky Way. There were two camps. According to one camp, the nebulae were structures within the Milky Way, probably protoplanetary systems or something similar, and the Sun was near the centre of MW. The other camp claimed that many of the nebulae were galaxies outside of the MW, and the Sun was at large distance from the centre of the MW. The reason they couldn't decide was mainly because they didn't know how to determine distances to the nebulae. The reason the first group got it wrong was quite interesting. And so we'll go through this debate, hoping we can learn from the mistakes. Before we do this, let's describe the main observables, and observing techniques.

2.1 The main observables

Star clusters There are two types of star clusters, *i.e.* groups of 10^3s to 10^5s of stars. **Globular Clusters** (GCs) are spherical, gravitationally bound systems of $10^5 - 10^6$ old stars. Most galaxies contain 10s to 1000s of GCs, spherically distributed around the parent galaxy. The MW has about 150 GCs. Stars form in the MW disk out of large molecular clouds. Sometimes, after the gas has been dispersed, the newly formed stars are gravitationally bound to each other, in what is called an **Open Cluster**. These typically contain of order 10^4 stars. So they differ from GC in that they are typically smaller, have young stars, and lie in the galactic plane. Star clusters are of great importance for studies of stellar evolution, since we know all the stars in such a cluster have (more or less) the same age, distance and initial composition, and hence they provide us with a sample of stars in which the only (or at least the main) difference is the stellar mass.

Star counts Star counts can be used to probe the number density of stars as a function of position. In the early days, William Herschel and his sister counted the numbers of stars as function of their magnitude in many (700!!) directions in the sky, by spending (most?) nights gazing through his telescope! Later, photographic plates were used for the same purpose. From the fact that the number of stars falls with magnitude much faster perpendicular to the disk, than in the plane of the disk, one deduced that the stars around the Sun are indeed distributed in a disk, with axes ratio about 5:1.

How does it work? Suppose you count the number of stars dN, in a given solid angle $d\Omega$, that have flux fainter than F, but brighter than $F-dF$ (with $dF>0$). If all stars have luminosity L, then stars with flux F are at distance r, given by $F=L/(4\pi r^2)$. Similarly, stars of flux $F-dF$ are at distance $r+dr$, given by $F-dF=L/(4\pi\,(r+dr)^2)$. If $dr\ll r$, then $dF=2F\,dr/r$. The volume of the fraction of a spherical shell that falls within a solid angle $d\Omega$, between r and $r+dr$ is $dV=d\Omega\,r^2\,dr$. Combining all this we get for the number of stars per unit solid angle, with flux between F and $F-dF$

$$\frac{dN}{d\Omega}=n(r)\,r^2\,dr=n(r)\,r^2\,\frac{r}{2F}\,dF=\frac{1}{2}\,(\frac{L}{4\pi F})^{3/2}\,n(r)\,\frac{dF}{F}\,. \qquad (2.1)$$

The function $dN(F)/d\Omega$ in the plane of the MW drops much slower with decreasing F than perpendicular to the plane, meaning $n(r)$ drops much slower within the disk, then perpendicular to it. Which is because the stars are distributed in a disk around the Sun, not spherically symmetric. This is how you can determine $n(r)$ by counting stars.

Standard candles are objects with a known absolute property, for example a known size, or known luminosity. It is then possible to determine the distance to the standard candle by measuring their angular size , or apparent luminosity, respectively. A most important example is that of **Cepheid variables**. Henrietta Leavitt studied variable stars in the Magellanic Cloud in 1912. She found a relation between the period P of the variation Δm and the magnitude m of these Cepheids. Since all these stars are at (nearly) the same distance, this must mean it is actually a relation between the *absolute* magnitude M and P. There are many types of variable stars, but Cepheids have many advantages as standard candles (a) the shape of their light curve (i.e. luminosity as function of time) is a very characteristic sawtooth pattern, (b) they are luminous - and so can be seen out to large distances and (c) the $P(M)$ relation has little scatter. Unfortunately, they are also relatively rare. **RR Lyrae** are similar to Cepheids, but occur in a different type of star. They are also used as standard candles. Of course, to get an absolute distance, we need somehow to find the distance to some Cepheids or RR Lyrae using another method, for example using the parallax.

Parallax Stretch your arm, point your index finger upward, and look with your left eye alone toward a distant wall. Now, look with your right eye alone: your finger seems to have moved with respect to the background. You've done (your first?) parallax measurement. If the wall is sufficiently distant then there is a relation between how much your finger appears to move (in degrees, say), the length of your arm, and the distance between your eyes. If you now one distance (e.g. between your eyes), you can determine the other (length of arm).

You can increase the distance between your eyes to twice the distance earth-Sun, by looking at the same object half a year apart. Since one can relatively easily measure angles to a fraction of an arcsec [1], we have a new distance unit: the pc. An object at 1pc distance has a parallax of 2arcsec. Let's compute how much this really is. Consider an equi-lateral triangle, with two sides of length 1pc third side 1AU. By definition 1pc/1AU=1arcsecin radians[2] Hence

$$1\text{pc} = \frac{180 \times 3600}{\pi} \text{AU} \approx 206265\text{AU}\,. \tag{2.2}$$

Unfortunately, this literally doesn't get use very far: the distance to even the nearest stars is of this order. We'll come back to the distance scale later.

The parallax and Cepheid variables are the first two steps in the *distance ladder*, which use one method (e.g. parallax) to calibrate another distance measure (e.g. Cepheids), that then can be used to calibrate another distance indicator, and so go to go to greater and greater distances. The Hubble space telescope recently measured Cepheids in the Virgo cluster (at a distance of

[1] An arcsec is a 60th of an arc minute, or 1/3600th of a degree. Now we can do much better!

[2] We've used the fact that this angle is very small.

$\sim$ 17Mpc), thereby providing the first accurate distance to that cluster of galaxies, and getting an accurate measurement of Hubble's constant in the process.

2.2 The main players, their measurements, and their mistakes

The famous Dutch astronomer, *Jacobus Kapteyn* and his collaborators, started a major project of counting stars around the 1900s, to determine once and for all the structure of the MW. He counted stars on photographic plates as function of apparent magnitude (brightness), in several directions.

His analysis basically confirmed Herschel's picture: the MW is a flattened elliptical system, with stellar density n decreasing away from the centre. n drops to half its central value at 150pc perpendicular to the MW plane, and 800pc in the galactic plane. The Sun is at 650pc from the centre. This is in fact wrong but for an interesting reason.

Kapteyn assumed that the flux received drops as $1/r^2$. One reason this could be wrong is because of absorption or scattering of the stellar light, in which case the flux will drop faster than $1/r^2$. Of course, not taking absorption into account would give you the wrong answer, and Kapteyn was fully aware of this. How can we test whether absorption is important? For *Rayleigh scattering* off atoms, there is a colour dependence: blue light is scattered more than red light. As a consequence fainter and hence more distant stars would be expected to be *redder* on average. Now Kapteyn found this indeed to be the case, yet the *amount* of reddening was too small for absorption to be important. The mistake in the reasoning was that the absorption is due to *dust*, not Rayleigh scattering. Dust absorption is much less colour dependent, and so the small reddening measured by Kapteyn actually implied a much greater dimming of the distant stars. Consequently the size of the MW as well as the position of the Sun in it came out wrong.

The American astronomer *Harlow Shapley* estimated the distances to the brightest globular clusters using their RR Lyrae variables, and found they were not centred around the Sun, but around a point 15kpc away in the direction of Sagittarius. This was in conflict with the Kapteyn universe – it seemed difficult to reconcile the Sun to be near the centre of the MW, yet the centre of the globular cluster system not to coincide with the centre of

MW? There was no clear way how to reconcile these different data, and so the issue remained unresolved for some time.

Meanwhile, another famous Dutch astronomer, *van Maanen*, found that some of the nebulae had proper motions, implying they were nearby, within the MW. (He compared the positions of the nebulae on plates taken several years apart). It was generally accepted these nebulae were probably proto-planetary systems, or something like that. van Maanen himself, but also *Edwin Hubble* later redid the proper motion measurements with better data (using a better telescope), and both failed to reproduce the proper motions. van Maanen probably just made a mistake! In the 1930s, Hubble made his even more famous discovery that the spectral lines of atoms in other galaxies are shifted to the red, by an amount proportional to the distance to the galaxy.

So it was a combination of mistakes (van Maanen), and not taking into account the importance of dust (Kapteyn), that led to the confusion. In the end, Hubble's better data clinched the matter. He resolved Cepheids in Andromeda, one of the bigger nebulae, and found a distance of 300kpc and so clearly outside of the MW. In one great swoop, the universe because much bigger, and in addition it was found to be expanding! It is difficult to underestimate the significance of this discovery.

The 1920s were of course a period of great economic depression. But from the scientific point of view, they must have been tremendously exciting. In only a couple of years, the general theory of relativity was formulated, quantum mechanics revolutionised physics, the MW was recognised to be just one of millions of other galaxies, the Universe was found to be big and expanding

It was not until the 1930s that *Trumpler* found what had been wrong with Kapteyn's results. He studied Open Clusters in the MW plane. He assumed them to be all of about the same size, and so got an independent relative distance for them. He used another estimate of their distance, based on 'spectroscopic parallaxes'. This involves predicting the luminosity of a star from its spectrum. He showed that the more distant clusters (as judged from their angular size), appeared to have unusually faint stars in them. Or equivalently, that the more distant clusters (as judged from their photometric parallax) were larger. Clearly, something was wrong here. He correctly interpreted this, as due to the fact that the light from these more distant stars was strongly attenuated. By comparing the reddening with the attenuation, it was clear that a given amount of reddening corresponded to much more dimming than you'd expect for Rayleigh scattering. And hence Kapteyn's scales were wrong.

Before I summarise this in a time-line, just think of the fact that Einstein formulated his theory of general relativity, on which the current cosmological models are still based, before this puzzle was resolved. In fact, our view of the Universe around Einstein's time, was just plain wrong. I guess he rightly paid little attention to what the astronomers of his days told him!

2.2.1 Time-line

1610 Galileo resolves the MW into stars

1750 Immanual Kant suggests that some of the other Nebulae are other galaxies, similar to the MW.

end of 1700s Messier and Herschel catalogue hundreds of Nebulae. Herschel counts stars, and deduces that the Sun lies near the centre of an elliptical distribution with axes ratio 5:5:1

1900-1920 Kapteyn counts stars, decides wrongly that extinction is unimportant, and deduces the MW to be 5kpc $\times$ 5kpc $\times$ 1kpc big, with the Sun at 650pc from the centre.

1912 Leavitt discovers the $P(L)$ relation for Cepheids.

1914 Slipher measures large (1000s km s^{-1}) velocities for some Nebulae, and finds evidence for rotation. The spectra he takes suggests presence of stars, not of gas. A clear indication these are not proto-planetary structures in the MW, but other galaxies.

1915 Shapley finds the centre of the MW's globular cluster system to be far away from Kapteyn's MW centre.

1920 van Maanen claims (erroneously) that some spiral nebulae have a large proper motion, suggesting they are within the MW.

1920 Shapley and Curtis debate publicly over the size of the MW, but the matter is not settled.

1923 Hubble resolves M31 (Andromeda) into stars, using the newly commissioned 100-inch telescope. Given the large inferred distance means that M31 must be outside the MW. He also discovers Cepheids, and the distance to M31 is estimated at 300kpc. So Andromeda us indeed another galaxy.

1926 Lindblad computes that Kapteyn's MW is so small, it cannot gravitationally bind its Globular Clusters. But Shapley's much bigger MW could.

1927 Jan Oort shows that several aspects of the local motion of stars can nicely be explained if the Sun (and the other nearby stars), is on a nearly circular motion around a position 12kpc away in the direction of Sagittarius. Nearly the same position as found by Shapley, and implying a much larger MW than Kapteyn's.

1927 larger MW picture, where many of the nebulae are extra-galactic MWs, gains general acceptance.

1929 Hubble discovers his expansion law. His derived value is a factor of 10 too large!

1930 Trumpler uses open clusters to show the importance of extinction, and explains why Kapteyn's measurement were faulty

1930-35 Hubble's new data confirm the modern picture of galaxies, and demonstrates van Maanen's measurements must have been wrong.

And so in a very short time indeed, three sweeping changes in our view of the visible universe took place:

- an 8-fold increase in the size of the MW
- the acceptance that many of the nebulae were external galaxies
- the universe is expanding

2.3 Absorption, scattering and reddening

What is the effect of dust on the light we detect from distant stars? Suppose the wavelength of the light is much smaller than the size of the dust grains (bricks, say) – then you could compute the dimming just from the geometric cross-section of the bricks along the line of sight. Consider a ray of light with intensity I. The amount of light of this ray absorbed per unit distance, will satisfy

$$\frac{dI}{dr} = -A\,I\,, \tag{2.3}$$

which expresses the fact that each distance dr will absorb a constant *fraction* $dI/I = -A\,dr$ of the light. This is what we want since, if the light source were twice as bright, then the bricks will stop twice as many photons. A is some constant, which will depend on the size of the bricks, and their number density. This equation is easy to integrate, and the solution is $I = I_0 \exp(-Ar)$. Taking the log of both sides, we find that the effect on the magnitude will be $\Delta m = A\,r$. So the relation between apparent (m) and absolute magnitude (M) will change from $m - M = 5\log(r) - 5$ to

$$m - M = 5\log(r) - 5 + A\,r\,, \tag{2.4}$$

due to the effect of absorption. So the unit of A is magnitudes per kpc. (Careful readers will notice that I have slightly changed the definition of A along the way, but you get the idea.)

If the size of the particles is of order of the wavelength λ of the light, then it is not a good approximation to use the geometric cross-section of the dust grains to estimate absorption. The amount of absorption (the value of A) will then depend on λ, and so $A_{\rm V} \neq A_{\rm B}$, say: absorption in the B band will differ from that in the V band, and so absorption will also lead to a colour change (reddening, since the longer wavelengths get *less* absorbed than the shorter ones).

Applying Eq.(2.4) to two bands, we find

$$\begin{aligned}
(m-M)_{\rm B} &= 5\log(r) - 5 + A_{\rm B}\,r \\
(m-M)_{\rm V} &= 5\log(r) - 5 + A_{\rm V}\,r \\
E_{\rm B-V} &\equiv (m_{\rm B} - m_{\rm V}) - (M_{\rm B} - M_{\rm V}) = (A_{\rm B} - A_{\rm V})\,r\,. \qquad (2.5)
\end{aligned}$$

The quantity $E_{\rm B-V}$ is called the colour excess, note that it is the difference between the observed and intrinsic colour of the star,

$$E_{\rm B-V} = ({\rm B}-{\rm V})_{\rm observed} - ({\rm B}-{\rm V})_{\rm intrinsic}\,. \tag{2.6}$$

Trumpler's measurements, and also laboratory measurements, show that for interstellar dust grains

$$E_{\mathrm{B-V}} \approx \frac{1}{3} A_{\mathrm{V}}\, r\,. \tag{2.7}$$

This is a crucial result. Reddening and hence $E_{\mathrm{B-V}}$, is easy to measure, and so if we do this for stars of known distance, we find[3] $A_{\mathrm{V}} \approx 1\mathrm{mag\,kpc^{-1}}$. If we now measure $E_{\mathrm{B-V}}$ for another star of *known* colour (from stellar evolutionary models say) we can estimate r.

[3]The amount of dust is not the same everywhere: there are regions where the absorption is much stronger, not surprisingly called dark clouds, and some directions along which the absorption is much less, a well known direction is called Baade's window.

2.4 Summary

After having studied this lecture, you should be able to

- Describe the Great Debate, give the points in favour of each camp, and discuss the problems with their arguments.
- Derive the relation between the stellar density and star counts
- Explain what a standard candle is.
- Explain why Cepheids and RR Lyrae variables are good standard candles.
- Explain what parallax is, and know how the distance to nearby stars can be measured using their parallax.
- Explain how parallax and Cepheids can be used to walk-up the distance ladder.
- Explain how Hubble's observations revolutionised our view of the MW and the realm of the Nebulae.
- Derive the effect of scattering by dust on the apparent magnitude and colour of distant stars. Know the approximate value of the interstellar absorption in the MW.

Chapter 3

The modern view of the Milky Way

Our observational picture of the MW has been clarified due to steady hard work and occasionally in great leaps with the advent of new technologies.

3.1 New technologies

3.1.1 Radio-astronomy

Radio astronomy used the new technology developed for radar and opened a new window for astronomers to gaze through. Roto-vibrational[1] transitions in molecules have low excitation levels, which are ideally suited to probe the low temperatures in interstellar clouds. Because the wavelengths of these transitions are long, one can build very large dishes to observe them, and also it is more straightforward to build interferometers.[2]

[1]Most transitions you know, which occur in the visible and UV-part of the spectrum, are *electronic* transitions: in emission, they occur because an electron changes from a higher energy state to a lower one, e.g. from $n = 2 \rightarrow 1$ for the hydrogen Lyman-α transition, and from $n = 3 \rightarrow 2$ for the hydrogen Hα transition. And in absorption, they go the other way around. But molecules also have excited levels due to their rotation or vibration. Since these are also quantum-mechanical the energy levels are quantised. And transitions between them correspond to rotational or vibrational transitions. Since the energy-levels are lower they correspond to longer wavelengths – typically mm and cm: in the radio.

[2]One way to think of this is to realise in good 'mirror' the departure from the ideal shape should not be larger than some small fraction of the wavelength of the radiation

An electron moving in a magnetic field will feel an acceleration, causing it to circle around the field lines. Such *synchrotron radiation* can also be observed in the radio.

Today, one regularly combines data from radio-observatories across the earth to get Very Long Baseline Interferometry, which gives superb angular resolution (10^{-6}arcsec!). Another great advantage of radio over optical observations, is that radio waves are not absorbed by dust. (By now you know why: it's because the wavelengths are much larger than the size of the dust grains.) And so, yo can use radio-waves to probe into very dense regions, for example where star formation occurs. One of the bigger problems facing radio-astronomy today, is mobile phones ...

3.1.2 Infrared observations

IR astronomy took-off relatively recently, with the advent of rockets and satellites to get the instruments above the atmosphere. The water vapour in the atmosphere will otherwise absorb a large fraction of the IR light. The construction of the camera for near IR observations is challenging. They need to be cooled, since otherwise the main source of signal is thermal radiation from the camera itself. I suspect that, again, a major driver behind their construction was not really astronomical, but over the past 10 years or so, it seems that much of the technology has becomes available to the astronomers.

IR-observations share with radio the great advantage not to be strongly absorbed by dust. In fact, much of the visible and UV-light absorbed by dust grains, is re-emitted in the IR and sub mm. And so we can use IR to probe the dust and look deep inside star forming regions. We thank the DIRBE instrument on the COBE (Cosmic Background Explorer) satellite for one of the best views of the Milky Way.

---

you plan to observe. So for optical telescopes, it should be of the order of 500nm, say, but for radio-waves, fractions of a mm, and up to a m. There is a down-side. The resolution you get, when making diffraction limited observations, is of order λ/D, where λ is the wavelength, and D the diameter of the telescope. And so, radio-telescopes have to be much bigger to obtain a similar resolution. In an interferometer, instead of building one giant dish – which would be very expensive, and difficult to move – one builds many smaller dishes, put far apart, and combines their output. The result is that the effective D is not the size of a dish, but rather the biggest distance between them. What you loose, is that the collecting area is only the sum of the sizes of all dishes, not D^2.

3.1.3 Star-counts, again

Although counting stars may seem boring, the Hipparcos satellite mission[3] was really a tremendous revolution, due to the incredible accuracy of its measurements. This satellite was launched in 1989 and specifically designed to measure very accurately the positions of stars, several times during the satellite's lifetime. For distant stars (some of which were in fact very bright, extra-galactic sources called quasars), this made it possible to construct an absolute reference system, that can be used define the positions of extra-solar objects. But for more nearby stars, Hipparcos could not only determine their parallax and hence their distance, but also, by comparing the positions at two different times, the *proper motion* of the stars with respect to the 'fixed' reference system.

The parallax is the ultimate step in creating a distance scale. We first need to determine the parallax to standard candles, like e.g. Cepheids, before we can use them to look further. So getting these parallax measurements accurate is of major importance.

The proper motion, i.e. the dance of nearby stars around us, tells us about the distribution of mass. For example, the oscillation of stars perpendicular to the disk of the MW tells us how much mass there is locally in the disk.

GAIA[4] is a planned mission to be launched before 2012. It will measure positions of stars up to 15-th magnitude with a resolution of 10^{-5}arcsec. Gaia's expected scientific harvest is 'of almost inconceivable extent and implication' (according to their web page). In total, about 10^9 MW stars will be measured (as compared to 10^4 measured by Hipparcos). Gaia will be so accurate that it can even measure proper motions of some of the nearest globular clusters and galaxies.

3.2 The components of the Milky Way

We've already briefly discussed what the major constituents of the MW are: stars, gas and dust. Now we'll describe how they're distributed. We think the MW is a pretty much ordinary spiral galaxy, of which we know several million others, but we know this one in a bit more detail. Even so, bare in mind that some of the details are uncertain. The three components are the

[3]http://astro.estec.esa.nl/Hipparcos/
[4]http://astro.estec.esa.nl/GAIA/

disk (there is a thin and a thick disk), *bulge* and *halo*.

3.2.1 The disk

Most of the MW's stars are in a thin disk, which represents about 70 per cent of the total star light of the MW. The centre of the disk is in the direction of Sagittarius, $\alpha = 17^h\,42^m\,29.3^s$, $\delta = -29^o\,59'$, at a distance of R_0. The latter is called the *solar galacto-centric distance*, and there is some debate about its value. $R_0 = 8.5$kpc is usually quoted, but some studies find $R_0 = 8.0$kpc. R_0 is usually determined from finding the centre of a set of objects which are presumed to be randomly distributed about the galactic centre, for example globular clusters, or variable stars such as RR Lyrae or Mira variables, or O and B stars. Another method uses the fact that there is an H_2O maser source is located near the centre of the MW. The maser emission is thought to come from a spherical shell, and one can measure both the radial and transverse speed of the shell. If the shell is expanding spherically, then one can determine its distance.

Besides stars, the disk also contains gas and dust. On top of the smooth disk are spiral arms, traced by young stars, molecular clouds, and ionised gas. The disk stars are in (nearly) circular motion around the centre. For the Sun, the circular velocity is ≈ 220km s^{-1}.

We can put limits on the ages of the *stars* in the disk, but have to keep in mind that the stars could possibly be older than the disk itself (i.e., they may have formed somewhere else). One neat age estimator are White Dwarfs. As you recall, these do not undergo nuclear fusion anymore, but are presently just cooling down. By measuring their current temperatures and cooling rates, we can compute how old they are. Some of the older White Dwarfs are thought to be $\sim 10 - 12 \times 10^9$ years old.

Let's describe the position of a star in the MW using cylindrical coordinates (R, φ, z), where R is the distance to the rotation axis, z the distance to the plane, and φ an angle. The density of stars in the disk goes approximately like $n(R, \varphi, z) \propto \exp(-R/R_h)\exp(-|z|/z_h)$. Since the disk is (nearly) axis-symmetric, $n(R, \varphi, z)$ is independent of φ. Away from the centre, n decreases exponentially, and at distance $R_h \approx 3.5$kpc, it has fallen to $1/e$ of it's central value. Also perpendicular to the disk, the fall-off is exponential, with $z_h \approx 0.3$kpc. R_h and z_h are called the *scale length* and *scale height* of the disk, respectively. These scale lengths really depend on the type of star you're using. Young stars are born out of gas which is much more concentrated toward $z = 0$, hence the scale height of young stars is smaller than of older stars.

Note that there isn't really an edge to the disk. It can be traced to a distance of around 30kpc. With a height of 0.3kpc, this is a ratio 100:1, which is thinner than a compact disk!

Given the exponential distribution in n, the luminosity profile (in $L_\odot$ per unit volume) of the disk is often modelled as

$$L(R, z) = L_0 \exp(-R/R_h) \frac{2}{\exp(z/z_h) + \exp(-z/z_h)}\,, \tag{3.1}$$

where L_0 is some normalisation constant. In the B-band, the total luminosity is $\sim 2 \times 10^{10} L_\odot$.

Because of the component I'll describe next, the disk is sometimes called the thin disk.

3.2.2 The thick disk

About 4 per cent of the MW's stars belong to a thicker disk, which is aligned with the (thin) disk, but has a scale height of $z_h \approx$ 1kpc. Its existence has been discovered only recently, in the 1980s. The difference with the thin disk is not just the kinematics of its stars: it turns out that thick disk stars are on average older than the stars of the thin disk, and have a lower metallicity [Fe/H]$\sim$-0.5, i.e. they have less Fe (and other metals as well, but we can usually measure Fe accurately) for a given amount of hydrogen (we'll see later what the notation [Fe/H] means).

3.2.3 The bulge

The bulge is a spheroidal stellar system with radius of $\sim$ 1kpc, located at the centre of the MW. In fact, its properties are quite similar to that of an elliptical galaxy: it is as if there is a small elliptical galaxy at the centre of each spiral. The surface brightness profile (in $L_\odot$ per unit area) of this component goes like

$$I(r) = I_e \exp[-7.6(r/r_e)^{1/4} - 1]\,. \tag{3.2}$$

This is a famous fit to the intensity profile of elliptical galaxies, introduced by *de Vaucouleurs* in 1948, and called the 'de Vaucouleurs' or '$r^{1/4}$'(r-to-the-one-quarter) profile. ($r = (R^2 + z^2)^{1/2}$ is the 3D distance to the centre.)

r_e is called the effective radius. Note that $I(r = r_e) = I_e$. The factor 7.6 is chosen such that half of the light comes from stars within r_e. For the MW, $r_e \approx 0.7$kpc.

Although the bulge is bright, we can't *see* much of it, due to the large amount of obscuring dust (in particular, you can't see it with the naked eye!). The total amount of extinction toward the centre is about 28 magnitudes in the visual! In other spirals, we can see the bulge very well (and recall that the classification of spirals is based on the bulge/disk ratio), but for the MW we need to go the the IR to get a nice picture of it. The MW's bulge is rather elongated with axis ratio 5:3, with strong evidence for a bar.

The stellar content of the bulge is a bit of a mixed bag, with both very old, low metallicity stars, but also young, metal rich stars. The mean metallicity is about [Fe/H]$\sim$ 0.3, i.e. twice the solar value. The mass of the bulge is $\sim 10^{10} M_\odot$, with a B-band luminosity of $\sim 3 \times 10^9 L_\odot$.

3.2.4 The stellar halo

The final stellar component of the MW is the stellar halo, composed of globular clusters (GCs), and 'field stars'. These stars, just like the stars in the globular clusters, are typically of very low metallicity, [Fe/H]~ -0.8. The current record holder of low metallicity has [Fe/H]≈ -5.

Unlike the disk, the halo rotates very little, if at all. As a consequence, halo stars all appear to move at high velocity (in fact, rather it is the Sun that is moving at ~ 220km s^{-1}), which is how they were discovered initially.

We know about 150 GCs in the halo. There is some evidence that there are two types, distinguished according to age, metallicity and position. The younger, more metal rich ones have a relatively flattened distribution, and may in fact be associated with the disk. The older, more metal poor ones, are spherically distributed.

There is quite some discussion on the current ages of GCs, with estimates for the oldest ones varying between 12 and 17Gyears (1Gyear$\equiv 10^9$ years). For a long time, these estimates were rather embarrassing for cosmologists, given that this was older than the best estimates for the age of the *Universe*. This discrepancy is now resolved, since the Universe is thought to be older now, and the ages of the clusters have come down a bit.

If the absolute age of the GCs is difficult to estimate, *relative* ages are more secure. And so it is rather more secure to say that there is a relatively large spread in the ages of the MW's GCs. Presumably, this tells us something about how the MW formed.

The density of halo stars (and of GCs as well) falls spherically as

$$n(r) = n_0(r/r_0)^{-3.5}\,, \tag{3.3}$$

and extremely distant field stars have been detected out to $r = 50$kpc. The total mass is actually quite low, $M \sim 10^9 M_\odot$ and luminosity $\sim 10^9 L_\odot$, or only 10 per cent of that of the disk.

3.2.5 The dark matter halo

We'll discuss evidence for dark matter later, but it appears to be the case that most, may be as much as 90 per cent, of the mass in galaxies is actually invisible. We don't know what it is made of, but this matter does not emit or absorb any light – hence *dark* matter, yet there seems to be a lot of it.

3.3 Metallicity of stars

Astronomers call all elements more massive than Helium[5] 'metals', and denote them by Z (X and Y being the hydrogen and helium abundance by mass, respectively). These are produced in stars. For some elements, like e.g. Carbon, we don't really know which stars are the dominant source.

Super Novae (SNe) are a major source of metals. There are two basic types: SNe of type II are explosions of very massive stars. Nuclear fusion during the explosion is responsible for many elements of the α-type, like Oxygen and Silicon[6]. SNe of type I, on the other hand, are thought to occur during mass-transfer in binary stars – these are the dominant producers of Iron, Fe. The chemical mix of the stars is therefore a fossil relict of the relative fractions of type I and type II SNe. Typically though, one measures just one metallicity, for example the Fe abundance since that is often the easiest to measure, and then assumes that the other elements scale with Fe. But we know for a fact that some stellar systems have a rather different relative abundance pattern of the elements.

[5]The Big Bang produced hydrogen, helium, and trace amounts of more massive elements

[6]And they produced the stuff we are made of

For the Sun, the total amount of metals by mass, is about $Z_\odot = 0.02$, or 2 per cent of the mass in the solar system is *not* hydrogen or helium. This is inferred not just from observations of the Sun, but also from the composition of comets.

For other stars, one usually compares the metallicity in units of the solar value, on a logarithmic scale,

$$[Fe/H] \equiv log_{10}\left[(Fe/H)/(Fe/H)_\odot\right] \; . \tag{3.4}$$

So if a star has [Fe/H]=0, it has the same Iron abundance as the Sun, for [Fe/H]=-1, it has one tenth the solar value.

One might expect the metallicity of a star to be related to when it formed. Indeed, very old stars formed before there had been many generations of stars[7], and hence before many SNe exploded, and so would have been formed from gas containing mostly hydrogen and helium. But stars forming now, will contract from gas that has already been polluted by SNe, and hence will be more metal rich. Within the MW, this seems to be born-out, at least to some extent. Stars in old GCs, for example, typically have [Fe/H]~ -1, so quite a bit below the solar value. And most of the old stars in the halo also have low metallicity. These are called *population II* stars. In contrast, stars in the disk usually have higher metallicity, and are called *population I.* There is no strict divide between those, some disk stars also have low Z for example.

Recently, there has been interest in the very first generation of stars that formed after the Big Bang. Those would have $Z = 0$! They are called *population III.* The halo star with the lowest metallicity currently known, has [Fe/H]≈ -5, and might well be one of the first stars to have formed in the MW.

The evolution of Z within the MW, or within galaxies in general, is called their chemical evolution (which is a misnomer, since the elements are produced in nuclear reactions which do not involve chemistry).

3.3.1 Galactic Coordinates

When describing the position of a star within the MW, it is useful to use Galactic Coordinates (l, b), $0^o \leq l \leq 360^o$, and $-90^o \leq b \leq 90^o$. From the vantage point of the Sun, the direction to the galactic centre is taken to be

[7] The Sun is thought to be a third generation star.

Table 3.1: Disks

	Neutral Gas	Thin Disk	Thick Disk
$M/10^{10}M_\odot$	0.5	6	0.2 to 0.4
$L_B/10^{10}L_\odot$		1.8	0.02
$M/L_{\rm B}$ $(M_\odot/L_\odot)$		3	
Diameter (kpc)	50	50	50
Distribution	$\exp(-z/0.16\text{kpc})$	$\exp(-z/0.325\text{kpc})$	$\exp(-z/1.4\text{kpc})$
$[FeH]$	> 0.1	-0.5 – 0.3	-1.6 – -0.4
Age (Gyr)	0 – 17	< 12	14 – 17

Table 3.2: Spheroids

	Central Bulge	Stellar Halo	Dark Matter Halo
$M/10^{10}M_\odot$	1	0.1	55
$L_B/10^{10}L_\odot$	0.3	0.1	0
$M/L_{\rm B}$ $(M_\odot/L_\odot)$	3	~ 1	-
Diameter (kpc)	2	100	> 200
Distribution	bar	$r^{-3.5}$	$(a^2 + r^2)^{-1}$
$[FeH]$	-1 – 1	-4.5 – -0.5	-1.6 – -0.4
Age (Gyr)	10 – 17	14 – 17	17

$l = 0$, $l = 180^o$ is the ante-centre, and $b = 0$ corresponds to the galactic plane. $b = \pm 90^o$ to the galactic poles. Use spherical trigonometry to convert between (l, b) and right ascension α and declination δ. The celestial equator is inclined wrt the plane of the MW, hence the planes $\delta = 0$ and $b = 0$ do not coincide.

3.4 Summary

After having studied this lecture, you should be able to

- Describe how radio and infra-red observations are being used to clarify our picture of the MW.
- Explain how the Hipparcos satellite was a major step in setting the scale of the MW, by performing accurate parallax and proper motions measurements, and fixing a reference frame with respect to distant objects.
- Describe the main components of the MW, thin and thick disk, bulge and bar, stellar halo, and their typical properties such as their density and luminosity profiles.
- Explain what is meant by the metallicity of a star, and know what is meant by the notation [Fe/H]=-1.
- Know how the Galactic Coordinate system (l, b) is defined, know the direction to the galactic centre $l = b = 0$, and the galactic plane $|b| < 5^o$.

Chapter 4

The Interstellar Medium

The interstellar medium (ISM) is the *stuff* between the stars: gas, dust, magnetic fields and cosmic rays. Stars form in molecular clouds, regions of the ISM which are dense and cold. During their lifetimes, they return material back to the ISM in the form of stellar winds and Planetary Nebulae, and at their deaths during super nova explosions. The detailed interplay between stars and the ISM is complicated and not well understood. Here we will discuss briefly how the gas and dust can be observed, describe the Jeans criterion for the collapse of gas into stars, and describe the physics of ionised regions of gas called HII regions.

4.1 Interstellar dust

Dust particles interact with light both through *scattering* and *absorption*. In both cases, there is a reduction in the amount of starlight you receive, described by Eqs. (2.5).

Before I describe the physics, let me tell you that you already know most of this. The molecules in the earth's atmosphere scatter light from the Sun. Because of the quantum mechanical properties of the molecules, they scatter blue light more efficiently than red light.

The colour of the sky during the day is blue, because it is (predominantly) the blue light from the Sun that is scattered toward you. At sunset or sunrise, the sun light passes through a lot more atmosphere than at midday. The increased path length causes the enhanced reddening of the sun. This is partly due to scattering, partly due to absorption on little dust grains. The same processes of scattering and absorption also occur in the ISM.

Scattering changes the direction of the incoming photon, but not its energy. This type of reflection can also induce polarisation of the light, if the dust grains are aligned in a given direction, for example due to a magnetic field. You know that scattering can cause polarisation: expensive sunshades are polarised, so as to block incoming polarised light reflected from the ocean, or road surface for example. So dust grains may scatter some of the light coming from a distant star out of the line-of-sight, thereby reducing the amount you detect, and hence decreasing the apparent luminosity of the star. Given the typical size of interstellar dust grains, blue light is scattered more than red light, and hence scattering also leads to reddening. Now if you look at a dust cloud, you may actually detect some of the light scattered off dust from a nearby bright star for example. And so such a *reflection nebulae* is typically blue (so for the same reason that the sky is blue, except it's scattering by dust (for the reflection nebula) vs by molecules (for the earth's atmosphere)).

When a dust grain *absorbs* a photon, it can sometimes undergo a change in structure, for example a molecular bond is broken or an atom or molecule gets into an excited state. The energy of the photon is converted into another form of energy, and is lost to us: the photon is *absorbed*. Again, the effect of absorption is to decrease the amount of light we detect from the star. Very often, the dust grain will emit other photons, but at much longer wavelengths, typically in the IR. For example in star forming regions, young stars produce large amounts of dust. Even though these stars may be very bright, we cannot see them at visual wavelengths because of the dust absorbs their light. The absorbed light is re-radiated in the IR, making the region very bright in the IR. And so you understand that the newly planned space-telescope which will work in the IR, will teach us a lot about star forming regions.[1]

The scattering and absorption by dust grains may seem to be just a nuisance for astronomers, given that they decrease the amount of starlight we observe. But because the absorbing properties depend on the size and composition of the dust, we can learn about dust properties by studying

[1]Many molecules in the earth's atmosphere, for example water, absorb infrared light in their molecular bands, which is why it is difficult to perform infrared observations from the ground. Some IR observatories have moved to the South pole, where there is much less water vapour.

their effect on star light. And so this is how we infer the typical size and composition of the grains. Although the amount of absorption typically decreases with increasing wavelength, there is much more absorption at some specific *resonances*, which teach us that some grains have graphite in them (graphite, like diamond, is a phase of carbon), and also silicates (we detect the resonance at the energy of the Si-O chemical bond).

4.2 Interstellar gas

We detect hydrogen gas in the ISM in the form of neutral, ionised and molecular form. In astronomy, these are denoted as HI (neutral), HII (ionised), and H_2 (molecular). This is a slightly confusing convention: you would think HII to be double ionised hydrogen – impossible of course – it means H^+. And when talking, you cannot distinguish HII from H_2 – both are pronounced as H-two. So, CIV is *triply* ionised carbon.

How much is there of each type? And more importantly, *why* do we sometimes find one form, sometimes the other?

4.2.1 Collisional processes

Suppose you have a collision between two H_2 molecules. If the collision is sufficiently violent, you can imagine that it will break the molecular bond and you have converted $H_2 \rightarrow 2HI$. Similarly, if you collide two hydrogen atoms (or a HI with an electron) with enough speed, it may lead to ionisation, HI+e→HII+2e. Since the energy required to break the molecular bond (11eV) is lower than to ionise hydrogen(13.6eV), the second collision needs to be more energetic than the first. And so you expect that *the higher the temperature of the gas – and hence the more energetic the collisions – the more you will shift from molecular, to atomic, to ionised gas.*

How about the reverse process? When an HI atom collides with another HI, or with an electron and gets ionised, then the system has converted kinetic energy into ionisation energy. When the ion recombines, HII + e → HI, it can be that the photon escapes from the cloud, in which case there has been a net loss of energy for the gas. Since the particles are now moving slower, the temperature of the gas $T \propto \langle v^2 \rangle$ is lower, or the gas has cooled down. This process is called *radiative cooling* and it is very important for star formation.

So we understand now why colder gas will tend to be molecular, whereas hotter gas is more likely to be atomic or even ionised. Now, if the *pressure* between these phases is nearly the same, then the colder gas (in which H_2 is favoured) will need to be denser than the hotter gas (atomic or, at higher temperature, ionised). And so you expect to find dense, cold clouds to be molecular, and hot, rarefied gas to be ionised – which is indeed what we observe, except that there is one more important process: photo-ionisation.

4.2.2 Photo-ionisation and HII regions

Interaction between a photon, from a nearby star say, and a molecule or atom, may lead to dissociation of the molecular bond, and/or ionisation of the atom. And so the process of molecule formation, which makes a cloud cool, get denser, and eventually form stars, can be completely reversed when the newly born star starts to ionise and heat its surroundings.

Recall from the first part of the course that massive stars, like O and B-type stars, are hot and hence emit lots of hydrogen ionising photons. These photons have a dramatic effect on the surrounding gas, and convert the hydrogen into its HII form. And so massive bright stars are often surrounded by *HII regions*. These regions of ionised Hydrogen gas delineate spiral arms, because star formation occurs predominantly in spiral arms and massive stars do not live for very long, and so are found close to where they form.

HII regions are important, not only because they are beautiful to look at[2], but because their physics is reasonably well understood. And by observing the relative strengths of the emission lines that occur in their spectra, one can deduce the properties of the HII region, such as its temperature, density, and the relative abundance of the elements.

You should remember two things about this, (1) understand the physics that determines the properties of the *spectrum* that you get from such a nebula, and (2) understand *what sets their size* (the so-called Strömgren radius)
Nebular spectra When an ion (for example HII, OIII, ...) recombines with an electron (and form HI, respectively OII), the electron does not necessarily have to fall directly to the lowest possible energy level, i.e. the ground

[2]One of the more famous HII regions is the Orion nebula, M42.

state. Taking the example of HI, the ground state would correspond to the electronic $n = 1$ state. Typically, the electron will cascade down to $n = 1$, and the relative probability for the intermediate steps can be computed from quantum mechanics. However, *this does not translate directly into the spectrum you observe*. For example, suppose the electron makes a transition $n = 2 \to n = 1$, and emits the corresponding photon. As this photon starts to move toward us, it may actually interact with another neutral hydrogen atom, and cause the electron of that atom to be excited from the $n = 1$ to $n = 2$ level. In which case the photon does not actually exit from the nebula! Suppose on the other hand, the photon makes an $n = 3 \to n = 2$ transition. This photon has much more chance of leaving the nebula, since most of the neutral HI will be in the $n = 1$ state, and not so much in the $n = 2$ state. And so that photon will escape from the nebula. So curiously, it may be that lines with a low quantum mechanical probability, dominate a nebular spectrum, since photons from more likely transitions are unable to escape. This is especially true for Planetary Nebulae spectra.[3]

For an HII region, the dominant wavelength photon that escapes results from the $n = 3 \to n = 2$ transition, denoted as Hα[4]. This red line is the reason HII regions appear to fluoresce red. And by observing galaxies through a filter that only lets Hα light through, one can easily find HII regions.

Strömgren spheres. Suppose a source of ionising photons such as a hot star, starts emitting ionising photons at a rate $\dot{N}$, in photons per second, and assume the source is surrounded by a homogeneous cloud of atomic hydrogen, with density n (in HI atoms per cm^3 say). The source will quickly ionise all hydrogen close to it. Let $R(t)$ be the radius of the *ionisation front*, within which most of the hydrogen is ionised, and outside of which the gas is mostly neutral. As R increases between R and $R + \Delta R$, the number N of atoms that need to be ionised is

[3] The strongest line from a Planetary Nebula is O[III], which is a *forbidden* transition. This means that the quantum mechanical probability is zero, and such lines are not observed in laboratory environments. The reason that the line does occur is because a collision with another particle makes the transition possible. But because the transition is forbidden, once the photon is produced, it has no trouble escaping the nebula.

[4] Recall that transitions to $n = 1$ are called the Lyman series and to $n = 2$ the Balmer series. Furthermore, a transition $n + k$ to n are 'counted' with the Greek alphabet, and so $n = 3 \to n = 2$ is called Hα, $n = 4 \to n = 2$ is Hβ, $n = 2 \to n = 1$ is Lymanα, $n = 3 \to n = 1$ is Lymanβ, etc.

$$(4\pi/3)n[(R+\Delta R)^3 - R^3] \approx 4\pi n R^2 \Delta R\,. \tag{4.1}$$

Since it takes the source a time $\Delta t = N/\dot{N}$ to produce this many photons, we find that the speed, v, of the front is

$$v = \frac{\Delta R}{\Delta t} = \frac{\dot{N}}{4\pi n R^2}\,. \tag{4.2}$$

Of course, this speed cannot be faster than the speed of light, and you see that, as the HII regions grows in size, the speed with which it grows decreases $\propto 1/R^2$.

Eventually some of the HII ions inside the ionisation front will start to recombine. Since extra photons are needed to re-ionise these, the speed of the front will start to decrease even more. Eventually, an equilibrium is reached, in which the number of photo-ionisations within R equals the number of recombinations. The stalling radius is called Strömgren radius, R_S. To compute R_S, consider a small volume of the HII region. Since a recombination is an interaction between an electron and an HII ion, the *rate* at which HII ions recombine is proportional to product of electron and ion density:

$$\text{recombination rate}{=}\alpha n_e N_{\rm HII}\,. \tag{4.3}$$

Since the recombination rate is in ions s^{-1} volume^{-1}, α has dimensions of volume s^{-1}. If the gas is composed purely of hydrogen, and is very highly ionised, then $n_e \approx N_{\rm HII} \approx n_{\rm H}$, where $n_{\rm H}$ is the density of hydrogen, either HII or HI. The total number of recombinations within radius R is then $(4\pi/3)\alpha n_{\rm H}^2 R^3$. In equilibrium, this is the rate $\dot{N}$ at which the source produces ionising photons, hence

$$R_S = (\frac{3\dot{N}}{4\pi\alpha})^{1/3} n_{\rm H}^{-2/3}\,. \tag{4.4}$$

For example, assume $n_{\rm H} = 5\times 10^3{\rm cm}^{-3}$ for the density of the cloud, and $\dot{N} = 10^{49}{\rm s}^{-1}$ for the ionisation rate of the star. Then $R_S \approx 0.21$pc, since $\alpha \approx 3.1\times 10^{-13}$ cm^3 s^{-1} at a temperature of $T = 8000K$ typical of HII regions.

4.2.3 21-cm radiation

We have just seen that you can observe ionised gas in HII regions because of its recombination radiation. So how can we observe HI?

A hydrogen atom is composed of a proton and an electron. Both these particles have a spin degree of freedom. Spin is a purely quantum mechanical concept, but there is some analogy with a bar magnet. So in a hydrogen atom, the spins of proton and electron can either be aligned or anti-aligned. In the bar-magnet analogy: either the north-poles of both magnets point in the same direction (aligned) or in the opposite direction (anti-aligned).[5] And just as in the bar-magnet case, there is a small energy difference between the two states, with the aligned state having higher energy. Transitions between states of different spins are also called hyper-fine interactions.[6]

When an HI atom with aligned spins is left on its own, it has some small probability for a spontaneous spin reversal to the lower energy state. When it does so, it emits a photon with wavelength 21-cm. The mean time it takes an HI to perform such a spin reversal is several million years. This makes it almost impossible to observe this transition in the laboratory, because collisions between the particles will de-excite the aligned spin well before there has been a spontaneous transition. But in the ISM, densities are far lower than can be generated in the lab. The 21-cm line is of great observational value. First of all, there is enough HI to observe the line. And once the line is produced, its low probability becomes an advantage, since the 21-cm radiation can pass through a lot of HI gas, without it being absorbed again. Given the long wavelength, 21-cm radiation is not absorbed by dust.

An external magnetic field shifts the energy levels (called the Zeeman effect). And so 21-cm radiation can be used to estimate the strength of the magnetic field in the ISM.

It was the Dutch astronomer Jan Oort who suggested one of his students, van der Hulst, to compute the wavelength expected for this transition. (Wouldn't you like to be given such a project?). When it was clear that it could be observed, he asked the government for money to build a radio-telescope to go an observe it.

---

[5]Two bar magnets can of course have any angle between them. But for quantum mechanical spins, this is impossible, and they are either fully aligned, or fully anti-aligned, with nothing in between! So the analogy only goes so far.

[6]Interactions between the spin and the *orbital* angular momentum of the electron also result in slightly different energy levels and transitions.

4.2.4 Other radio-wavelengths

We've seen how we expect gas to become mostly molecular in high density regions. And so we cannot observe these with 21-cm observations, since also the hydrogen will become molecular. Fortunately, many molecules also emit radiation at radio-wavelengths. Typically these are caused by rotational transitions.[7]

Which transition is strongest depends a lot on the density of the cloud. The most commonly studied transition is from the CO molecule, with wavelength 2.6cm. At higher densities, other molecules such as CH, OH and CS become observable.

The result of many years of investigation is that the properties within gas clouds can vary widely. In most of the outer parts of the disk, densities are low, and most gas is in the form of HI with typical density of order $1\mathrm{cm}^{-3}$. Closer in, we find a variety of clouds within the HI gas, which differ in total mass, and density. *Giant Molecular Clouds*, or GMCs for short, are enormous complexes of gas and dust, with total masses up to $10^6 M_\odot$, temperatures $T \sim 20$K, and densities 100-300cm^{-3}. These clouds often have sub-condensations which are much denser. Thousands of GMCs are known in the MW, mostly in the spiral arms.

It is thought that most, if not all, star formation in the MW occurs in GMCs. To understand what determines whether a cloud can remain stable, or will become unstable and undergo star formation, brings us to the next concept: the Jeans mass.

4.2.5 The Jeans mass

I trust you have already encountered the concept of Jeans mass when studying stars. Let me just phrase the same thing slightly differently, starting with the difference between stable and unstable equilibria.

When a system is in equilibrium, forces balance. In a *stable* equilibrium, small changes remain small, whereas in an *unstable* equilibrium, small changes are amplified and grow.

For example, consider small perturbations in an infinite, self-gravitating fluid, with density ρ and temperature T. If the scale λ of the perturbation

[7] Energy can be stored in rotation of the molecule. In a quantum mechanical description, the amount you can store is quantised. Rotational transitions are transitions between different rotation speeds of the molecule.

is small, then you know what happens: the perturbation will be a sound wave. This is why you can hear what I am saying. But what happens for perturbations on larger scales? I'll explain why eventually gravity will become dominant.

Consider the material inside one perturbation length λ. The thermal energy K enclosed within it, is of order $K \sim M \mathrm{k_B} T$, where $M = \rho\lambda^3$ is the mass within the perturbation (I'm neglecting constants of order unity, like the $4\pi/3$ and such – we'll do better later). The gravitational energy U within the perturbation, is of order $U \sim G\,M^2/\lambda \sim G\,M\rho\lambda^2$. For sufficiently small λ, the gravity energy is small with respect to the thermal energy, and can be neglected: these are the usual sound waves. But for larger λ, the ratio $U/K \sim \lambda^2$ gets bigger, and so eventually, gravity will dominate! The Jeans mass is the critical mass above which gravity dominates. For perturbations below the Jeans mass, pressure forces dominate, and so the perturbation will re-expand when being compresses. But for perturbation more massive than the Jeans mass, gravity dominates, and so a perturbation will collapse even further when compressed, leading to run-away collapse.

This derivation follows CO.[8] Putting in all the constants, we find that K and U are given by[9]

$$\begin{aligned} K &= M\frac{\mathrm{k_B}T}{(\gamma-1)\mu\mathrm{m_H}} \\ U &= -\frac{3}{5}\frac{GM^2}{\lambda}\,. \end{aligned} \tag{4.5}$$

Here, $\mathrm{k_B}$ is Boltzmann's constant, γ the ratio of specific heats of the gas (e.g. $\gamma = 5/3$ for a mono-atomic gas), and μ the mean molecular weight ($\mu = 1$ for a pure neutral hydrogen gas, $\mu = 0.5$ is the hydrogen is fully ionised). K is the product of the mass of the cloud, M, with the thermal energy per unit mass, $u = (\mathrm{k_B}T)/((\gamma-1)\mu\mathrm{m_H})$. The potential energy $U \sim GM^2/\lambda$. The factor in front, 3/5 is appropriate for a spherical, homogeneous density perturbation. Now, in virial equilibrium, $2K = |U|$, So whenever $|U|$ is larger, gravity dominates, and so the critical length λ_J is whenever

[8] In different books, you'll find slightly different numerical pre-factors in the definition of the Jeans length and Jeans mass.

[9] An even better way would be to find the dispersion relation that relates the speed of a wave to its wavelength.

$$2K = |U|\,. \tag{4.6}$$

Now we can use $M = (4\pi/3)\rho\lambda^3$ in Eqs.(4.5+6.1) to obtain the critical Jeans length λ_J, and the Jeans mass M_J as

$$\lambda_J = \left(\frac{5\mathrm{k_B}T}{2\pi(\gamma-1)\mu\mathrm{m_H}\rho G}\right)^{1/2} \tag{4.7}$$

$$M_J \equiv \frac{4\pi}{3}\rho\lambda_J^3 = \left(\frac{10\mathrm{k_B}T}{3(\gamma-1)\mu\mathrm{m_H}G}\right)^{3/2}\left(\frac{3}{4\pi\rho}\right)^{1/2}. \tag{4.8}$$

So note that in a homogeneous fluid, *the Jeans mass is the mass contained in a volume with radius the Jeans length. The latter is such that perturbations larger than the Jeans length will collapse under their own gravity.*

Fragmentation A nice application of this concept is to investigate what happens during collapse of a cloud. Let's say that the Jeans mass is given by $M_J = AT^{3/2}/\rho^{1/2}$, where I've assembled all constants into a new constant A. Now assume you have a cloud with mass $M = M_J$ that starts to collapse, and hence ρ increases. In general T will rise as well, and so M_J will change. Now suppose M_J *decreases* to M_J'. Following our earlier discussion, this would mean that, if there were smaller perturbations within the cloud already (substructure), then those with masses larger than M_J' will start collapsing on their own – the cloud may fragment. Given our expression for the Jeans mass, $M_J' < M_J$ will happen whenever $T^{3/2}$ increases slower than $\rho^{1/2}$. Let's parametrise the dependence of T on ρ as $T \propto \rho^{\gamma-1}$. Then $M_J' < M_J$ requires $\gamma < 4/3$. For adiabatic, mono-atomic gas, $\gamma = 5/3$ and you don't expect fragmentation. But if radiative cooling can keep the gas isothermal, $\gamma = 1$ and you expect fragmentation.

So you've learnt how important cooling is for the formation of stars. After all, the mass of the GMCs in which stars form, is far bigger than stellar masses, and so we need to understand why stars are so much smaller than GMCs.

4.3 Summary

After having studied this lecture, you should be able to

- Describe how we know the properties of interstellar dust from scattering and absorption of star light.
- Explain why we find different ionisation states of interstellar gas, depending on density, temperature, and ionising background.
- Derive the Strömgren radius for an HII region, and explain the concept.
- Explain the origin of the hydrogen 21-cm line, and explain its importance in understanding the structure of the MW.
- Explain the concept of Jeans mass, and its relation to fragmentation of clouds.

Chapter 5

Dynamics of galactic disks

The stars in the Milky Way disk are on (almost) circular orbits, with gravity balancing centripetal acceleration. Given that most of the light of the disk comes from its central parts, we would expect the circular velocity in the outer parts of the disk to fall with distance as appropriate for Keplerian motion. We should also be able to compute how velocities of stars in the solar neighbourhood depend on direction. Observations do not follow these expectations at all, which leads to the startling conclusion that most of the mass in the Milky Way is invisible.

5.1 Differential rotation

5.1.1 Keplerian rotation

Suppose you want to describe the orbit of a star in the outer parts of the MW disk. Since most of the light is then enclosed within the orbit, you could expect that most of the mass would be interior to the orbit as well. Newton's law then tells us that there is a relation between the circular velocity, V_c, the radius, R, of the orbit, and the mass, M, of the MW,

$$\frac{V_c^2}{R} = \frac{GM}{R^2}\,. \tag{5.1}$$

Newton's law guarantees us that the force GM/R^2 is independent of the actual density distribution of the MW, as long as it is cylindrically symmetric.

The run of circular velocity with radius, *i.e.* the function $V_c(R)$, is called the *rotation curve* of the MW. And so we expect $V_c(R) \propto R^{-1/2}$, or the disk is in *differential rotation*[1], with V_c decreasing with increasing R. Since the period P of the orbit $P = 2\pi R/V_c$, we expect that $P \propto R^{3/2}$, just as is the case for the motion of planets around the Sun.

If the stars around the Sun are on circular orbit, then there are relations between the relative velocity of the star, its distance, and its direction in the disk. These are described by Oort's constants.

5.1.2 Oort's constants

Suppose you're on a circular orbit with radius R_0 around the MW, and you measure the velocities of stars around you. For simplicity, assume all stars are also on circular orbits. Stars on orbits with radius $R < R_0$ will go faster than you do, and so will over take you, and the opposite for stars with $R > R_0$. Jan Oort computed how the velocities you measure for these stars depend on their distance and the direction you see them in.

For the geometry, look at Figure (5.1)[2]. To keep the notation the same, let's call the circular velocity at the position of the observer $V_{c,0}$, and at the position of the star V_c. The radial and tangential velocity of the star with respect to the observer (at the position of the Sun) are

$$\begin{aligned} V_r &= V_c\cos(\alpha) - V_{c,0}\sin(l) \\ V_t &= V_c\sin(\alpha) - V_{c,0}\cos(l)\,. \end{aligned} \tag{5.2}$$

Now, in the right-angled triangle on the figure, you can convince yourself that

$$\begin{aligned} d + R\sin(\alpha) &= R_0\cos(l) \\ R\cos(\alpha) &= R_0\sin(l) \\ R_0 &= d\cos(l) + R\cos(\beta) \approx d\cos(l) + R\,\text{when}\ d \ll R_0\,, \end{aligned} \tag{5.3}$$

where β is the angle Sun-MW Centre-Star. Combining these equations gives

[1]The inner parts go around faster than the outer parts, so differential rotation, as opposed to say solid body rotation.

[2]We are describing the motion of stars *in the plane of the disk*, hence the galactic coordinate $b = 0$

$$
\begin{aligned}
V_r &= (\Omega - \Omega_0) R_0 \sin(l) \\
V_t &= (\Omega - \Omega_0) R_0 \cos(l) - \Omega\, d\,,
\end{aligned} \tag{5.4}
$$

where $\Omega_0 = V_{c,0}/R_0$ is the angular velocity of the observer, and $\Omega = V_c/R$ the angular velocity of the star. If $|R - R_0|$ is small, we can perform a Taylor expansion to find $\Omega \approx \Omega_0 + (d\Omega/dR)(R - R_0)$ and substitute this in the previous expression. Finally we can introduce Oort's constants

$$
\begin{aligned}
A &\equiv -\frac{1}{2}\left[\frac{dV_c}{dR}|_{R_0} - \frac{V_{c,0}}{R_0}\right] \approx 14.4 \pm 1.2 \mathrm{km\,s^{-1}\,kpc^{-1}} \\
B &\equiv -\frac{1}{2}\left[\frac{dV_c}{dR}|_{R_0} + \frac{V_{c,0}}{R_0}\right] \approx -12.0 \pm 2.8 \mathrm{km\,s^{-1}\,kpc^{-1}}\,.
\end{aligned} \tag{5.5}
$$

We'll discuss in a moment how the measured values are found. With these definitions we obtain after some algebra,

$$
\begin{aligned}
V_r &= A\, d\, \sin(2l) \\
V_t &= A\, d\, \cos(2l) + Bd\,.
\end{aligned} \tag{5.6}
$$

Let's see what this means. For a star toward the galactic centre or anti-centre ($l = 0$ and $l = 180^o$ respectively), we find the radial velocity to be zero, and the tangential velocity to be $(A+B)\, d$. For example with $d = 1$kpc, $V_t \approx 2.4$km s^{-1}.

To measure A and B, we need to measure the radial and tangential velocities of stars of known distance.

Finally, let's see what we would expect to find for a Keplerian disk, for which $V_c \equiv V_{c,0}(R_0/R)^{1/2}$, and hence $dV_c/dR = -(1/2)V_c/R$. At the position of the Sun, we know $V_{c,0} \approx 220$km s^{-1} and $R_0 \approx 8.5$kpc hence

$$
\begin{aligned}
A_{\mathrm{Kepler}} &= \frac{3}{4}\frac{V_{c,0}}{R_0} = 19.4 \mathrm{km\ s^{-1}\,kpc^{-1}} \\
B_{\mathrm{Kepler}} &= -\frac{1}{4}\frac{V_{c,0}}{R_0} = -6.5 \mathrm{km\ s^{-1}\,kpc^{-1}}\,.
\end{aligned} \tag{5.7}
$$

This does not compare well with the measured values in Eqs.(5.5). Assuming we got V_c/R correct, let's compare the measured and computed values for dV_c/dR,

$$
\begin{aligned}
\frac{dV_c}{dR}|_{\text{measured}} &= -(A+B) = -2.4 \pm 3.0 \text{km s}^{-1}\,\text{kpc}^{-1} \\
\frac{dV_c}{dR}|_{\text{Keplerian}} &= -\frac{1}{2}\frac{V_c}{R} = -13.0 \text{km s}^{-1}\,\text{kpc}^{-1}\,. \qquad (5.8)
\end{aligned}
$$

So what appears to be wrong is that the circular velocity, as computed for Keplerian fall-off, decreases much more rapidly with increasing R, than the values measured from Oort's constants. We've assumed $R_0 = 8.5$kpc, but for the two values to agree, we would require R_0 to be larger by a factor 6! Clearly, the discrepancy is much larger than the uncertainties in our values of either V_c or R_0. We shall see later that there is other evidence that suggests a much more dramatic failure of our underlying assumptions.

We've so far assumed stars to be on exactly circular orbits. This is of course only an approximation, and real stars will have velocities which differ slightly from $V_c(R)$: they will have (small) *peculiar* velocities in all three Cartesian directions. The reference system that is on a circular orbit with velocity $V_c(R)$ is called the *local standard of rest* (LSR)[3], and, for example the Sun moves with a velocity of ~ 16km s^{-1} with respect to its LSR. This is called the *solar motion*. To determine A and B, we need to measure V_r and V_t as a function of d and l *statistically*, and take into account that we also need to determine the solar motion.

When Oort did his measurements of the constants A and B, he discovered that some stars had very large deviations from the expected values of V_r and V_t. He called them *high velocity stars*. He also correctly identified their origin, they are stars in the MW halo. Unlike the disk, the halo does not rotate, and so the high velocities of these stars are due to the rotation velocity of the Sun.

5.1.3 Rotation curves measured from HI 21-cm emission

The HI 21-cm emission is able to penetrate virtually the entire galaxy, and so can be used to measure the rotation curve $V_c(R)$. The biggest problem

[3]Note that this is not an inertial reference frame!

with this is how to determine d, the distance to the cloud that produces the absorption.

A solution is to recognise from Eqs.(5.2) that V_r has a *maximum*

$$V_{r,\max} = V_c - V_{c,0}\sin(l)\,, \tag{5.9}$$

for any l with $|l| < \pi/2$, which occurs for $\alpha = 0$, at which point

$$R = R_0 \sin(l)\,, \tag{5.10}$$

is a minimum, and [4] $d = R_0\cos(l)$. Combining these, we find

$$\frac{dV_c(R)}{dR} = \frac{dV_{r,\max}(l)}{dl}\,/R_0\,\cos(l) + V_{c,0}/R_0\,. \tag{5.11}$$

Performing this analysis essentially confirmed Oort's measurements, i.e. $|dV_c/dR|$, the rate of change of the circular velocity with R, is far smaller than you'd expect for a Keplerian rotation curve. In fact, from measurements of Cepheids with $R > R_0$ it became apparent that $dV_c/dR \approx 0$, i.e. **the Milky Way's rotation curve is nearly flat**, or $V_c(R) \approx$ const, as opposed to $V_c \propto (M/R)^{1/2}$ for Keplerian fall-off – it is as if there is considerable mass outside the solar circle.

5.2 Rotation curves and dark matter

If the rotation curve is flat, $V_c(R) = \textit{constant}$, it means there must be more mass in the outskirts of spirals than we've assumed so far. Let's compute how much more.

Assume the density ρ of the MW is $\rho = \rho_0(R_0/R)^\alpha$, where ρ_0 and R_0 are constants. For such a density distribution, the mass within a given radius R, $M(<R) = \int_0^R 4\pi\rho(R)R^2dR \propto R^{3-\alpha}$. Computing $V_c(R)$, you easily find that a flat rotation curve requires $\alpha = 2$, hence $\rho(R) \propto 1/R^2$ is required for a flat rotation curve. And so, although the *light density* drops significantly in the outer parts of the MW, the *mass density* drops much slower with increasing R.

More recently, rotation curves have been measured using the 21-cm line in many other spiral galaxies, with always the same result: although the

[4] This only works when $|l| < \pi/2$, i.e. toward the centre, since in the outer parts there is no unique orbit that you can associate with a given velocity.

rotation curve rises from the centre outward, at sufficiently large radius R_f it invariably becomes flat, even though most of the light is enclosed within R_f. So why don't we see this mass? Two solutions seem possible. Firstly, for some strange reason, the mass-to-light ratio of the stars outside R_f increases very much, so that the little amount of light we see, actually represents much more mass than we think. Or secondly, there is some type of invisible mass – dark matter – that dominates the mass-density in the outer parts of spiral galaxies.[5] Both seem rather contrived, yet evidence for dark matter seems to reoccur again and again in other places as well. Note that this dark matter needs to be (1) massive, (2) invisible, *i.e.* it does not *absorb* light, nor does it *emit* light, it simply does not interact with photons.

Already in the 1930s, the Swiss astronomer Fritz Zwicky said that the galaxies in a cluster of galaxies move too fast for the amount of matter accounted for in the galaxies themselves. In fact, they moved so fast that gravity could not contain them to remain inside the cluster. So Zwicky postulated that there must be some type of invisible matter inside galaxy clusters. We'll discuss evidence for dark matter in elliptical galaxies, in clusters, and finally in the Universe as a whole. The case is sufficiently strong that experiments have been set-up all over the world, to try to detect this mysterious type of mass in the laboratory.

5.3 The Oort limit

Jan Oort performed another interesting measurement: count stars as function of their height above the disk. From this he was able to estimate how much mass was in the plane of the MW. By counting stars in the plane of the disk as well, he computed a mass-to-light ratio, which was also a bit on the high side, suggesting the presence of dark matter in the disk, even at the position of the Sun.

Here is a simplified way how to do this. First let me prove that for a uniform disk, the force on a star is *independent* of the distance to the disk.

[5] There is a third way out: may be gravity does not behave as expected on these large scales, or for these small accelerations. This is not so easy to dismiss as you might think: we have no measurements on other scales, for example in the solar system, or in the laboratory, that can test the small acceleration regime that applies on galactic scales. The theory of **Mo**dified **N**ewtonian **D**ynamics (MOND) is able to provide very good fits to measured rotation curves with a small modification of gravity that cannot be probed in other regimes.

As an intermediate step, let's compute the gravitational acceleration on a star, at distance d from the centre of a ring of matter with radius ω and thickness $d\omega$, and of surface density Σ. From symmetry, the acceleration is toward the middle of the ring, and it has magnitude

$$da(d,\omega) = -2\pi\, G\, \Sigma \frac{\cos(\alpha)\omega\, d\omega}{\omega^2 + d^2}\,, \tag{5.12}$$

where α is the angle at the position of the star, between the vertical and the ring, so $\tan(\alpha) = \omega/d$. Now the acceleration due to the whole disk is found by summing over all such rings, hence

$$a(d) = -2\pi\, G\, \Sigma \int_0^\infty \frac{\cos(\alpha)\omega\, d\omega}{\omega^2 + d^2} = -2\pi\, G\, \Sigma \int_0^{\pi/2} \sin(\alpha) d\alpha = -2\pi\, G\, \Sigma\,, \tag{5.13}$$

independent of d! So the equation of motion for the star is

$$\frac{d^2 z}{dt^2} = -2\pi\, G\, \Sigma \equiv -f\,. \tag{5.14}$$

This is just the same as that for a ball thrown vertically on the earth's surface, so the solution is $z = z_0 + \dot{z}_0 t - (1/2) f\, t^2$. From this, one can easily obtain the fraction of stars you expect that have z between $[z_1, z_1 + \delta]$ and between $[z_2, z_2 + \delta]$, which depends only on f. So, if you measure this from your data, you can determine Σ. The value Oort found is called the 'Oort limit' since the MW disk needs to have at least this surface density.

5.4 Spiral arms

Before leaving the subject of spiral galaxies, we should address what is in fact probably the first question you'd like to know about them: *why* do they have spiral arms? So let's start by reviewing what we know about spiral arms.

Firstly, spiral arms are really evident when we look at massive stars and HII regions – these really delineate the spiral structure. But there is more: we also detect the arms in HI, and even better in molecular gas. In fact, GMCs are probably the best tracers of the pattern. So may be they have something to do with star formation?

However, they are also present in the K band where we are probing *old* stars. So first conclusion: it's really the whole galaxy that takes part in the spiral pattern. The *density contrast*[6] is not very large, probably less than a factor of two. The contrast in *surface brightness* can be quite a bit higher though, since the young stars in the arms are massive and hence bright.

The winding problem Given that the disk is in differential rotation, a spiral arm cannot be a material structure (meaning that it is always the same stuff in the arm going round and round) – because of it were, we would have a *winding problem*. To realise why this is so, consider the following thought experiment.

Suppose you paint all stars along a radius in the disk green, and follow how those green stars move (see figure (5.2). If the circular velocity is a constant, V_c, then the period of an orbit is proportional to its radius, $P \propto R$. Now concentrate on two green stars, one at distance R (star 1), the other at distance $2R$ (star 2), having periods $P_1 = P$ and $P_2 = 2\,P$. After a time P, star 1 will be at the original position, whereas star 2 will be at the other side of the galaxy. After time $2P$, both stars will be back at their original position, and so the green line of stars in between them will have circled the galaxy, i.e. *the initially straight line of green stars now encircles the galaxy.* Applying this to stars originally in a spiral arm, it is clear now that very quickly, the spiral arm will wind tighter and tighter. Considering how loosely wound some spiral arms are, this would imply they are all very recently formed, placing us at a peculiar instant in time where spiral structure forms.

In cases where we were are able to determine the sense of rotation of the pattern, it turned out that the spiral arms are *trailing*, that is, at least they are rotating in the same sense as the spiral pattern of our green stars.

So what's the solution? May be it's good to realise that there is quite a variety in spiral patterns. Grand Design spirals have two really well defined relatively open arms. Some spirals have *flocculent* arms, which are not very long (i.e. you can't trace them for very long), but often there are many in the disk. So may be we're looking at more than one possible origin.

The theory of *spiral density waves* suggest that a density wave rotates with a constant pattern speed in the disk. You can think of such a wave as a natural mode of oscillation of the disk, much as you have sound waves in air, or harmonic waves in a violin string. If the rotation period of a star is shorter than of the pattern, then the star will occasionally overtake the wave. The

[6] the relative difference in density at given r from the centre, within vs outside of the arm

gravitational force of the matter in the wave will make the star linger a bit near the pattern, before it leaves the arm again. An analogy often made is that of a slow lorry on the motor way causing a pile-up of faster cars behind it. Although most cars go fast, the speed with which *the pile-up* moves is determined by the slow lorry.

Another process which we *know* to occur, is the triggering of spiral arms by *galaxy encounters*. When two galaxies come close together, they will exert tidal forces on each other.[7] This effect can be convincingly demonstrated with numerical simulations. And so, although these arms may quickly disappear when left on their own (winding problem), the satellite galaxy may continually re-excite the spiral arms. Another hint comes from the fact that in barred galaxies, the spiral arms emanate from the ends of the bar (have a look at NGC 1365 if you don't believe me). It's is thought that Grand Design spirals are either triggered by a tidal encounter, or by the rotating bar.

Goldreich and Lynden-Bell said about flocculent spirals: 'a swirling hochpotch of spiral arms is a reasonably apt description' . So here we're probably faced by some kind of gas-dynamical instability in the disk.

So the conclusion about spiral arms is that there may be several processes that give rise to spiral arms, and depending on the particular spiral galaxy, one of these may dominate. So tidal encounters are probably responsible for the beautiful Grand Design spiral arms. The more messy pattern of flocculent spirals is probably due to a gas-dynamical instability in the disk. And once excited, a spiral pattern may live for a long time due to the spiral-density wave.

[7]The tidal force exerted by the moon causes the tides in the earth's oceans.

5.5 Summary

After having studied this lecture, you should be able to

- Derive the rotation curve for a Keplerian disk
- Derive the equations for the radial and tangential velocity of stars on circular orbits in a disk in differential rotation, and derive expressions for Oort's constants.
- Compute Oort's constants A and B for a Keplerian disk. Explain how A and B are measured in the MW.
- Explain how the 21-cm emission line can be used to estimate the rotation curve of the MW.
- Explain why both Oort's constants, and the rotation curve measured from 21-cm emission, suggest the presence of dark matter in the outer parts of the MW.
- Explain how Oort derived a limit on the amount of dark matter in the MW disk from the vertical motion of stars.
- Describe why spiral arms cannot all be material structures by explaining the winding problem.
- Discuss solutions to the winding problem.
- Explain the density wave theory of spiral arms.

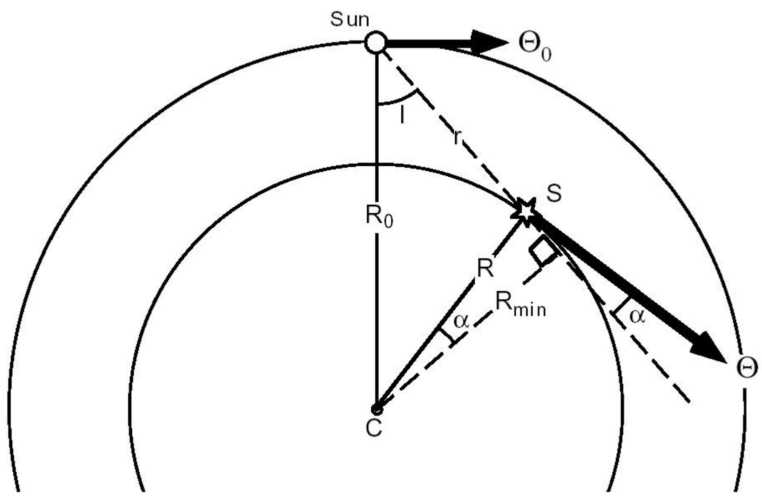

Figure 5.1: (Taken from http://www.csupomona.edu/˜jis/1999/kong.pdf) The observer is moving on a circular orbit with velocity Θ_0 and radius R_0, and observes a star at distance r, and galactic longitude l, on circular orbit with velocity Θ and radius R. In the text, r is called d and the circular velocities are V_c and $V_{c,0}$ for the star and the Sun, respectively.

Figure 5.2: A spiral pattern made out of stars will rapidly wind-up in a disk undergoing differential rotation.

Chapter 6

The Dark Halo

Measurements of the rotation curve from HI 21-cm emission, analysis of the motions of stars in the solar neighbourhood with Oort's constants, and the Oort limit, all suggest the presence of a large amount of invisible 'dark matter' in the MW. Given such a startling conclusion, it may be a good idea to look for other evidence for dark matter in galaxy haloes.

6.1 High velocity stars

A number of high-velocity stars near the Sun have measured velocities[1] up to $v_\star \approx 500\text{km s}^{-1}$. Their existence provides us with a probe of the galaxy's mass, if we assume that these stars are still bound to the MW: it means that *the escape speed*[2] *at the solar circle,* $v_e > v_\star$.

6.1.1 Point mass model

If we assume there is no dark halo, then it is easy to find a relation between the circular velocity V_c and the escape speed v_e. If we assume all mass M

[1]The quoted velocity is wrt to the centre of mass velocity of the MW. Do not confuse these with Oort's high velocity stars, which are typically low mass, low metallicity stars in the *Galactic Halo*. The velocities of Oort's stars are of order 200km s^{-1}. The present high velocity stars are typically *A*-type stars, presumably born in the disk, that have acquired their high velocity following a super nova explosion.

[2]Stars travelling at the local escape speed are just able to leave the potential well and escape to infinite.

of the MW is interior to the solar radius $R_\odot$ [3], then $V_c^2 = GM/R_\odot$, and the potential $\Phi = -GM/R_\odot = -V_c^2$. The velocity of a star just able to escape from the gravitational well is the *escape speed*, v_e. So since that star has zero total energy,

$$0 = E = \frac{1}{2}v_e^2 + \Phi = \frac{1}{2}v_e^2 - V_c^2\,. \tag{6.1}$$

So in this case, $v_e = 2^{1/2}\,V_c \approx 311\text{km s}^{-1}$ (Using $V_c = 220\text{km s}^{-1}$.) So for a point mass model of the MW, these high velocity stars cannot be bound to the MW – suggesting once more the presence of more mass in the outskirts of the MW than expected, given the rapid decrease in the distribution of light.

6.1.2 Parameters of dark halo

Given the failure of the point mass MW model, let's assume there to be a dark halo. In addition, let us assume the MW to be spherical, with a constant circular speed V_c out to some radius $R_\star$. Since $V_c^2 = GM/R =$ constant, we have that

$$\begin{aligned} M(R) &= \frac{V_c^2\,R}{G} \quad \text{when } R < R_\star \\ &= \frac{V_c^2\,R_\star}{G} \quad \text{when } R \geq R_\star\,. \end{aligned} \tag{6.2}$$

Now the equation for the potential Φ is

$$\frac{d\Phi}{dR} = \frac{GM}{R^2} = \frac{V_c^2}{R}\,, \tag{6.3}$$

for $R \leq R_\star$. Integration gives $\Phi = \text{constant} - V_c^2 \ln(R_\star/R)$. We can determine the value of the constant at $R = R_\star$, since then $\Phi = -G\,M/R_\star = -V_c^2$. Hence

$$\Phi(R) = -V_c^2\,\left[1 + \ln(R_\star/R)\right]\,. \tag{6.4}$$

Using Eq.(6.1) for the escape speed, we obtain

[3] $R_\odot$ is the radius of the circular orbit around the MW, at the position of the Sun – the solar galacto-centric radius.

$$v_e^2 = 2\,V_c^2\,[1 + \ln(R_\star/R)]\ . \tag{6.5}$$

For the Sun, $R \approx 8.5$kpc, $V_c = 220$km s^{-1}, $v_e \geq 500$km s^{-1}, implying $R_\star \geq 41$kpc, hence

$$M(R = R\star) \geq 4.6 \times 10^{11} M_\odot\,. \tag{6.6}$$

Since the total luminosity of the MW is of order $1.4 \times 10^{10} L_\odot$ (in the V -band), it means[4] that the mass-to-light ratio is at least 30. Given that the mass-to-light ratio of the MW *stars* is around 3 or so, our reasoning again suggests that a dominant fraction of the MW mass is not in stars.

Another estimate of the MW mass comes from the motion of neighbouring galaxies in the *Local Group*.

6.2 The Local Group

The MW is located in a rather average part of the Universe, away from any dense concentrations of galaxies. The 'Local Group' consist of the MW, Andromeda (M31), and about 40 small, irregular galaxies, gravitationally bound to each other.

6.2.1 Galaxy population

The MW has a couple of small galaxies gravitationally bound to it, for example the Large and Small Magellanic Clouds. The Magellanic stream is a stream of gas which originates from the Large Magellanic Cloud, which can be traced over a significant fraction of the whole sky. It may result from the tidal forces exerted by the MW.

The Sagittarius Dwarf galaxy is the neighbouring galaxy closest to the MW. It was discovered as recently as 1994, having eluded detection since it happens to be at the other side of the MW bulge and therefore is hiding behind a large amount of foreground stars. Sagittarius has ventured so close to the MW that the tidal forces will probably tear it apart in the near future. Its discovery was actually hampered by its close proximity, since the large area on the sky ($\sim 8^o \times 2^o$) made it difficult to pick out against the bright

[4]Remember: we've only derived a lower-limit to the MW mass, hence a lower-limit to the mass-to-light ratio.

MW foreground. Even away from the galactic plane (where dust obscuration limits our view), new members of the LG continue to be identified. The study of these small galaxies is exciting, since they allow us to constrain models for galaxy formation and evolution.

The Andromeda galaxy M31 is about twice as bright as the MW, and is the brightest member of the Local Group (LG). It is very similar to the MW, and so also has its own set of small galaxies orbiting around it. These galaxies, together with another 20 small galaxies, form the LG. M31 and the MW are by far the most massive galaxies in the LG.

The LG is almost certainly gravitationally bound to other nearby groups, and so does not really have a well defined edge. Galaxies in the outskirts of the LG may in fact belong to other groups.

The distribution of galaxies in the LG is rather untidy and lacks any obvious symmetry. Most of the galaxies are found either close to the MW, or close to M31. There is a third, smaller condensation of galaxies, hovering around NGC 3109.

The motion M31 can be used to estimate the mass of the Local Group, using the Local Group *timing argument*.

6.2.2 Local Group timing argument

The dynamics of M31 and the MW can be used to estimate the total mass in the LG. From the Doppler shifts of spectral lines, one can measure the radial velocity of M31 with respect to the MW[5],

$$v = -118\text{km s}^{-1}\,. \tag{6.7}$$

The negative sign means that Andromeda is moving *toward* the MW. This may be surprising, given that most galaxies are moving apart with the general Hubble flow. The fact that Andromeda is moving toward the MW is presumably because their mutual gravitational attraction has halted, and eventually reversed their initial velocities. Kahn and Woltjer pointed out in 1952 that this leads to an estimate of the masses involved.

Since M31 and the MW are by far the most luminous members of the LG we can neglect in the first instance the others, and treat the two galaxies as

[5]What one measures is the radial velocity wrt to the *Sun*. Since the Sun is on a (nearly) circular orbit around the MW, one needs to correct the measured heliocentric velocity to obtain the radial velocity of Andromeda wrt the MW.

an isolated system of two point masses. Since M31 is about twice as bright as the MW, and given that they are so similar, it is presumably also about twice as massive. If we further assume the orbit to be radial, then Newton's law gives for the equation of motion

$$\frac{d^2r}{dt^2} = -\frac{GM_{\rm total}}{r^2}\,, \tag{6.8}$$

where $M_{\rm total}$ is the sum of the two masses. Initially, at $t = 0$, we can take $r = 0$ (since the galaxies were close together at the Big Bang).

The solution can be written in the well known parametric form

$$\begin{aligned} r &= \frac{R_{\rm max}}{2}(1 - \cos\theta) \\ t &= \left(\frac{R_{\rm max}^3}{8\,G\,M_{\rm total}}\right)^{1/2} (\theta - \sin\theta)\,. \end{aligned} \tag{6.9}$$

The distance r increases from 0 (for $\theta = 0$) to some maximum value $R_{\rm max}$ (for $\theta = \pi$), and then decreases again. The relative velocity

$$v = \frac{dr}{dt} = \frac{dr}{d\theta}\Big/\frac{dt}{d\theta} = \left(\frac{2\,G\,M_{\rm total}}{R_{\rm max}}\right)^{1/2} \left(\frac{\sin\theta}{1 - \cos\theta}\right)\,. \tag{6.10}$$

The last three equations can be combined to eliminate $R_{\rm max}$, G and $M_{\rm total}$, to give

$$\frac{v\,t}{r} = \frac{\sin\theta\,(\theta - \sin\theta)}{(1 - \cos\theta)^2}\,. \tag{6.11}$$

v can be measured from Doppler shifts, and $r \approx 710$kpc from Cepheid variables. For t we can take the age of the Universe. Current estimates of t are quite accurate[6], but even using ages of the oldest MW stars, $t \sim 15$Gyr gives an interesting result. Using these numbers, we can solve the previous equation (numerically) to find $\theta = 4.32$ radians, *assuming M31 is on its first approach to the MW*[7].

[6]From properties of the micro-wave background radiation.

[7]Equation (6.11) has no unique solution for θ, since it describes motion in a periodic orbit. On its first approach, θ should be the *smallest* solution to the equation.

Substituting yields $M_{\rm total} \approx 3.66 \times 10^{12} M_\odot$, and hence for the MW mass, $M \approx M_{\rm total}/3$

$$M \approx 1.2 \times 10^{12} M_\odot \,, \tag{6.12}$$

comfortably higher than our lower limit Eq.(6.6).

Since the luminosity of the MW (in the V band) is $1.4 \times 10^{10} L_\odot$, the corresponding mass-to-light ratio for the MW is around $\Gamma = 100$. Furthermore, the estimate of M is increased if the orbit is not radial, or M31 and the MW have already had one (or more) pericenter passages since the Big Bang.

If all stars in the MW and M31 were solar mass stars, we would expect $\Gamma = 1$. Now even in the solar neighbourhood, most stars are less massive than the sun, and so the mass-to-light ratio for the *stars* is about 3 or so[8]. So the very large mass inferred from the LG dynamics strongly corroborates the evidence from rotation curves and Oort's constants, that most of the mass in the MW (and presumably also in M31) is dark.

From these numbers, we can also estimate the extent $R_\star$ of such a putative dark halo. If the the circular velocity $V_c = 220{\rm km\ s}^{-1}$ out to $R_\star$, then from $V_c^2 = GM/R_\star$ we find

$$R_\star = \frac{GM}{V_c^2} \approx \frac{G\,10^{12} M_\odot}{(220{\rm km\ s}^{-1})^2} \approx 100{\rm kpc}\,. \tag{6.13}$$

If, as is more likely, the rotation speed eventually drops below $220{\rm km\ s}^{-1}$, then R is even bigger. Hence the extent of the dark matter halo around the MW and M31 is truly enormous. Recall that the size of the stellar disk is of order 20kpc or so, and the distance to M31 $\sim$ 700kpc. So the dark matter haloes of the MW and M31 may almost overlap.

[8]Recall that the luminosity of stars scales with their mass quite steeply, $L \propto M^\alpha$, with $\alpha \approx 5$, and hence $M/L \propto M^{1-\alpha}$. See your notes on stars, p. 34.

6.3 Summary

After having studied this lecture, you should be able to

- Show that in a point mass model of the MW, the high velocity stars are not bound.
- Estimate the parameters of a dark halo, assuming the high velocity stars are bound to the MW.
- Describe the properties of the Local Group in terms of the galactic content.
- Estimate the mass and extent of the dark halo of the MW from the Local Group timing argument.

Chapter 7

Elliptical galaxies. I

Elliptical galaxies are spheroidal stellar systems with smooth luminosity profiles which contain old, metal rich stars. They have a large range in sizes. Usually, rotation is unimportant in ellipticals, unlike for spirals. They often contain low density, very hot gas that emits X-rays. The properties of this hot gas, and also the stellar dynamics, suggest that also in ellipticals most of the mass is in dark matter.

7.1 Luminosity profile

In contrast to spirals galaxies, elliptical have smooth surface brightness (SB) profiles, ellipsoidal in shape. The SB profile of some elliptical are shown in Figure (7.1). Note how in the center the isophotes are well defined and elliptical, whereas in the outer parts they become less smooth (noisier) because the SB decreases. Also note the presence of foreground MW stars.

A plot of the average surface brightness as function of radius for these same ellipticals is shown in Figure (7.2). The drawn line which fits the data well, is an $r^{1/4}$ fit, Eq. (3.2). Often, these profiles are plotted as SB vs $r^{1/4}$ in which case the profile is a straight line, as in Figure (7.3). For most ellipticals this profile shape fits their SB-distribution well. A more general fit which sometimes used, is $I(r) \propto \exp[-(r/r_e)^{1/n}]$, where $n = 4$ corresponds to the de Vaucouleurs $r^{1/4}$ fit.

Some of the giant elliptical galaxies that are invariably found in the centers of big clusters of galaxies have an extended halo that is not fit well by the de Vaucouleurs profile. Such galaxies are called cD galaxies. Fig. (7.4)

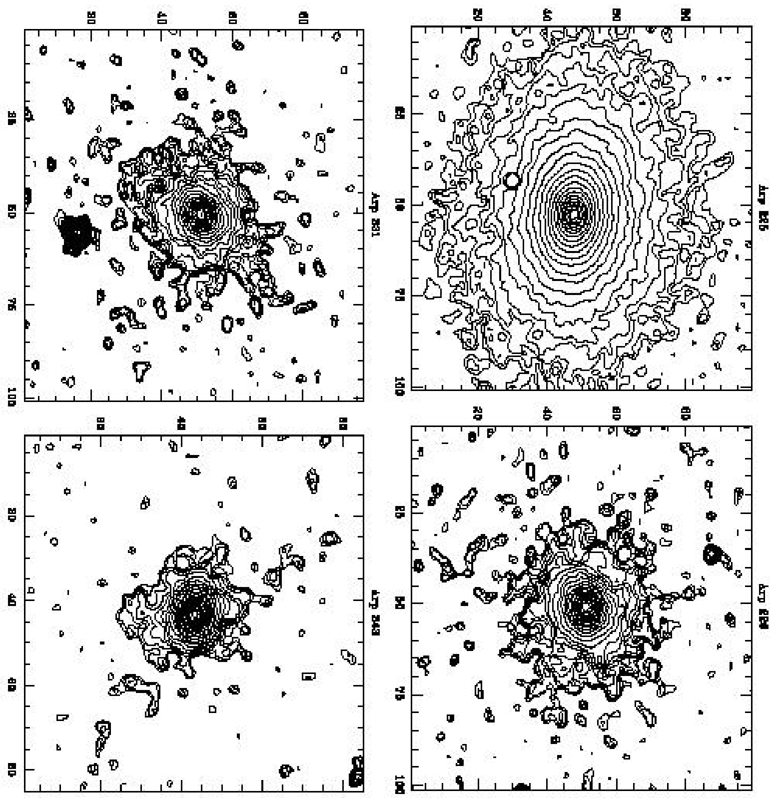

Figure 7.1: (Taken from astro-ph/0206097) Lines of constant surface brightness (isophotes) in the K band for four elliptical galaxies. In successive isophotes, the surface brightness increases by 0.25 magnitudes.

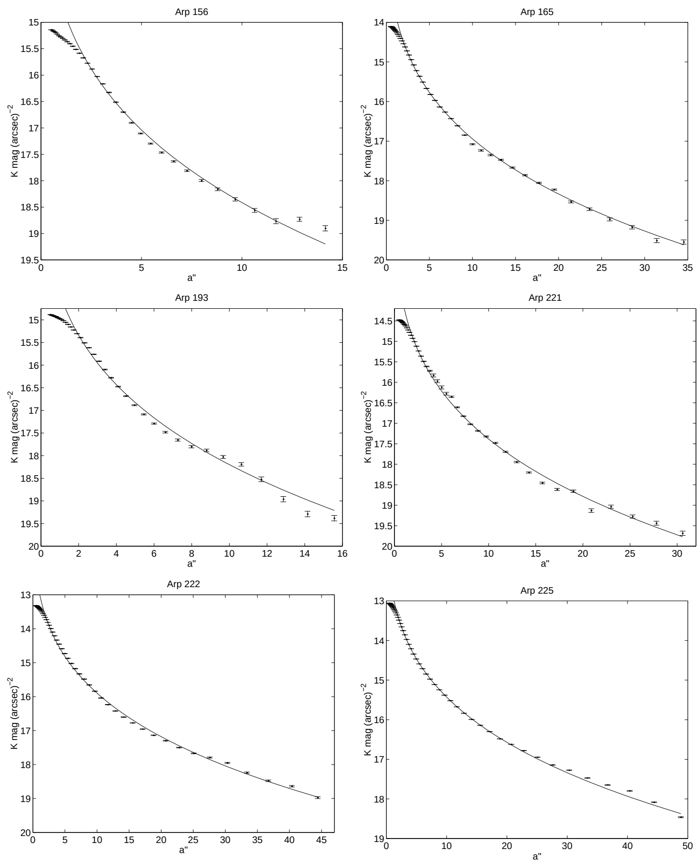

Figure 7.2: (Taken from astro-ph/0206097) Surface brightness as function of distance to the center. The drawn line is the best $r^{1/4}$ fit.

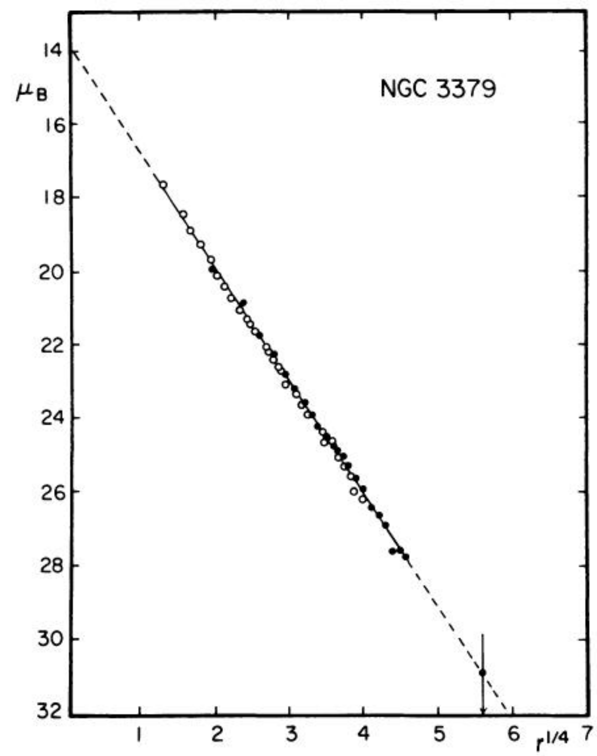

FIG. 2.—Mean E-W luminosity profile of NGC 3379 derived from McDonald photoelectric data. ●, Pe 4 data with 90 cm reflector; ○, Pe 1 data (M + P) with 2 m reflector. Note close agreement with $r^{1/4}$ law.

Figure 7.3: Luminosity profile for galaxy NGC 3379 (open symbols) and $r^{1/4}$ fit (drawn line). Taken from de Vaucoleurs & Capaccioli, ApJS 40, 1979.

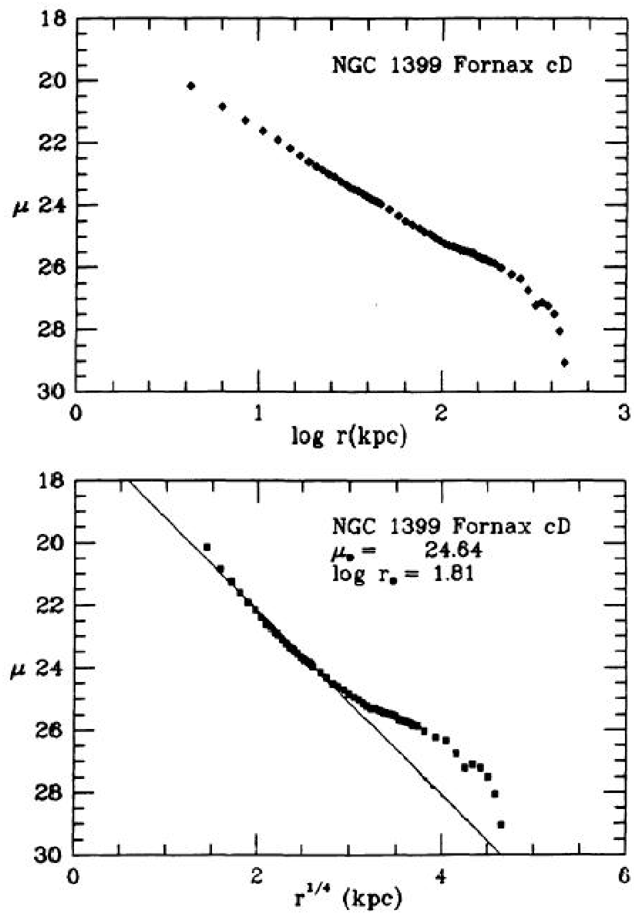

Figure 7.4: SB profile for NGC 1399 from Schombert (filled symbols; ApJS, 1989). The outer parts of the profile do not follow the $r^{1/4}$ fit (bottom panel), but are nearly a straight line in a SB-log(r) (i.e. a log − log) plot (top panel), meaning the profile is close to a power-law.

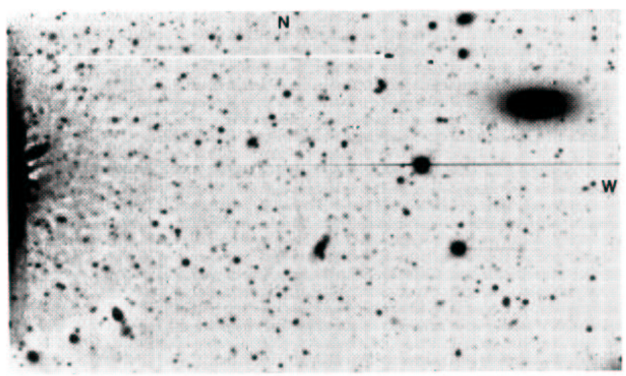

Figure 7.5: Image of NGC 1399 with smooth, best fit $r^{1/4}$ profile subtracted (Bridges et al., AJ 101, 469 (1991). The centre of NGC 1399 is at the left, the object toward the top right is another galaxy in the galaxy cluster. The other extended objects in the image are other galaxies. Clearly seen are hundreds of high SB unresolved objects, which are GCs in the halo of NGC 1399.

show the profile of NGC 1399, which is the cD galaxy in the nearby Fornax cluster of galaxies. This very extended halo of stars is truly enormous: it can be traced to around a Mpc. Even though the distance to Fornax is large, it means that the extent on the sky of NGC 1399 is about the size of the moon! This gigantic size has prompted the suggestion that may be we should associate such a halo with the cluster of galaxies, rather than with any individual galaxy. This suggestion has some support from the observations of some clusters where we observe a faint halo, but *without* a galaxy in the center!

Es also have many globular clusters. Figure (7.5) is an image of NGC 1399, where an $r^{1/4}$ fit to the SB-profile of the galaxy has been subtracted. Clearly visible are 100s of high SB objects, indistinguishable on this plate from foreground stars, which are in fact GCs in the halo of the galaxy. The GC luminosity function (i.e., the number of clusters as function of their magnitude) has been used as a standard candle: by measuring the luminosity function of GCs of a galaxy, and comparing it to the luminosity function of a galaxy with known distance, one can obtain a distance estimate.

Why do ellipticals have an $r^{1/4}$ SB-profile? As I explained earlier, there seems to be a strong connection between the density of galaxies (in the sense of how many galaxies you find per Mpc^3 in a given region, say) and their types: high density regions invariably contain a much larger fraction of Es than low density regions, where Ss dominate. So, may be interactions between galaxies have something to do with it?

Recent numerical simulations of collisions between two spiral galaxies do produce objects with SB-profiles not too different from $r^{1/4}$ appropriate for an E. And although this is a good hint that we're on the right track, in fact, also in the simulations we don't really know *why* this is. In a famous paper, Donald Lynden-Bell suggested that during a galaxy merger, the gravitational potential Φ which the stars feel, changes very rapidly and this might produce a characteristic density profile[1]. He termed this process **violent relaxation**. And so the suggestion is that, as (predominantly spiral) galaxies fall into galaxy clusters, they will violently interact with other galaxies, transforming them somehow into Es. And although this is certainly an important piece of

[1]rather, a characteristic *distribution* function, the density of stars in six-dimensional position and velocity space.

the puzzle, it can't be the whole story, because the stellar populations of Es and Ss differ.

7.2 Stellar populations and ISM

Stars in Es tend to be older and more metal rich than those in Ss, and as a consequence Es are redder than Ss. You can easily understand why this is so. We discussed this before: if the atmosphere of a star contains more metals, then the dust and the gas will scatter preferentially the *blue light* hence the stars appears redder. That's one. Secondly, if you have young stars in a stellar population, then the more massive stars may still be alive[2], and these are hot and hence blue. So when a given stellar population gets older, it will become redder as well. And in an E galaxy, both effects occur.

The fact that a stellar population appears blue either because it is young, or because it is metal poor, is called the *age-metallicity degeneracy*: just observing the colour of a population, you cannot tell a young but metal rich population apart from an old but metal poor one. To make sure, you have to have independent constraints on th metallicity Z, for example from stellar spectra, or, the astronomer's favourite, from HII regions.

Dust and gas Although I said that a characteristic difference between Es and Ss is that Es don't contain dust or gas, and have no star formation, this is of course not complete true, they usually have some, but much less than Ss. Many Es contain strong dust lanes, the Sombrero galaxy is a beautiful example (Fig. 7.6). Often, there is almost no connection between the orientation of the dust lane, and any other of the galaxy's parameters. It is thought that such dust lanes are evidence that the E recently swallowed a small galaxy. In fact, one sometimes observes very faint rings of stars around an E, and also this could be the result of 'galactic cannibalism'. A good example is shown in Figure (7.7).

[2]Recall that massive stars have short lifetimes.

Figure 7.6: Image of the ‘Sombrero’ galaxy, with its striking dust lane.

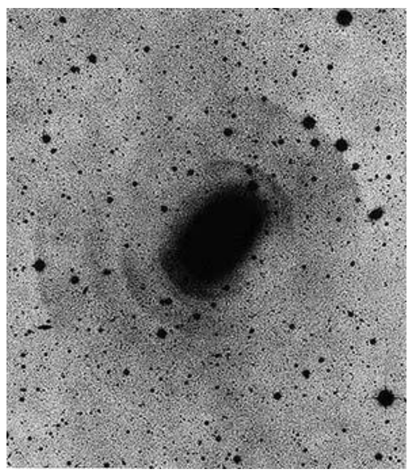

Figure 7.7: Image of NGC 3923 taken by David Malin. Several faint shells of stars appear in the outer parts of the galaxy. (see http://www.ast.cam.ac.uk/AAO/images/general/ngc3923.html)

7.3 X-rays

Elliptical galaxies emit X-rays. I want to briefly discuss *why*, *how* X-rays are observed, and *what* we can learn from them.

Fortunately for us, the earth's atmosphere blocks X-rays. Unfortunately for astronomers, this means we need to be above the atmosphere, in a balloon, rocket or satellite, to observe extra-solar X-rays.

How does one built an X-ray telescope? X-rays do not reflect off mirrors the same way that visible light does. Because of their high energy, X-ray photons penetrate into the mirror in much the same way that bullets slam into a wall. Likewise, just as bullets ricochet when they hit a wall at a grazing angle, so too will X-rays ricochet off mirrors. These properties mean that X-ray telescopes must be very different from optical telescopes. Basically, they contain many almost parallel plates, that gently nudge the incoming X-rays into the new direction you want them to go in. See http://chandra.harvard.edu/xray_astro/ for a nice diagram and more detailed info.

The sun (we're really close!) was the first source to be detected in the X-rays, in 1949. The technology improved in great leaps, and it came somewhat as a surprise to detect X-rays from Es that were not associated with *point sources.* Why am I talking about point sources? Recall from the stellar part of the course the concept of *X-ray binary.* In these binary stars, hot gas associated with mass transfer radiates in the X-ray band. Such (binary) stars are also seen in Es. But subtracting these sources from the X-ray image, we still detect *extended X-ray emission* from the galaxy. The process that produces them is called *thermal bremsstrahlung.*

7.3.1 Thermal bremsstrahlung

How are the X-rays produced in this hot gas? Since the gas temperature is very high, most elements are highly ionised, and so the gas is really a hot plasma. The negatively charged electrons in this plasma experience a force when they pass close to an ion, and an accelerating charge emits electromagnetic dipole radiation: this is the radiation that we observe. Since the electron looses energy in the encounter, it will slow down. Hence the name thermal bremsstrahlung[3]. And as the electron slows down, the plasma also looses energy, and so the gas cools.[4] The electron is not bound to any particular ion throughout the interaction and this process is therefore also called free-free radiation. **This is all you need to know about this**, but if you want more details, here they come.

radiated power

Let's try to obtain an expression for the radiated power. The aim of this derivation is to show that the power emitted is $\propto n_e\, n_i\, T^{1/2}$, where n_e and n_i are the electron and ion densities, and T is the temperature. You could have guessed the first bit, since we're talking about an encounter between *two* particles. So the exercise is how to get the $T^{1/2} \propto v$ dependence.

If the deflection angle is small, we can simple neglect it, and say that the electron moves at constant speed v along a straight line with impact parameter b. The acceleration is then

$$\mathbf{a}(t) = \frac{Ze^2}{m_e d^3}\mathbf{d} = \frac{Ze^2}{m_e(b^2+v^2t^2)}\mathbf{d}/d\,, \tag{7.1}$$

where Z is the charge of the ion in units of the electron charge e, and $\mathbf{d}$ is the distance between electron and ion. $t=0$ is the time of closest approach, and so the encounter lasts from $t\to-\infty$ to $t\to+\infty$.

The dipole electric field at distance r from the electron is

$$|\mathbf{E}(t)| = \frac{e\,a}{c^2\,r}\sin(\theta) = \frac{Z\,e^3\sin(\theta)}{m_e\,c^2\,r\,(b^2+v^2t^2)}\,. \tag{7.2}$$

This is the electric field as function of time t, but we want it as a function of frequency ν. The trick to obtain $E(\nu)$ is by Fourier transforming $E(t)$,

[3] German for 'braking radiation'.

[4] Also the ions collide of course, but since they are much more massive than the electron, their acceleration is far smaller, and hence their radiation is negligible.

$$
\begin{aligned}
E(\nu) &\equiv \int_{-\infty}^{\infty} E(t)\exp(2\pi\, i\nu\, t)dt \\
&= \frac{Ze^3\sin(\theta)}{m_e c^2\, r}\int_{-\infty}^{\infty}\frac{\exp(2\pi\, i\nu\, t)}{b^2+v^2t^2}dt \\
&= \frac{Ze^3\sin(\theta)}{m_e c^2\, r}\frac{1}{b\, v}\int_{-\infty}^{\infty}\frac{\exp(2\pi\, ix(\nu\, b/v))}{(1+x^2)}\, dx \\
&= \frac{Ze^3\sin(\theta)}{m_e c^3 b^2 v^2}\,\frac{\pi}{bv}\,\exp(-2\pi\nu\, b/v)\,. \qquad (7.3)
\end{aligned}
$$

Along the way, I've introduced the variable $x = vt/b$. The integral $\int \exp(2\pi x\alpha)/(1+x^2)\, dx =$ some function of α is a little tricky to perform, and I don't expect you to be able to do it. What is important is to note that this change of variables gives rise to a factor $1/bv$.

Now we're done, because it means that the power radiated by this single electron per unit frequency is

$$
\frac{dW}{d\nu} \propto \int E^2(\nu)\, d\Omega \propto \frac{1}{(b\, v)^2}\exp(-4\pi\nu\, b/v)\,. \qquad (7.4)
$$

So the spectrum is independent of ν for low frequencies $\nu \ll v/b$, and cuts off exponentially for large ν. Now the number of encounters with impact parameter between b and $b+db$, in time dt is $n_e 2\pi b\, dbvdt$, and so this gives the power as

$$
\frac{dW}{d\nu\, dt\, dV} \propto n_e\, n_i\, v^{-1}\exp(-4\pi\nu\, b/v) \propto n_e\, n_i\, T^{-1/2}\exp(-4\pi\nu\, b/v)\,. \qquad (7.5)
$$

So the origin of the $1/v$ is that the energy radiated per encounter $\propto 1/v^2$ whereas as the rate of encounters is $\propto v$.

To obtain the final result, we still need to integrate over the impact parameter b, and average over all impact velocities v (for example by assuming a Maxwell-Boltzmann distribution for v). The result is that the total power emitted is

$$
\text{power} \propto n_e\, n_i\, T^{1/2}\,. \qquad (7.6)
$$

Notice that the total power is much more dependent on the density of the gas ($\propto$ density *squared*), than on its temperature ($\propto T^{1/2}$)

7.3.2 Spectrum

On top of the thermal bremsstrahlung emission spectrum, observed X-ray spectra also exhibit spectral lines. Given the high temperature of the gas, it won't come as a surprise that we observe lines of highly ionised species – Fe-lines are especially important in the wave-bands observed by modern X-ray telescopes. And just as we did for HII regions, we can try to constrain the physical conditions in the hot gas, by modelling the relative strength of different transitions. This is one way to determine the temperature T of the hot gas.

7.4 Evidence for dark matter from X-rays

An important application of X-ray observations is to probe the gravitational potential of the galaxy, and hence infer its density distribution. Consider a thin shell of hot gas. The gravitational force due to mass enclosed by the shell (consisting of stars, gas and dark matter) will pull the shell inward. At the same time, if the pressure decreases outward, then the pressure gradient will try to push the shell outward. In *hydrostatic equilibrium*, these forces balance. By estimating the pressure gradient from X-rays, we can infer the gravitational force and hence look for evidence for dark matter. Let's assume for simplicity that the elliptical galaxy is spherically symmetric.

Let $M(< r)$ be the mass inside the shell of radius r (consisting of stars, gas and dark matter: $M = M_\star + M_{\rm gas} + M_{\rm dm}$), and dr its thickness The gravitational force on the shell is

$$F_g(r) = \frac{GM(< r)M_s}{r^2}\,, \tag{7.7}$$

where $M_s = 4\pi r^2 \rho_{\rm gas}(r) dr$ is the mass of the shell.

Let $p(r)$ be the pressure in the hot gas. The pressure force on a surface is $p(r)$ times the surface area, $4\pi r^2$. The net outward force is the difference between the outward and inward forces:

$$F_p(r) = 4\pi r^2 (p(r) - p(r + dr)) = -4\pi r^2 \frac{dp}{dr}\, dr \tag{7.8}$$

where the pressure was expanded in a Taylor series. Combining the last three equations gives

$$\frac{GM(<r)}{r^2}4\pi r^2\rho_{\rm gas}(r)dr = -4\pi r^2\frac{dp}{dr}dr\,, \tag{7.9}$$

hence

$$\frac{GM(<r)}{r^2} = -\frac{1}{\rho_{\rm gas}(r)}\frac{dp}{dr}\,, \tag{7.10}$$

or alternatively in terms of the gravitational potential Φ

$$\nabla\Phi = -\nabla p/\rho_{\rm gas}\,. \tag{7.11}$$

This should look familiar from hydrostatic equilibrium in stars!

By modelling the spectrum as function of position we can determine $T(r)$, and since the X-ray intensity $\propto \rho^2_{\rm gas}T^{1/2}$, we can also determine $\rho_{\rm gas}(r)$. Hence $p \propto rho_{\rm gas}T$, and so we have determined the rhs of Eq.(7.11). From that, one can reconstruct the gravitational potential, and hence infer the mass distribution.

The temperature T is observed to remain approximately constant, in which case we can obtain an approximate expression for the density profile. The pressure is

$$p = \frac{{\rm k_B}\,T}{\mu\,m_{\rm p}}\,\rho_{\rm gas}\,. \tag{7.12}$$

If we take T *constant* then the rhs of Eq. (7.10) becomes $-\frac{{\rm k_B}\,T}{\mu\,m_{\rm p}}d\ln(\rho_{\rm gas})/dr$, and hence

$$GM(<r) = -\frac{{\rm k_B}T}{\mu\,m_{\rm p}}r^2\frac{d\ln\rho_{\rm gas}}{dr}\,. \tag{7.13}$$

Taking the derivative of both sides with respect to r, gives

$$G4\pi r^2\rho_{\rm tot} = -\frac{{\rm k_B}T}{\mu\,m_{\rm p}}\frac{d}{dr}(r^2\frac{d\ln\rho_{\rm gas}}{dr})\,, \tag{7.14}$$

since $dM(<r)/dr = 4\pi\,r^2\,\rho_{\rm tot}$. $\rho_{\rm tot}$ is the *total mass density*, *i.e.* due to stars, gas and dark matter.

Now assume that the gas density and total mass density are proportional, $\rho_{\rm gas} \propto \rho_{\rm total}$, and try a solution of the form $\rho_{\rm totcl} = \rho_0\,(r_0/r)^\beta$. Substitution shows that this is indeed a solution when $\beta = 2$, in which case

$$\rho_{\rm tot}(r) = \frac{{\rm k_B}\,T}{2\pi\,G\mu m_{\rm p}}\,r^{-2}\,. \tag{7.15}$$

So a measurement of T can constrain the **total** mass density $\rho_{\rm tot}$. Estimating the stellar mass density from the luminosity, the gas density from the X-ray emissivity, we find $\rho_{\rm dm} = \rho_{\rm total} - \rho_\star - \rho_{\rm gas}$. More detailed modelling of this type confirms that also elliptical galaxies have most of their mass in the form of dark matter.

7.5 Summary

After having studied this lecture, you should be able to

- describe the surface brightness profiles of Es in terms of the de Vaucouleurs profile.
- explain why we think that this profile results from galaxy encounters.
- recall that Es have typically hundreds of GCs.
- explain why dust lanes in Es and shells of stars around Es are thought to be evidence for galactic cannibalism.
- contrast the stellar population and ISM of Es with those of Ss.
- describe how X-rays are detected, and explain the process with which X-rays are produced in the hot gas in Es. (But not be able to recall the equation, or reproduce the calculation.)
- explain how X-ray observations can be used to infer the gravitational potential in Es.

Chapter 8

Elliptical galaxies. II

This chapter is a bit harder, with more mathematics than I would like to use. I **do not** expect you to be able to reproduce all these derivations, only the ones stressed in the Summary. The aim of this chapter is twofold: (*i*) explain why you can derive fluid-like equations to describe the structure of ellipticals, and (*ii*) use these fluid equations to explain why elliptical are elliptical in shape.

We've seen that the stars in spiral disks are all on nearly circular orbits. Although some of the smaller Es sometimes rotate as well, the angular momentum of stars is not sufficient to balance gravity. So we need to look for another mechanism to prevent collapse. Basically, it's the *velocity dispersion* of the stars that balances gravity, much like it is a pressure gradient in the gas that balances gravity in a star. For this reason, Es (but also spiral bulges, and globular clusters for example) are called *hot stellar systems*.

The Jeans's equations relate the stellar velocity dispersion, $\sigma_{ij}^2 = \langle (v_i - \langle v_i \rangle)(v_j - \langle v_j \rangle) \rangle$ with the gravitational potential of the system, and are the stellar analogue of the equation of hydrostatic equilibrium $\nabla\Phi = -\nabla p/\rho$ in a star. They do not describe the properties of the orbits of a single star, but rather assume one can use averaged properties. To investigate whether such an approach makes sense, I first need to introduce the concept of relaxation time.

8.1 Relaxation in stellar systems

As a star moves through a stellar system, it will feel the gravitational force due to all other stars. Is the motion of this star mainly determined by the average gravitational force of all other stars combined, or is it mostly sensitive to the force due to near stars?

To see that this would make a huge difference, consider the case where a single star (a test particle) moves through a smooth density distribution with spherically symmetric density distribution $\rho(r)$, where $r = |\mathbf{r}|$. In such a time-independent potential both the angular momentum, $\mathbf{L} = m\,\mathbf{r} \times \mathbf{v}$, and energy, $E = v^2/2 + \Phi$, of the star will be conserved along its orbit. Now suppose such a density distribution to be represented by a finite number of stars. As our test star moves along its orbit, it will be deflected by encounters with these other stars. In such a situation, neither $\mathbf{L}$ nor E will be conserved, and so the properties of its orbit will change in time. Clearly, the longer the star orbits in this density distribution, the more E will differ from its initial value. *The relaxation time T_E is a measure of how long the star remembers its initial energy.* We will see that the more stars are there in the galaxy, the longer is the relaxation time. You might have expected this, since for an infinite number of stars, the density is smooth, $\mathbf{L}$ and E are constant, and hence $T_E \to \infty$.

To estimate T_E, we will compute the change in energy ΔE of a star due to a single encounter with another star, as function of the impact parameter b and impact velocity v (the velocity at infinity, before the encounter started), and then sum over encounters. This derivation was first done by Chandrasekhar, the derivation here is slightly different from BT (p. 188).

Let's compute the change in velocity $\delta\mathbf{v}$ of this star. Because we want to treat the case where the effect of encounters is small, we'll approximate the orbit as a straight line, traversed with constant velocity v. The force perpendicular to the orbit is

$$\mathbf{F}_\perp = \frac{Gm^2}{b^2 + (vt)^2}\cos(\theta) \approx \frac{Gm^2}{b^2}\left[1 + (\frac{vt}{b})^2\right]^{-3/2}, \tag{8.1}$$

where I've assumed equal mass stars. The approximation $\cos(\theta) = b/r \approx b/vt$ is valid for b small. Combined with Newton's law

$$m\dot{\mathbf{v}}_\perp = \mathbf{F}_\perp\,, \tag{8.2}$$

yields the net change $|\delta\mathbf{v}_\perp|$ over a single encounter:

$$|\delta\mathbf{v}_\perp| \approx \frac{Gm}{bv}\int_{-\infty}^{\infty}(1+s^2)^{-3/2}\,ds \;=\; \frac{2Gm}{b\,v}\,. \tag{8.3}$$

Thus the change is roughly equal to the force at closest approach, Gm/b^2, times the time this force acts, $\Delta t \sim b/v$. Because of symmetry, encounters are equally likely to change the velocity in the positive direction, as in the negative direction. Therefore the mean change $\langle\delta\mathbf{v}_\perp\rangle = 0$. We will therefore consider the rate of change of $\delta\mathbf{v}_\perp^2$.

To compute the rate of change of $\delta\mathbf{v}_\perp^2$, due to many encounters, proceed as follows. Suppose the test star moves with velocity v through a stellar system with radius R, and (uniform) number density of stars, n. When travelling for a time δt, the number of encounters with impact parameter between b and $b+db$, is simply the volume of the cylindrical shell, $dV = 2\pi\,b\,db\,\times\,v\delta t$, times the number density of stars, n. Here, $v\delta t$ is the length of the cylinder, $2\pi\,b\,db$ is the surface area of the shell. Since each of these encounters leads to a change in velocity squared by an amount $\delta\mathbf{v}_\perp^2$, we find for the rate of change

$$\frac{d^2\delta\mathbf{v}_\perp^2}{dt\,db} = n\,v\,2\pi\,b\,\times\,(\frac{2Gm}{b\,v})^2\,. \tag{8.4}$$

This is the product of the *rate* of encounters, $n\,v\,2\pi\,bdb$, with the velocity change per encounter. Now we can integrate over impact parameter to obtain for the rate of change in velocity

$$\frac{d\delta\mathbf{v}_\perp^2}{dt} = \int_0^\infty n\,v\,2\pi\,b\,(\frac{2Gm}{b\,v})^2\,db = 2\pi\,nv\,(\frac{2Gm}{v})^2\,\int_0^\infty\frac{db}{b}\,. \tag{8.5}$$

Unfortunately this integral diverges, both at small impact parameters (which are few in number, but cause large velocity changes), and at large impact parameters (which cause small velocity changes yet are very numerous). How to overcome this divergence?

For a real stellar system, there cannot be encounters with impact parameter larger than the size R of the system. For the smallest impact parameter b, we can take $b_{\rm min} = Gm/v^2$, for which the *change* in velocity is comparable to the *initial* velocity (cfr.Eq. 8.3). Recall that our derivation assumed small changes anyway. With this caveat in mind we replace the previous equation by

$$\frac{d\delta \mathbf{v}_\perp^2}{dt} \approx 2\pi\, nv\, (\frac{2Gm}{v})^2 \int_{b_{\rm min}}^R \frac{db}{b} = 2\pi\, nv\, (\frac{2Gm}{v})^2\, \ln(\frac{R}{b_{\rm min}})\,. \tag{8.6}$$

We can simplify this by assuming that the stellar system is in virial equilibrium, in which case the kinetic energy of all stars combined, $K \sim N\, m\, v^2/2$ is half of the potential energy, $U \sim G\,(N\, m)^2/R$, where m is the mass of a star. Therefore

$$v^2 \approx \frac{GNm}{R}\,. \tag{8.7}$$

Combining this with the expression for $b_{\rm min}$ it is easy to show that $R/b_{\rm min} = N$, the total number of stars.

So far we have computed the *rate* of change of the velocity. To judge whether the change of velocity is large or small, we can multiply the rate with the *crossing time* of the system, $t_{\rm cr} \equiv R/v$. It is a measure of how long it takes to cross the stellar system, on average. The velocity change during a single crossing of the stellar system is therefore

$$\begin{aligned} \delta \mathbf{v}_\perp^2 \sim \frac{d\delta \mathbf{v}_\perp^2}{dt}\, t_{\rm cr} &= 2\pi\, nv\, (\frac{2Gm}{v})^2\, \ln(N) \times \frac{R}{v} \\ &= \frac{6\ln(N)}{N}\, v^2\,. \end{aligned} \tag{8.8}$$

The first line combines Eq.(8.6) with our finding $R/b_{\rm min} = N$ and the definition of the crossing time. The second step assumes virial equilibrium, Eq. (8.7). We are finally in a position to define the *relaxation time* $T_{\rm relax}$, as the time it takes for encounters to change the velocity by order of itself, hence

$$\frac{v^2}{T_{\rm relax}} \equiv \frac{d\delta \mathbf{v}_\perp^2}{dt} = \frac{6\ln(N)}{N}\, \frac{v^2}{t_{\rm cr}} \tag{8.9}$$

hence we get as our final result

$$T_{\rm relax} = \frac{N}{6\ln(N)}\, t_{\rm cr}\,. \tag{8.10}$$

For systems of the same mass and size, and hence a *given* crossing time, we find that the relaxation time increases with N as $T_{\rm relax} \propto N/\ln(N)$. Put differently, the effect of encounters decreases with increasing particle numbers as $N/\ln(N)$.

For a Globular Cluster with $N = 10^5$ say, $r = 10$pc, and $v = 10$km s^{-1}, $T_{\rm cross} \approx 1$Myr. Putting in the numbers[1], we find $T_{\rm relax} \approx 10^9$ years, much smaller than the ages of the stars in the GC. Hence we expect stellar encounters to be important in shaping the structure of GCs.

However, for a big elliptical galaxy, $N = 10^{11}$, say. The crossing time is of order 20kpc/200km s^{-1} $\approx 10^2 M$yr, and $T_{\rm relax} \sim 10^{13} M$yr which is *much* larger than the age of the Universe ($\sim 10^{10}$yr). Therefore we can completely neglect encounters between stars, when describing the equation of stellar dynamics in a galaxy, and derive fluid-like equations for the behaviour of a galaxy. The stars will be like the gas particles in the usual gas equations. This is what we'll do next.

8.2 Jeans equations

8.2.1 A continuity equation

We can derive fluid-like equation for the gas of stars in an elliptical galaxy, using the concept of a *continuity equation*. A continuity equation expresses conservation of a quantity. This is a general concept, let's do a simple example first: conservation of number of cars on a motor way.

Suppose you're standing on a hill above a tunnel on a motor way, and you count the number of cars entering the tunnel on one side and the number that leaves the tunnel on the other side.

If n is the number density of cars in the tunnel (in cars per km say), and Δl is the length of the tunnel, then the number of cars in the tunnel is obviously $N = n\,\Delta l$. The change in N between two times is

$$\Delta N = [n(l, t+\Delta t) - n(l,t)]\,\Delta l \approx \frac{\partial n(l,t)}{\partial t}\,\Delta l\,\Delta t\,. \tag{8.11}$$

[1]A handy approximate relation is 1km/$s \approx$ 1pc/Myr.

Assuming there is no other exit, this is also the change between the number of cars entering and leaving the tunnel,

$$\Delta N = [n(l,t)\dot{l}(l,t) - n(l+\Delta l,t)\dot{l}(l+\Delta l,t)]\,\Delta t \approx -\frac{\partial n(l,t)\,\dot{l}(l,t)}{\partial l}\,\Delta l\,\Delta t\,. \tag{8.12}$$

The first term $n(l,t)\dot{l}(l,t)\,\Delta t$ is the number of cars entering the tunnel in time Δt, the second term is how many are leaving in that time. $\dot{l}(l,t)$ obviously denotes the speed of the cars.

Combining these two equation, we get the following continuity equation:

$$\frac{\partial n(l,t)}{\partial t} + \frac{\partial n(l,t)\,\dot{l}(l,t)}{\partial l} = 0\,. \tag{8.13}$$

A little bit of thought shows that, in a N-dimensional case (spaceships through a worm-hole?), the more general equation would be

$$\frac{\partial n(\mathbf{l},t)}{\partial t} + \frac{\partial n(\mathbf{l},t)\,\dot{\mathbf{l}}(\mathbf{l},t)}{\partial \mathbf{l}} = 0\,, \tag{8.14}$$

so there is a partial derivative $\partial/\partial l_i$ for every Cartesian coordinate l_i, and n is multiplied by $\dot{l}_i$ in the second term, which is the velocity of the spaceships in the i-direction.

8.2.2 Boltzmann's equation

We're interested in finding an equation that describes the distribution of stars in a galaxy. So we want to know is how many stars there are in a small volume $d\mathbf{x}$ around a position $\mathbf{x}$, that have velocity in a small interval $d\dot{\mathbf{x}}$ around a given velocity $\dot{\mathbf{x}}$. This density of stars, $f(\mathbf{x},\dot{\mathbf{x}},t)$, is called the *distribution function*. If we don't care about the velocity of stars, then we can integrate f over all velocities, to obtain the density of stars,

$$n(\mathbf{x},t) = \int f(\mathbf{x},\dot{\mathbf{x}},t)\,d\dot{\mathbf{x}}\,. \tag{8.15}$$

Now we only have to realise that stars acts as our cars on the motor way, and hence f should satisfy the continuity equation (8.14). The way to do it is to substitute for the l coordinates in that equation the positions and velocities of the stars,

$$\begin{aligned} \mathbf{l} &\equiv (\mathbf{x}, \mathbf{v}) \\ \dot{\mathbf{l}} &\equiv (\mathbf{v}, \dot{\mathbf{v}}) = (\mathbf{v}, -\nabla\Phi) \,, \end{aligned} \tag{8.16}$$

where I've written $-\nabla\Phi$ for the gravitational acceleration $\dot{\mathbf{v}}$ of the star.

So this $\mathbf{l}$ vector has 6 coordinates, $l_\alpha = \mathbf{x}$ for $\alpha = 1 \to 3$, and $l_\alpha = \mathbf{v}$ for $\alpha = 4 \to 6$. Now these coordinates are very special, since

$$\begin{aligned} \sum_{\alpha=1}^{6} \frac{\partial \dot{l}_\alpha}{\partial l_\alpha} &= \sum_{i=1}^{3} \left(\frac{\partial v_i}{\partial x_i} + \frac{\partial \dot{v}_i}{\partial v_i} \right) \\ &= \sum_{i=1}^{3} -\frac{\partial}{\partial v_i} \left(\frac{\partial \Phi}{\partial x_i} \right) \\ &= 0 \,. \end{aligned} \tag{8.17}$$

Here, $\partial v_i / \partial x_i = 0$ since the positions and velocities of the stars are independent variables, and the second term is zero because the gravitational acceleration $\partial\Phi/\partial x_i$ does not depend on velocity.

We can use this equation to simplify the continuity equation, and to obtain

$$\frac{\partial f}{\partial t} + \sum_{\alpha=1}^{6} \dot{l}_\alpha \frac{\partial f}{\partial l_\alpha} = 0 \,. \tag{8.18}$$

Rewriting in terms of positions and velocities, we obtain the **collisionless Boltzmann equation**,

$$\frac{\partial f}{\partial t} + \sum_{i=1}^{3} \left(v_i \frac{\partial f}{\partial x_i} - \frac{\partial \Phi}{\partial x_i} \frac{\partial f}{\partial v_i} \right) = 0 \,. \tag{8.19}$$

or in vector notation

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{r}} - \nabla\Phi \cdot \frac{\partial f}{\partial \mathbf{v}} = 0 \,. \tag{8.20}$$

8.2.3 Moments of the Boltzmann equation

Although the collisionless Boltzmann equation (CBE for short) provides us with a complete description of the stellar dynamics problem, it is rather unwieldy for practical applications. The Jeans equations, which give a clearer physical picture of what is going on, are *moments* of the Boltzmann equation. Here is how to obtain them.

Suppose we integrate Eq. (8.19) over all velocities. The first term will become

$$\int \frac{\partial f(\mathbf{r},\mathbf{v},t)}{\partial t}\, d\mathbf{v} = \frac{\partial}{\partial t}\int f(\mathbf{r},\mathbf{v},t)\, d\mathbf{v} \equiv \frac{\partial n(\mathbf{r},t)}{\partial t}\,. \tag{8.21}$$

The integral of the distribution function over velocities, $n \equiv \int f\, d\mathbf{v}$ is just the density of stars. And we're allowed to interchange the integral over velocities, with the derivative wrt time, since the range of velocities over which we integrate does not depend on t.

Integrating the second term, gives

$$\int v_i \frac{\partial f}{\partial x_i}\, d\mathbf{v} = \frac{\partial}{\partial x_i}\int v_i\, f\, d\mathbf{v} \equiv \frac{\partial (n\bar{v}_i)}{\partial x_i}\,, \tag{8.22}$$

where I've introduced the mean velocity $\bar{v}_i$ as

$$n\,\bar{v}_i = \int v_i\, f\, d\mathbf{v}\,. \tag{8.23}$$

Now the last term is zero, since

$$\int_{-\infty}^{\infty} \frac{\partial f}{\partial v_i} dv_i = f|_{v_i=-\infty}^{v_i=+\infty}\,, \tag{8.24}$$

since there are no stars with infinite velocity.

And so the first moment of the CBE is a continuity equation for the density,

$$\frac{\partial n}{\partial t} + \frac{\partial (n\bar{v}_i)}{\partial x_i} = 0\,. \tag{8.25}$$

Now the second moment is what we're after. The trick is to *first* multiply the CBE with v_j, and *then* integrate over all velocities. The calculation proceeds exactly as for the density, and the result is after a bit of algebra:

$$n\,\frac{\partial \bar{v}_j}{\partial t} + n\,\bar{v}_i\,\frac{\partial \bar{v}_j}{\partial x_i} = -n\,\frac{\partial \Phi}{\partial x_j} - \frac{\partial n\,\sigma_{ij}^2}{\partial x_i}\,. \tag{8.26}$$

where

$$\sigma_{ij}^2 \equiv \overline{v_i v_j} - \bar{v}_i \bar{v}_j\,. \tag{8.27}$$

This was a lot of mathematics to come to a simple equation: suppose the system is in a steady state, $\partial/\partial t$=0, and there are no streaming motions, so that $\bar{v}_i = 0$. Then the previous equation simplifies to

$$\frac{\partial \Phi}{\partial x_j} = -\frac{1}{n}\frac{\partial n\sigma_{ij}^2}{\partial x_i}\,. \tag{8.28}$$

This is a more general version for the equation of hydrostatic equilibrium. To see this, note that the matrix σ_{ij} is symmetric: $\sigma_{ij} = \sigma_{ji}$. For such symmetric tensors, we can always rotate the coordinate system such that the tensor becomes diagonal, i.e. $\sigma_{ij}^2 = 0$ for $i \neq j$. Now, suppose that the velocities are *isotropic*, so that $\sigma_{xx} = \sigma_{yy} = \sigma_{zz} \equiv \sigma^2$. Then we find that

$$\frac{\partial \Phi}{\partial x_j} = -\frac{1}{n}\frac{\partial n\sigma^2}{\partial x_j}\,. \tag{8.29}$$

which is *exactly* the same as the equation for hydrostatic equilibrium, if $n\sigma^2$ is identified with the pressure p. Because of this analogy, the tensor $n\,\sigma_{ij}^2$ is called the **pressure tensor**.

In conclusion: the behaviour of collisionless stellar systems is similar to that of self-gravitating gas spheres. The role of temperature is taken over by that of the stellar velocity dispersion. It is the high velocity dispersion of the stars in an elliptical (and also in the bulge of spirals), which balances the gravitational pull. Since the required velocities are so high, such systems are called 'hot stellar systems'. Recall that in disk galaxies, it is the ordered motion of the stars – i.e. the rotation of the disk – which supports the system against gravity. Such systems are dynamically cold, i.e. the stellar velocity dispersion is small compared to the streaming motions. For disks, the 'velocity dispersion' refers to the small velocities that stars have with respect to their 'local standard of rest' (10s of km s^{-1}), versus the rotational velocity of the disk, 200km s^{-1}.

In a gas the pressure is isotropic and hence stars are spherical. In contrast in stellar systems, the pressure tensor $n\,\sigma^2_{ij}$ is in general *not* isotropic, and so the pressure gradient can be larger in one direction (x, say) than in another direction (y). So even if the potential is spherical, the stellar distribution may be more extended in x than in y, and the system will be elliptical in shape.

8.3 Summary

After having studied this lecture, you should be able to

- to derive an estimate for the stellar dynamical relaxation time in a spherical system of N stars.
- explain why we can describe the dynamics of elliptical galaxies in terms of fluid equations.
- explain the difference in dynamics between an elliptical (hot stellar system) and a spiral (cold, rotating stellar system)
- explain what is a continuity equation, and derive the collisionless Boltzmann equation, Eq.(8.20), stating the underlying assumptions.
- explain that the ellipticity of ellipticals does not results from rotation, but from the anisotropy of the pressure tensor.

Chapter 9

Groups and clusters of galaxies

Galaxies are not distributed homogeneously in the Universe. Some regions in the Universe have a galaxy density (i.e. the number of galaxies per Mpc^{-3}) significantly lower than the mean, and some a significantly higher density. The main reason for this is that gravity amplifies small inhomogeneities: if a region is slightly denser than its surroundings (for example a group of galaxies), it will tend to attract more matter than the surrounding regions, and hence become even more massive (turn into a galaxy cluster, say). This is the essence of hierarchical structure formation: small structures merge together to make more massive one.

The galaxies huddle together in groups of a few galaxies – such as the MW and M31 in the Local Group – and some are in clusters with hundreds of galaxies. The groups and clusters themselves are clustered in super clusters (i.e. clusters of clusters). But there are no clusters of super clusters: on sufficiently large scales the Universe is not a fractal but becomes homogeneous.

9.1 Introduction

Recall that the MW, M31, and some 40 small irregular galaxies within 2Mpc or so from the MW, are part of a gravitationally bound system, called the Local Group. Several other such bound groups occur at larger distances from the MW, and searches over larger volumes have revealed the existence of 1000s of such groups of galaxies.

Within about 20Mpc from the MW are several clusters of galaxies, containing 10s of large galaxies, and hundreds of smaller ones each. The *Virgo* cluster of galaxies contains the big elliptical M87. We already discussed the big cD galaxy NGC1399 at the centre of the Fornax galaxy cluster. The typical total mass of these clusters is estimated at around $10^{14} M_\odot$. At 90Mpc we find the Coma cluster of galaxies, with a mass of may be as large as $10^{15} M_\odot$. Clusters such as Coma contain the most massive known galaxies.

In the 1950s, George Abell and collaborators eye balled photographic plates to look for rich clusters of galaxies, by finding regions with a large over density of galaxies. He found some 4500, and ranked them from poor to rich, according to the number of galaxies they had. Paul Hickson later performed a similar exercise looking for 'Hickson' compact galaxy groups.

Clusters of galaxies are the most massive virialised structures in the Universe. Their crossing time[1] is sufficiently small that each galaxy has had the opportunity to cross the system several times, and hence the system has had time to dynamically relax to a near equilibrium situation, i.e. the system is close to being in virial equilibrium. Although more massive bound systems exist, these have crossing times which are so large that, given the age of the Universe, they have not had time to become dynamically relaxed.

9.1.1 Evidence for dark matter from galaxy motions

Evidence for the existence of dark matter in galaxy clusters dates from the 1930s. The Swiss astronomer Fritz Zwicky used the velocity dispersion of galaxies in clusters as determined from Doppler shifts to estimate their dynamical mass. Suppose all galaxies have the same mass m, then the kinetic energy of the system of N galaxies is

$$K = \frac{1}{2} \sum m\, v^2 = \frac{1}{2} M\, \sigma^2 \,, \tag{9.1}$$

where $M = N\, m$ is the total mass of the cluster galaxies, and the velocity dispersion σ^2 is defined as

$$\sigma^2 = \frac{\sum m\, v^2}{M} \,. \tag{9.2}$$

The potential energy in the system is of order

[1]If v is the typical velocity of a galaxy in a cluster of galaxies, and R is the radius of the cluster, then the crossing time $T_{\rm cr} = R/v$, i.e. it is the typical time a galaxy takes to cross the cluster once.

$$U = -\frac{G\,M^2}{R}\,, \tag{9.3}$$

where R is a measure of the size of the system. When the system is in *virial equilibrium*, $2K = |U|$ hence

$$M = \frac{R\,\sigma^2}{G}\,. \tag{9.4}$$

Note that the *measured* rms velocity is just the line-of-sight one, say $\sigma_z^2 = \langle(v_z - \langle v_z\rangle)^2\rangle$, but if the velocities are isotropic, then $\sigma^2 = 3\sigma_z^2$. So from that, and from an estimate of R, one can estimate the dynamical mass M.

Zwicky estimated the masses of the galaxies, and compared the result with the dynamical mass. Not only was the dynamical mass much larger, the velocities of the cluster galaxies was so high that, given the amount of mass inferred to be in the galaxies, the cluster was not even a gravitationally bound system. Left on its own, it would cease to be a high density region of galaxies on a short time scale because the galaxies would fly apart very rapidly. So either clusters were just chance superpositions of galaxies, or there was much more mass in them than was inferred from just the galaxies. Zwicky concluded that most of the mass in clusters must be invisible.

9.2 Evidence for dark matter from X-rays observations

Just as elliptical galaxies, clusters emit X-rays due to thermal bremsstrahlung produced in the highly ionised gas bound by the gravitational potential well of the cluster. This can be used to estimate the mass required to bind this gas, as we did when deriving Equation (7.13) from assuming hydrostatic equilibrium in an isothermal gas. This type of analysis again strongly suggests the presence of dark matter. From the pictures of the X-ray gas, it is also immediately clear that the X-rays come from all over the cluster, they are not obviously associated with individual galaxies.

From the X-ray intensity distribution, one can also estimate the amount of hot gas present in the cluster. From modelling the colours of the galaxies, one can estimate the mass in stars. Comparison of these two shows that *hot gas is the dominant baryonic mass by a large factor*. So most of the baryonic mass in a cluster of galaxies is actually in hot gas and not in galaxies at all.

The temperature of the gas is so high that it balances the gravitational force of the dark matter in the cluster. It gets to this high temperature by *shock heating*. This is how it works: as the cluster is accreting matter from its surroundings, newly accreted gas falls into the cluster at high velocity. The velocity is high, because the gas has been accelerated by the gravitational force from the tremendous amount of mass in the cluster. The high velocity gas slams into the stationary hot gas and converts most of its kinetic energy into thermal energy: this is *an accretion shock*.

Suppose a parcel of gas starts at infinity with zero velocity, and falls into a cluster with mass M and radius R. By the time it reaches the outskirts of the cluster at radius R it will have an in fall velocity

$$\frac{1}{2}v_{\rm infall}^2 = \frac{G\,M}{R}\,, \tag{9.5}$$

where I've used energy conservation. When it hits the cluster gas, it will convert this kinetic energy into thermal energy, hence it will be heated to a temperature T

$$\frac{3}{2}\frac{{\rm k_B}T}{\mu} = \frac{1}{2}m_{\rm p}\,v_{\rm infall}^2\,. \tag{9.6}$$

Here, μ is the mean molecular weight of the gas. Combining the last two equations, we find that the temperature of the cluster will be

$$\frac{3}{2}\,\frac{{\rm k_B}T}{\mu m_{\rm p}} = \frac{G\,M}{R}\,. \tag{9.7}$$

This temperature is called the *virial temperature*.

Where does all this gas come from? The X-rays we observe from clusters imply that the hot gas is loosing energy and hence is cooling. So from the observed X-ray luminosity, we can estimate the cooling time of the gas $t_{\rm cool}$,

$$\frac{dT}{dt} \equiv \frac{T}{t_{\rm cool}}\,. \tag{9.8}$$

Given the present cooling rate, dT/dt, the gas will have lost it's thermal energy after a cooling time, $t_{\rm cool}$. The measured values of $t_{\rm cool}$ turn out to be longer than the age of the Universe for most clusters[2]. This means that *once the gas gets hot it will remain hot*, and cannot form into galaxies anymore. So clusters of galaxies are in fact regions where galaxy formation is strongly suppressed because the gas is too hot to cool and form stars.

But why did it not form galaxies *before* it became hot? This is actually a rather hotly debated issue at the moment, but here's a hint:

9.3 Metallicity of the Intra-cluster medium.

The X-ray spectrum of cluster gas consists of the usual featureless thermal bremsstrahlung continuum, with additional emission lines mostly from inner-shell transitions of the Fe ion, see Figure 9.1. By modelling the spectrum, one can estimate the amount of Fe present in the gas. Comparison with the total amount of gas gives a mean metallicity $M_{\rm Fe}/M_{\rm gas} \approx 1/3$ of the solar value[3].

This is highly surprising: it means that the gas in clusters has a higher metallicity than for example globular cluster *stars*, or even the stars and gas in most dwarf galaxies. The only way we know how to produce Fe is in stars – in fact we think that most of the Fe is produced during the thermo-nuclear deflagration of a Super Nova. Somehow Fe produced in stars, themselves presumably inside a galaxy, was able to escape from that galaxy and end-up in the intra cluster medium. If the metals were able to escape, then may be also some of the explosion energy was able to escape the galaxy, and heat the surrounding gas, preventing gas cooling and galaxy formation. So the presence of hot metal enriched gas in clusters, suggests that star formation in proto-clusters provided a negative feedback mechanism, which kept most of the gas at temperatures too high for efficient galaxy formation.

9.4 The dark matter density of the Universe

Clusters provide a very nice way to estimate the total amount of dark matter in the Universe. The argument goes like this.

[2]Except in the inner parts of some cooling flow clusters

[3]Recall that the sun contains a mass fraction of 0.02 of metals.

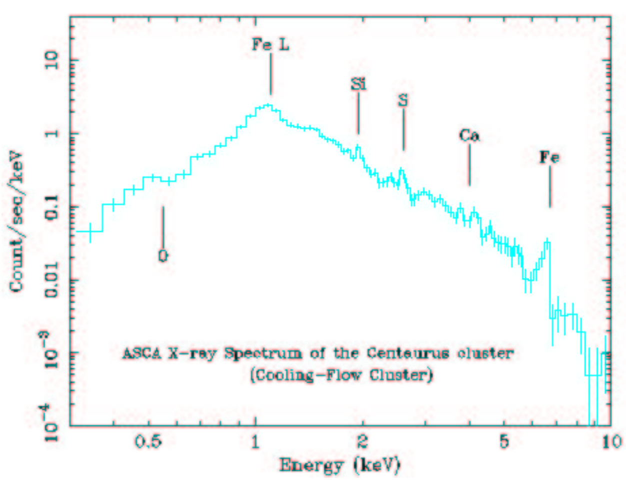

Figure 9.1: X-ray spectrum of the central region of the Centaurus cluster, taken by the ASCA X-ray satellite. It shows some emission lines and an absorption edge (which are labelled) due to different components of the hot intra cluster medium.

Suppose the Universe starts-out smooth, with a (nearly) constant ratio ω of dark matter to gas:

$$\omega = \rho_{\text{dark matter}}/\rho_{\text{gas}} \,. \tag{9.9}$$

As a cluster starts to form, the density of dark matter and of gas will increase. But the potential well of the cluster is so deep, that it is probably a good approximation to assume that none of the dark matter, nor the gas, can ever escape the gravitational pull of the cluster. This means that the ratio of the total dark matter to total gas mass *in the cluster*

$$\frac{M_{\text{cluster-in-dark-matter}}}{M_{\text{cluster-in-gas}}} \approx \omega \,. \tag{9.10}$$

We could do slightly better by also including the mass in stars – but that is a small correction anyway.

So we can find ω by determining $M_{\text{cluster-in-dark-matter}}$ from dynamics, and $M_{\text{cluster-in-gas}}$ from the X-ray emissivity. The result is $\omega \approx 6$.

Next we need an estimate of the mean baryonic density, ρ_{gas}. An elegant way is by determining the *Deuterium* abundance of gas[4]. Deuterium is produced in the Big Bang and destroyed in stellar burning. So if we find Deuterium, we know it is left over from the Big Bang. Analysis of the spectra of quasars (we will discuss quasars soon) allows us to measure the Deuterium abundance quite accurately. How does this help us? The missing link is that *the gas density* ρ_{gas} *determines how much Deuterium is produced in the Big Bang*. Figure 9.2 shows how the abundance (with respect to hydrogen) of elements produced in Big Bang nucleo-synthesis, as function of the baryon density. The result is that the mean baryon density corresponds to of order $\rho_{\text{gas}}/m_{\text{p}} \approx 2.2 \times 10^{-7}$ hydrogen atoms per cm^{-3} (i.e. 3.75×10^{-31} g cm^{-3}. So the density of paper/density of the screen you are reading this on, is about 10^{30} times higher than the mean!

And so we're done: the Deuterium abundance determines the gas density through the known nuclear reactions that occurred during Big Bang nucleo-synthesis. And the dark matter density is $\omega\,\rho_{\text{gas}}$ with $\omega \sim 6$ determined from clusters.

[4]Recall that the nucleus of a Deuterium atom differs from that of ordinary Hydrogen in that it has a proton and a neutron.

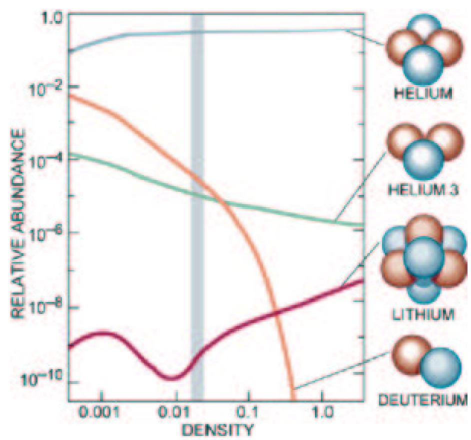

Figure 9.2: The production of various elements during Big Bang nucleo-synthesis as a function of the baryon density, taken from http://astron.berkeley.edu/˜mwhite/darkmatter/bbn.html.

9.5 Summary

After having studied this lecture, you should be able to

- Define what is a cluster and a group of galaxies by listing some of their properties.
- Explain how galaxy motions suggest the presence of dark matter in clusters.
- Explain how X-ray observations suggest the presence of dark matter in clusters.
- Explain the origin of the high temperature of the cluster gas.
- Explain how we know that most of the baryons in a cluster are in hot gas, and not in stars.
- Explain why the high observed metallicity of the cluster gas suggests that some kind of stellar feedback mechanism prevented most of the gas turn into stars before it fell into the cluster.
- Explain how clusters have been used to estimate the mean dark matter density of the Universe.

Chapter 10

Galaxy statistics

Having studied some properties of spiral and elliptical galaxies in detail, we now take a step back to examine the statistics of the different types. The two degree-field galaxy redshift survey (2dF), and the Sloan Digital Sky survey (SDSS), are two very recent *galaxy redshift* surveys. In a galaxy redshift survey, one first identifies galaxies in images of the sky, and then takes a spectrum for each of them. Because the Universe is expanding, spectral lines of stars in a distant galaxy are shifted from the laboratory toward longer wavelengths: they are 'redshifted'. There is a relation between the distance to the galaxy and its redshift, and so a galaxy redshift survey provides us with the 3D position of a set of galaxies, about 2×10^5 for the 2dF, about 10^6 for the SDSS. So what kind of general trends do they uncover? We will discuss some of these here.

The *distribution* of galaxies probably traces to some extent the distribution of dark matter in the Universe. For example, we've seen that there is evidence for a lot of dark matter in the Universe as a whole (from the cluster argument), but also in individual galaxies, groups and clusters. And so, since we find that galaxies are distributed very inhomogeneous in groups and clusters, presumably also the dark matter is distributed in an inhomogeneous fashion. How? And what process determines this in the first place?

We have a good idea of how the dark matter is distributed on a cosmological scale, from some general considerations, backed-up by observations of the micro-wave background (i.e. the sea of photons left-over from when the Universe was still hot). Given this input, we can use numerical simulations to predict how dark matter is clustered on large scales. The nice part of the story is that this theory is highly predictive, and does predict structures in the galaxy distribution very similar to what we observe: a filamentary pattern delineating large regions of space devoid of galaxies, called 'voids', with regions of high galaxy density, clusters of galaxies, at the intersection of the filaments.

10.1 The density-morphology relation

In the previous chapter we saw how some galaxies are in low density environments, some are in higher density *groups*, and others in dense *clusters* containing several hundred galaxies. Figure 10.1 shows that the *fraction* of elliptical and S0 galaxies increases rapidly with increasing galaxy density. In other words, in a region of high galaxy density, such as a dense group or cluster of galaxies, most galaxies tend to be S0 or E, and very few are of type S. Vice versa, in regions of low galaxy density, most galaxies tend to be of the type S. This is called **the density-morphology relation**, since it **relates the density of galaxies in a given region with their typical morphology** (i.e. S, E, S0). Clearly this is telling us something on the nature of galaxy formation and evolution.

Recall from the movie I showed of the motion of galaxies in a cluster, that encounters between galaxies are very frequent in a region of high galaxy density. Such tidal encounters could be responsible for converting S-type galaxies into S0 and E, but other processes might operate as well. Recall that we already discussed that this cannot be the whole story, since the stellar populations of Ss and Es tend to differ as well.

10.2 Galaxy scaling relations

10.2.1 The Tully-Fisher relation in spirals

The circular velocity as function of radius – the rotation curve – in spirals tends to be flat, i.e. the rotation velocity is independent of radius sufficiently far out in the disk. We argued in section 5.2 that this was evidence for the presence of dark matter in spiral disks, since it implies lots of mass in the

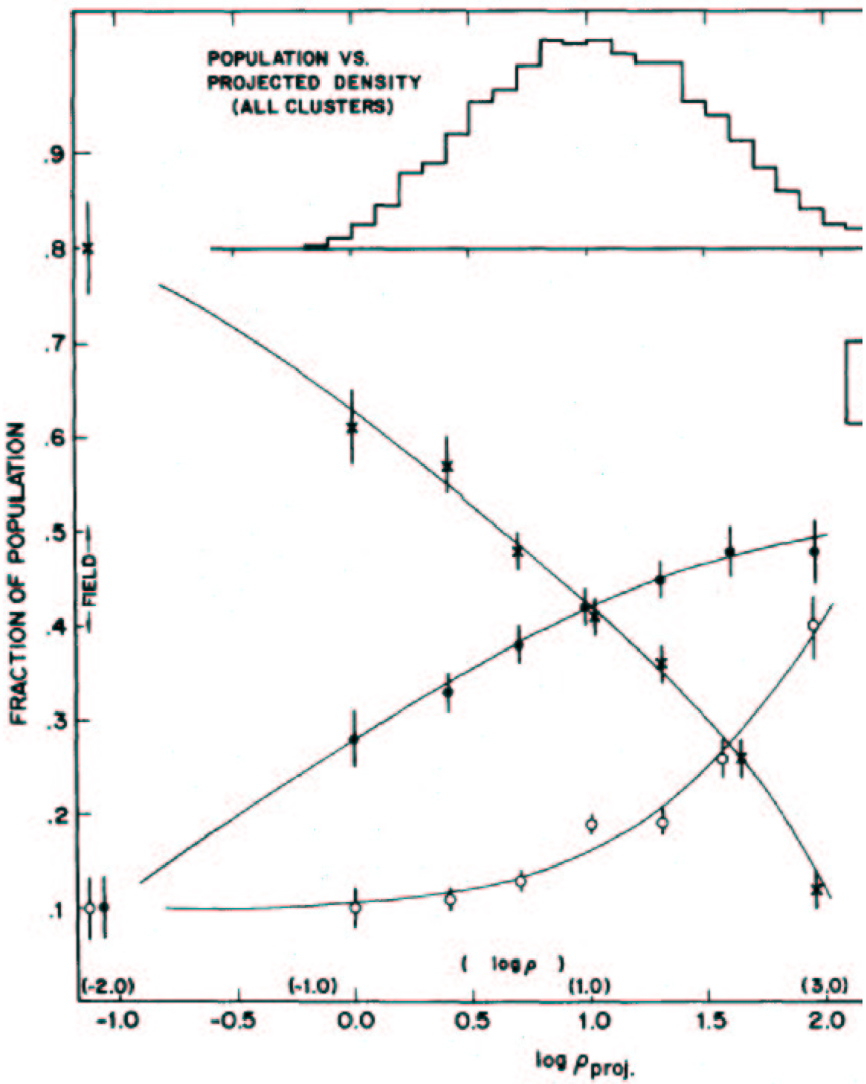

Figure 10.1: The fraction of Elliptical (E), S0, and Spiral + Irregular (S+I) galaxies as function of the logarithm of the projected density, in galaxies Mpc^{-2}. The data shown are for all cluster galaxies in the sample and for the field. Also shown is an estimated scale of true space density in galaxies Mpc^{-3}. The upper histogram show the number distribution of the galaxies over the bins of projected density. Es and S0s tend to populated regions of high galaxy density, and Ss regions of low galaxy density. From Dressler, ApJ 236, p. 351 (1980).

outer parts, whereas we see very little light there. Figure 10.2 compares the measured rotation velocity V_c (written W in the figure) with the total absolute magnitude $M = -2.5\log(L)$+constant, for a sample of spiral galaxies. **Brighter spirals** (more negative M) **rotate faster** (larger W). Since the figure shows there to be a linear relation between M and V_c, it implies a power-law relation between the intrinsic luminosity L and the circular velocity V_c, called the **Tully-Fisher relation**:

$$L \propto V_c^\alpha\,, \quad \alpha \approx 4\,. \tag{10.1}$$

The slope α of this relation depends somewhat on which colour is used (i.e., does L refer to a V -band or I -band luminosity).

What is the origin of this relation? Remember[1] that the circular velocity V_c is determined by the enclosed mass $M(<R)$,

$$V_c^2 = \frac{G\,M(<R)}{R}\,. \tag{10.2}$$

Let us introduce a global mass-to-light ratio, [M/L], as

$$[\mathrm{M/L}] \equiv \frac{M(<R)}{L}\,, \tag{10.3}$$

then

$$V_c^2 \propto [\mathrm{M/L}]\,\frac{L}{R}\,. \tag{10.4}$$

.

The mass-to-light ratio [M/L] will depend on the type of stars in the galaxy (which determines L), and the amount of dark matter (which mainly determines M). The intensity I is the luminosity per unit area, $I = L/(\pi\,R^2)$, or solving for R

$$R \sim \left(\frac{L}{I}\right)^{1/2}\,. \tag{10.5}$$

Combining the last two equations gives

$$L \propto \frac{V_c^4}{[\mathrm{M/L}]^2\,I}\,. \tag{10.6}$$

[1] It is unfortunate that absolute magnitude and mass are both denoted by M.

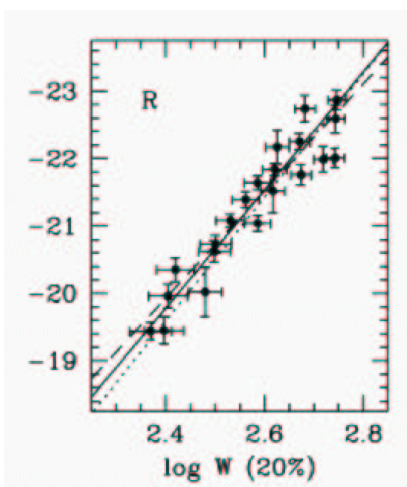

Figure 10.2: R-band absolute magnitude M of spiral galaxies as function of the maximum rotational velocity $W = V_c$ (in km s^{-1}). *Brighter* spiral galaxies (more negative M; recall that magnitude M is related to luminosity L as $M = -2.5\log(L)$+constant) rotate *faster*.

So the Tully-Fisher relation is reproduced, with $\alpha = 4$, if the intensity I and mass-to-light ratio [M/L] are *independent* of the luminosity of the galaxy. Notice that this is not a *derivation* of the Tully-Fisher relation: we are just trying to understand what is required for spirals to follow this relation. The fact that [M/L] and I are independent of L implies that somehow stars and dark matter are closely linked, or in other words: the star formation history is largely determined by the mass of the dark halo of the galaxy.

10.2.2 The Faber-Jackson relation in ellipticals

The Faber-Jackson law, Figure 10.3, relates the velocity dispersion σ of the stars in an E-type galaxy with the total luminosity L,

$$L \propto \sigma^4 . \tag{10.7}$$

Compare with the Tully-Fisher relation for spirals, Eq. (10.1): the velocity dispersion σ takes the role of the circular velocity V_c.

Comparison of Figure 10.3 with Figure 10.2 shows that the *scatter* in the Faber-Jackson relation is quite a bit bigger than the scatter in the Tully-Fisher relation: at a given value of σ, there is a range of ± 2 magnitudes in $M_{\rm B}$ (i.e. a factor 2.5 in L), versus a few tenths of a magnitude for a given value of V_c in the Tully-Fisher relation. So ellipticals follow a similar relation as spirals, with more luminous and hence more massive, ellipticals having a large velocity dispersion, but the relation is not as tight as the corresponding $V_c(L)$ relation in spirals.

From all the parameters we can measure for an elliptical, the total luminosity L, the velocity dispersion σ, the effective radius R_e and the intensity I_e (which both enter in the de Vaucouleurs profile, Eq. (3.2)), only three are independent. So we can try to obtain a tighter relation by introducing a second parameter in Eq. (10.7), for example R_e. This is how it goes: suppose that the galaxy is in virial equilibrium, then its kinetic energy $M\,\sigma^2$ will be proportional to its potential energy M^2/R_e, and so we expect

$$\sigma^2 \propto \frac{M}{R_e} . \tag{10.8}$$

Introducing again the mass-to-light ratio, [M/L] this can be written as

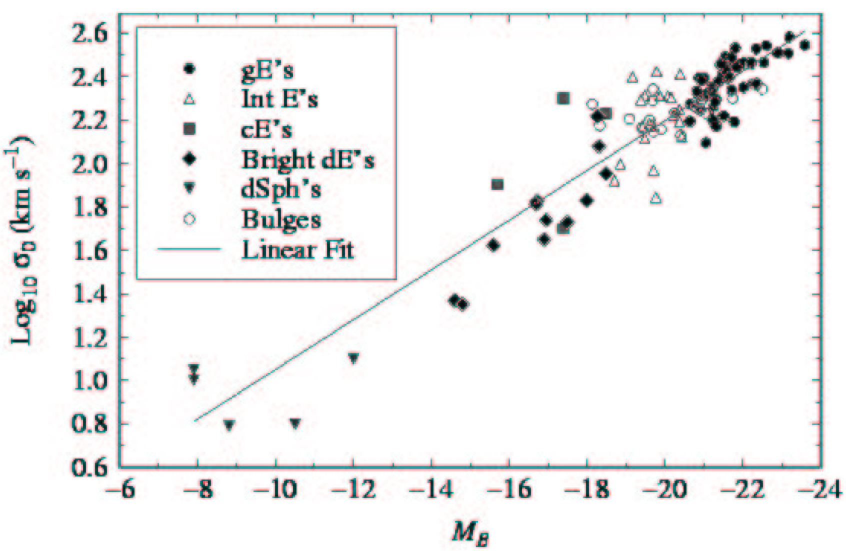

Figure 10.3: B-band absolute magnitude M for a sample of elliptical galaxies plotted versus the velocity dispersion σ (in km s^{-1}). *Brighter* ellipticals galaxies (more negative $M_{\rm B}$; recall that magnitude M is related to luminosity L as $M = -2.5\log(L)$+constant) have a *larger* velocity dispersion.

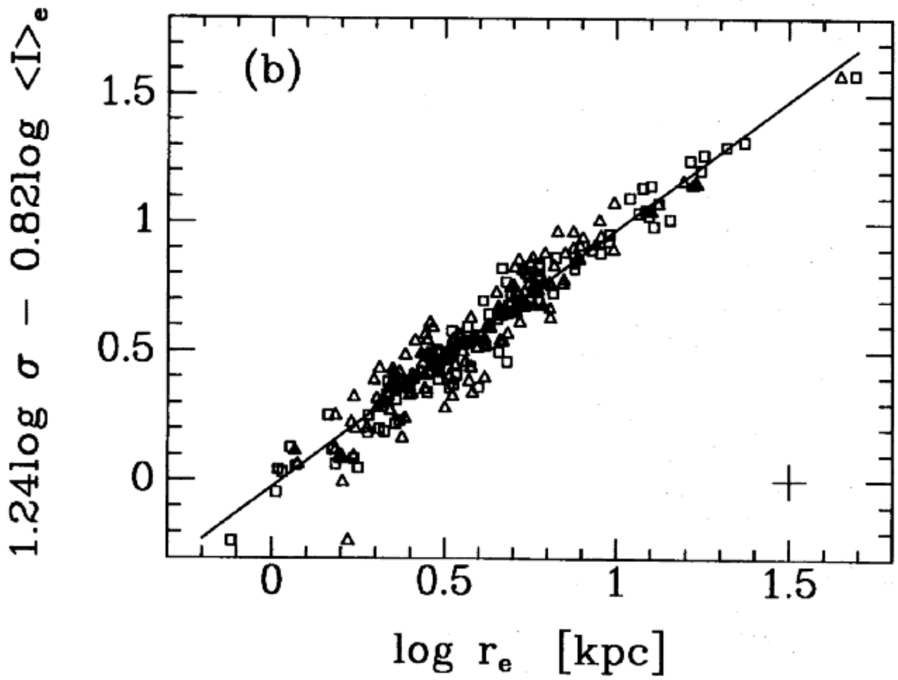

Figure 10.4: Scale radius R_e versus $\sigma^{1.24}\, I_e^{-0.82}$ for the same sample of Es plotted in Figure 10.3. The galaxies follow this relation much better, with much less scatter then the Faber-Jackson relation in Fig.10.3.

$$\sigma^2 \propto [\mathrm{M/L}]\frac{L}{R_e} \propto \frac{L^{1+\alpha}}{R_e}\,, \tag{10.9}$$

where I've assumed [M/L] to depend on L as

$$[\mathrm{M/L}] \propto L^{\alpha}\,. \tag{10.10}$$

The intensity $I_e \propto L/R_e^2$, hence

$$R_e \propto \sigma^{2/(1+2\alpha)}\, I_e^{-(1+\alpha)/(1+2\alpha)}\,. \tag{10.11}$$

Figure 10.4 shows that

$$R_e \propto \sigma^{1.24}\, I_e^{-0.82}\,, \tag{10.12}$$

for the same sample of ellipticals which was plotted in Figure 10.3. The galaxies follow this relation with very little scatter. Comparing the last two equations shows that if we assume $\alpha \approx 0.25$, then $R_e \propto \sigma^{1.33} I_e^{-0.82}$ – very nearly the same as what the figure suggests. Put differently, if the mass-to-light ratio of ellipticals depends on luminosity as $M/L \propto L^{0.25}$, then we can understand why ellipticals follow the relation plotted in Figure 10.4 so tightly.

The relation Eq. (10.12) is called the **fundamental plane** of elliptical galaxies. In four dimensional L, σ, R_e and I_e space, galaxies do not occupy the whole space, but are restricted to a 3-dimensional surface defined by relation (10.12), hence the name fundamental plane.

10.3 Tully-Fisher and Fundamental plane relations as standard candles

The Tully-Fisher relation Eq. (10.1) relates the luminosity L of a galaxy to its circular velocity. Now suppose we were able to measure the proportionality constant, by determining the luminosity for galaxies with known distance. Then by just measuring the circular velocity of a spiral galaxy, we could infer its luminosity hence absolute magnitude M, and consequently from its apparent magnitude m obtain the distance r, since

$$m - M = 5\log(r) - 5\,. \tag{10.13}$$

Therefore, the Tully-Fisher relation can be used as a standard candle, since it allows us to determine the luminosity from the distance independent quantity V_c. Note also that V_c is relatively easy to determine from spectra, and so we have found a good way of measuring distances to distant galaxies!

The fundamental plane relation plotted in Figure 10.4 can also be used as a standard candle: all we need to do is measure the velocity dispersion of the stars σ from a spectrum, and determine the intensity I_e (recall that the surface brightness and hence also I_e is distance independent). By putting the E galaxy onto Figure 10.4, we then find R_e. And so from the apparent size of the galaxy, we can estimate its distance.

Both these methods are widely used, since measuring velocities is relatively easy from a good quality spectrum, and can be done even for faint and/or distant galaxies. And since the scatter in the relations is small, we can obtain a relatively accurate distance estimate.

10.4 Galaxy luminosity function

From a big galaxy redshift survey, we can make a histogram of the number of galaxies with absolute luminosity between L and $L+dL$, where L is either the total (or bolometric) luminosity, or the luminosity in some band, e.g. the B-band, per unit volume. Figure 10.5 shows that the number of faint galaxies increases like a power law, whereas there is an abrupt cut-off toward the brighter galaxies. This behaviour is well captured by the **Schechter** function

$$\frac{dN}{dL/L_\star} = N_\star \left(\frac{L}{L_\star}\right)^\alpha \exp(-L/L_\star)\,. \tag{10.14}$$

This function has three parameters, $L_\star$, $\alpha \sim -1$ and $N_\star$. For luminous galaxies, it has an exponential cut-off for galaxies brighter than $L_\star$. For less luminous galaxies, the number increases toward fainter galaxies, as $L^\alpha \sim 1/L$. Finally, $N_\star$ is a normalisation constant, which determines the mean number density of galaxies.

Values are $N_\star \approx 0.007\mathrm{Mpc}^{-3}$, the steepness of the faint-end slope $-1.3 \leq \alpha \leq -0.8$, and the exponential cut-off sets in at $L_\star \approx 2 \times 10^{10} L_\odot$ in the B-band. From Table 3.1, we find that the total B-band luminosity of the MW is about $1.8 \times 10^{10} L_\odot$, so the MW is just slightly fainter than an $L_\star$ galaxy.

If the faint-end slope is steeper than $\alpha = -1$, then the total number of galaxies per unit volume, $n = \int_0^\infty (dN/dL)\,dL$ diverges, because there are so many faint galaxies. In reality, this does not happen of course, since there must be some finite number of small galaxies. The mean luminosity density can be found from integrating

$$l = \int_0^\infty L\,\frac{dN}{dL}\,dL\,. \tag{10.15}$$

This is the mean luminosity per unit volume. Combining this with the mean density of the universe as obtained in section 9.4, we can estimate the mass-to-light ratio for the Universe as a whole!

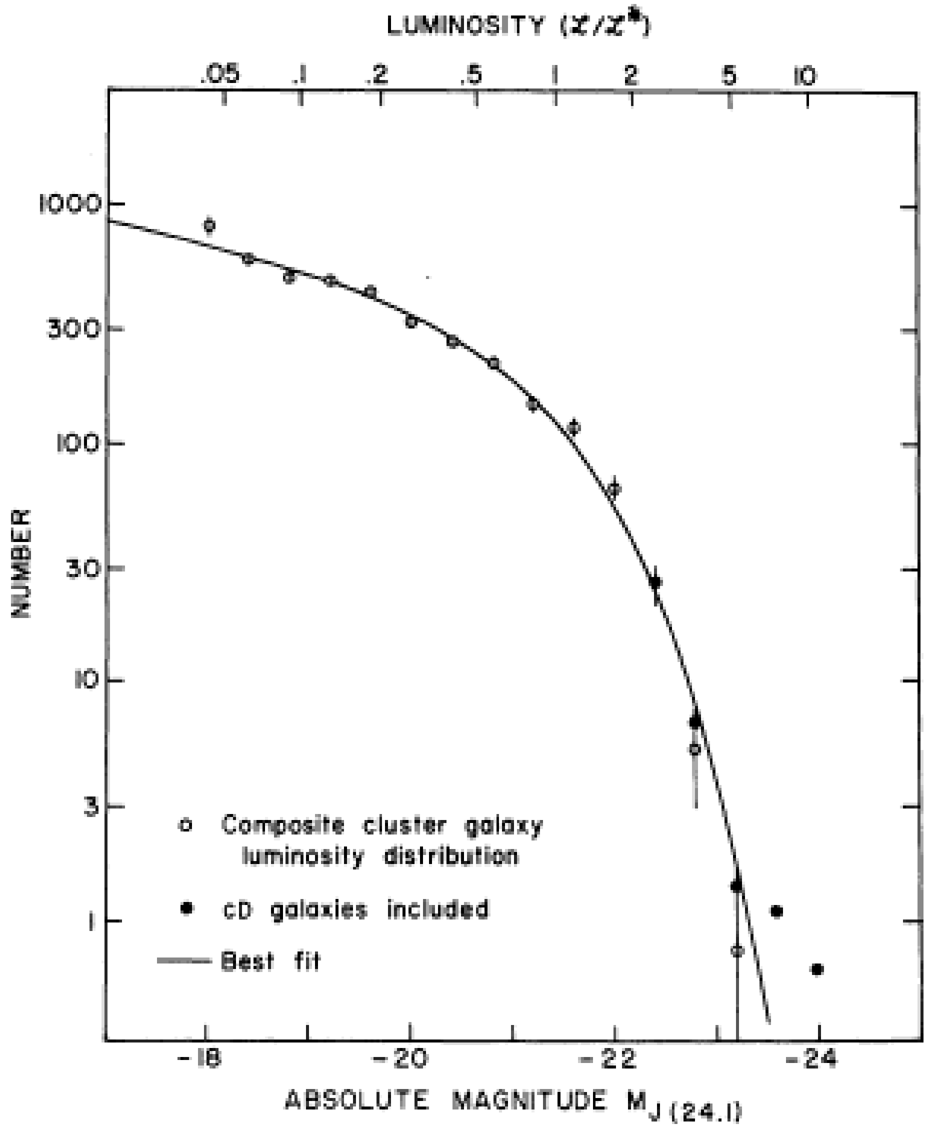

Figure 10.5: Histogram depicting the number of galaxies per Mpc^3 with given luminosity L in a given colour band. At magnitudes fainter than -18, the number increases as a power in L, but for magnitudes brighter than -22, the number of galaxies cuts-of exponential. The drawn line is the best-fit Schechter function.

Finally, Figure 10.6 shows how the Schechter luminosity function is composed of contributions from the various galaxy types, E, S0 and the different types of spirals, and irregular galaxies. Top panel is for a sample of galaxies away from big clusters, bottom panel refers to the Virgo cluster. Here we recognise the density-morphology once more: most of the brighter galaxies in the top panel are Spirals, whereas in the cluster sample, E an S0s play a much bigger role.

10.5 Epilogue

The classification of galaxies that we have used so far, in terms of spirals and ellipticals, is based largely on how these galaxies appear in pictures. With several 10^5 galaxies in modern surveys, it is not really possible to eye-ball all of them in order to classify them. In addition, it is not even necessarily the best way, and certainly not the only way, of classifying galaxies and identifying different types. With so many spectra and images available, new statistical techniques are presently being developed, where computer programs which do not suffer from human prejudices, such as principle component analysis and neural networks, decide how galaxies should be grouped and classified.

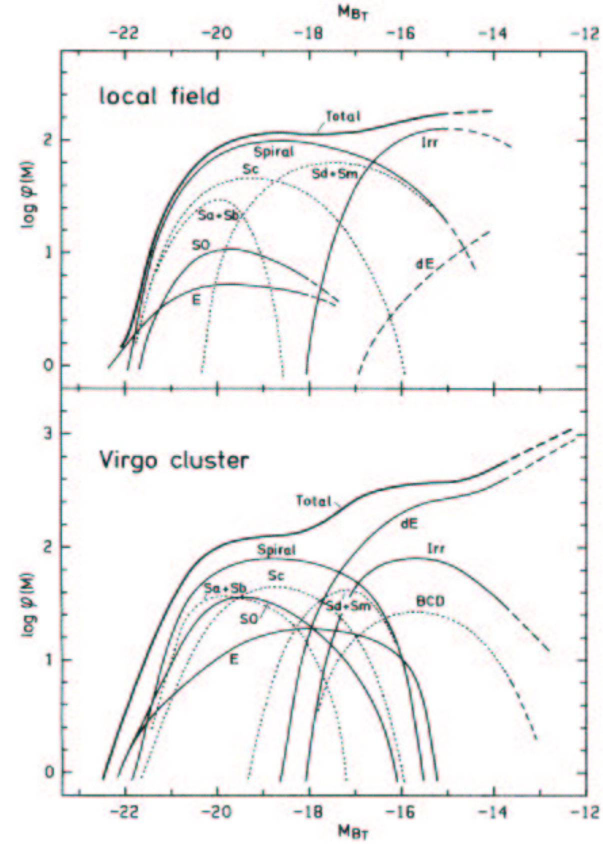

Figure 1 The LF of field galaxies (top) and Virgo cluster members (bottom). The zero point of $\log \varphi(M)$ is arbitrary. The LFs for individual galaxy types are shown. Extrapolations are marked by dashed lines. In addition to the LF of all spirals, the LFs of the subtypes Sa+Sb, Sc, and Sd+Sm are also shown as dotted curves. The LF of Irr galaxies comprises the Im and BCD galaxies; in the case of the Virgo cluster, the BCDs are also shown separately. The classes dS0 and "dE or Im" are not illustrated. They are, however, included in the total LF over all types (heavy line).

Figure 10.6: This plot shows how the galaxy luminosity function for galaxies in a cluster (bottom panel is for the Virgo cluster) or galaxies not in a cluster ('field galaxies', top panel), is made-up from the mix of galaxy types.

10.6 Summary

After having studied this lecture, you should be able to

- describe the density-morphology relation for galaxies
- explain what is the Tully-Fisher relation, and derive what it implies for the mass-to-light ratio and surface brightness of spirals
- explain the Faber-Jackson and Fundamental Plane relations for Ellipticals, and derive what the fundamental plane relation implies for the mass-to-light ratio of ellipticals
- explain how Tully-Fisher and Fundamental Plane relations are used as standard candles.
- explain what is the luminosity function of galaxies and sketch it.

Chapter 11

Active Galactic Nuclei

Some galaxies harbour immensely powerful sources of radiation in their centres, detected from the longest (radio) to the shortest wavelengths (X-rays): these are called Active Galactic Nuclei (or AGN for short). We think they are powered by the accretion of matter onto super massive black holes (SMBH), the same process that powers some galactic X-ray binaries. However, the SMBH have masses ranging from 10^6 to $10^9 M_\odot$, and so, unlike galactic black holes, cannot be stellar remnants. We think that *most* or even all reasonably massive galaxies have a SMBH lurking in their centres, but only a small fraction of these are 'active' at any given time. The evidence that the MW has its own SMBH is particularly impressive.

Where do these SMBHs come from? Are they really black holes (i.e. objects with an event horizon) or are they just very massive, extremely dense objects (MDO)? Why are some associated with strong radio-sources (radio-loud AGN), but others are not? How do radio-loud AGN power the immense radio-lobes we observe? What is the effect of the AGN on the galaxy itself? Most of these questions remain without a clear answer, even today.

I will start by discussing some of the observations of AGN. Then I'll take you through the evidence that these really are black holes, but leave it up to you to decide whether you believe it or not. My own view is that the evidence for super massive and very dense objects is very convincing – but I have not seen evidence for an event horizon (yet).

11.1 Discovery and observational properties

Carl Seyfert reported in 1943 that a small fraction of galaxies have a very bright nucleus, which is a source of broad emission lines produced by atoms in a wide range of ionisation stages. These nuclei appear unresolved, i.e. they are stellar in appearance.

After World War II, radio surveys discovered a very bright radio source, called Cygnus A. Given the accurate radio-position of the source, the optical counterpart was identified with a peculiar-looking cD-galaxy, whose centre is encircled by a ring of dust.

The spectral lines in the optical spectrum of Cygnus A are redshifted to $\Delta\lambda/\lambda_{\rm laboratory} = 0.057$. In terms of a Doppler shift, this corresponds to a recession velocity of 16000km s^{-1}. From this, one can estimate the distance to Cygnus A[1], $d \sim 240$Mpc, and hence compute the implied luminosity of the radio source, which turned out to be roughly ten times the *total* energy output (i.e. over all wavelengths) of the MW.

In 1960, Thomas Matthews and Alan Sandage were trying to find the optical counterpart of another bright radio-source, called 3C 48[2]. They found a 16th magnitude star like object, with a very curious spectrum, displaying broad emission lines that they could not identify with any known element or molecule. Sandage put it very scientifically: 'This thing is exceedingly weird'. In 1963, another such weird object, 3C 273 turned-up. Given they looked like stars, in that they were point like (unlike galaxies which are extended), they were called quasi-stellar objects, or QSOs.

In the same year, the Dutch astronomer Maarten Schmidt recognised that the pattern of lines in 3C 273 was similar to the Balmer series[3] of the Hydrogen atom – but only if they were redshifted to the (improbably large ?) value of $\Delta\lambda/\lambda = 0.16$, implying a recession velocity of $0.16c$ and a distance of 630Mpc. The distance to 3C 48 was even greater: a recession velocity of

[1]The redshifting of the spectral lines is due to the expansion of the Universe. By measuring the expansion rate, one can convert redshift to distance, assuming a cosmological model.

[2]The C stands for Cambridge – this is the third Cambridge radio-survey.

[3]The Balmer series is the series of electronic transitions in an atom which starts with the Hα line ($n = 3 \rightarrow n = 2$), then Hβ ($n = 4 \rightarrow n = 2$) etc.

0.367c and a distance[4] of 1800Mpc.

For a distance of $d = 630$Mpc to 3C 273, the distance modulus $m - M = 5\log(d/10) \approx 39$. The apparent magnitude of 3C 273 is $m = 12.8$ (in the V-band), hence the absolute magnitude is $M = -26$. Since the Sun's absolute magnitude $M_\odot = +5.$-ish, it implies that 3C 273 is $10^{0.4\times31} \sim 10^{12}$ times brighter than the Sun, or 100 times as bright as the entire MW. Give or take a factor of a few. What kind of object can be so small yet so tremendously powerful?

Since then, many more QSOs have been discovered, we know several hundred thousands of them by now. The current[5] record holder has $\Delta\lambda/\lambda = 6.3$. In terms of recession velocity, this is 6 times the speed of light (!), showing that interpreting such a redshift in terms of Doppler shift is not a good idea[6]. The luminosity of the brighter QSOs can range up to 10^5 times the luminosity of the MW.

The shear luminosity of QSOs alone already suggests we are looking at something unusual. In addition, some of these QSOs have radio-lobes which are tens of times the size of a galaxy. What is the engine that powers them?

11.2 The central engine, and unification schemes

QSOs certainly are very bright. But are they *unusually* bright? Yes. The argument is neat, and based on the *variability* of QSOs. The luminosity of QSOs varies on a range of time-scales – hours, days, and months, depending on the wavelength at which you observe them. But variations in the X-ray properties tend to have very short time-scales, of order of minutes. Now this implies that the X-rays come from a small region – of the order of light minutes. To come to this conclusion, all we need is to say that the information

[4] The 'velocity' of such distant objects is not a real velocity in the sense that something is moving. The velocity results from space expanding, and it is often better to not convert this to velocity at all – just use the cosmological model to find the relation between the redshift $z = \Delta\lambda/\lambda$ and distance.

[5] March 2003. The text book we're using, OC, states we know 5000 of them. And that we know 'several' with redshift greater than 4. This just shows you how rapidly this field is evolving.

[6] You may argue we should use special relativity. In fact, we need general relativity to make proper sense of this.

about the variation has to be able to travel across the object that is producing the emission, and the fastest the information can travel is the speed of light, c. Hence the engine is small, of the order of light minutes. Recall that the distance to the Sun is 8 light minutes[7]. And so we need to find a small engine, size of order a solar system, that can produce as much energy as all the stars in 100 galaxies *combined*.

The most efficient way we know of for producing energy is by mass accretion onto a compact object[8]. Recall from the stellar part of this course, that stellar X-ray binaries produce their energies this way. The estimated efficiency (in terms of the rest mass energy mc^2 of the fuel converted into radiation) is 10 per cent, which you can compare to the second most efficient way, nuclear fusion, at 0.7 per cent.

Recall also from those lectures that there is a maximum luminosity – the Eddington luminosity – that a spherical object in equilibrium can produce. If the luminosity is higher, than the radiation pressure becomes larger than the gravitational force, and the object can no longer remain bound. Comparing the Eddington luminosity with the QSO's luminosity, gives us a lower limit to the masses involved, of order $10^8 M_\odot$. Given such a small size, and such a large mass, Donald Lynden-Bell and Martin Rees suggested that accretion onto a massive black hole must be the energy source powering QSOs.

There is a problem though. A black hole is characterised by just two numbers: its mass, and its spin (i.e. its rotational energy). But there is a wide variety of QSOs out there. For example, some have jets, and strong radio sources, others don't. The time variability is by no means regular. How can that be, if there is only two parameters that characterise the engine?

The gas is thought to accrete onto the SMBH by passing through an accretion disk, where it has to loose enough angular momentum to be able to accrete. This introduces a third parameter: the orientation of the disk with respect to the observer. The most popular unification schemes[9] suggest

[7]To do this properly, we need to take (special) relativistic effects into account. But if you do everything properly, you get a similar answer to what we found.

[8]Actually, annihilation of matter–anti-matter is even better!

[9]i.e. theories that try to explain QSO activity in terms of a single model – accretion onto a SMBH.

that orientation effects determine many of the observed properties of the QSO. But there is by no means a simple answer to understand the wide variety of properties.

11.3 Evidence for a SMBH

I[10] am going to take you through several lines of evidence that QSOs are indeed powered by a SMBH. You may think that the fact that we need an engine the size of the solar system that produces as much energy as 100 galaxies together, would seem enough of a proof on its own. So let me discuss just one alternative: suppose there is a dense cluster of stars in the centre of a galaxy hosting a QSO, and what we are seeing is the combined effect of several tens of SNe explosions. A single SN at its peak luminosity is about as bright as a whole galaxy, so we need about 100 SNe and we're done. Is this a reasonable model for AGN?

The model has withstood for several decades, but it does have some pitfalls. For example, we don't know how such SNe would generate the tremendous radio-lobes that some QSOs have. But then, we don't know how a SMBH does that either. As Luis Ho puts it: 'our confidence that SMBHs must power AGN largely rests on the implausibility of alternative explanations'. Most of the arguments in the next sections just suggest the presence of a very massive, dense object (MDO) in the centre of galaxies (not just in galaxies harbouring a QSO – in fact, some of the best evidence is in galaxies which do not have an active QSO), but not really require the MDO to be a *black hole*, i.e. an object with an event horizon.

11.3.1 Photometry

The density of stars near a MDO will rise very sharply in the region where the MDO dominates the mass. This basically comes from Jeans' equations expressing hydrostatic equilibrium, in the form

$$\frac{G\,M_{\rm DO}}{r^2} = -\frac{\nabla \rho \sigma^2}{\rho}\,. \tag{11.1}$$

[10]I will follow a nice review paper by Luis Ho, see http://nedwww.ipac.caltech.edu/level5/March01/Ho/Ho.html for the full text.

If the velocity dispersion does not rise very rapidly, $\sigma^2 \approx$ constant, then the density must rise very rapidly, $d\ln(\rho)/dr \propto -1/r^2$ hence $\rho \propto \exp(1/r)$. So a rapid increase in density of light toward the centre could signify the presence of a SMBH.

However, the region where the MDO dominates may be very small, of order $r_{\rm MDO} \approx GM_{\rm DO}/\sigma^2$. As viewed from Earth for a QSO at distance D, the angular extent will be $\approx$ 1arcsec$(M_{\rm DO}/2\times10^8 M_\odot)\,(\sigma/200\text{km s}^{-1})^{-2}\,(D/5\text{Mpc})$. Such a small extent is hard to see from the ground (due to atmospheric seeing), and Hubble Space Telescope (HST) will be needed.

Now some galaxies do indeed show a dramatic increase in brightness – a cusp – toward the centre, even at HST resolution. Unfortunately, it is also known that the velocity dispersion σ^2 is not constant. So we cannot claim this proves the presence of a MDO – it's a good indication though.

11.3.2 Stellar kinematics

The previous method was not perfect, since we did not take into account that σ may vary. Now close enough to the MDO where it dominates the mass, we should be able to detect the Keplerian rise in velocity dispersion $\sigma \propto r^{-1/2}$. So people tried to determine $\sigma(r)$ close to the centre.

However again there is a loop hole: $\sigma(r)$ can grow as well in case the velocity distribution becomes anisotropic, e.g. in spherical coordinates $\sigma_r^2 \neq \sigma_\theta^2 \neq \sigma_\phi^2$. Again using the Jeans' equations, the radial variation in mass can be related to the stellar velocity dispersion as[11]

$$\begin{aligned} M(<r) &= \frac{V^2\,r}{G} + \frac{\sigma_r^2\,r}{G} \\ &\times \left[-\frac{d\ln\nu\sigma_r^2}{dr} + (\frac{\sigma_\theta^2+\sigma_\phi^2}{\sigma_r^2} - 2)\right]. \end{aligned} \tag{11.2}$$

V is the circular velocity. Clearly, if we were allowed to *choose* the three components of σ^2, we could balance a large increase in density with the velocity tensor becoming more anisotropic, even in the absence of a MDO.

Observed galaxies do indeed show a large increase in σ close to the centre, but we cannot rule out that this is due to anisotropic velocities, and not to the presence of a MDO. So once more, the evidence is a bit circumstantial.

[11]I just state the result and don't expect you to derive it.

11.3.3 Stellar kinematics in the MW centre

The MW centre has a weak AGN in its centre, identified with the radio source Sagittarius A.

Over the past couple of years, Genzel and collaborators measured the kinematics of stars close to the MW centre from near-IR spectra[12] They also determined very accurate positions of the stars. By doing this over several consecutive years, they were able to measure not just the 3D proper motions of the stars, but also measure *accelerations*. In fact they were able to follow one star over nearly a whole orbit, when it swung passed the centre at velocities approaching 1000km s^{-1}.

So first of all, they demonstrated that there is indeed a very massive 'dark' object in the centre of the MW. Using Kepler's laws, they were able to put very stringent limits on the mass of the central concentration, $M_{\rm DO} \approx 2.6 \times 10^6 M_\odot$, in a region smaller than 0.006pc, implying an enormous mass density higher than $2 \times 10^{12} M_\odot/\mathrm{pc}^3$. This leaves *almost* no room to escape the conclusion that the MDO is indeed a SMBH, at least in the MW.

The presence of the large mass is also supported by the presence of other very high velocity stars. From that, and from the fact that the putative centre, i.e. the radio source Sagittarius A itself does not seem to move, Genzel derived a lower limit to the density of $3 \times 10^{20} M_\odot/\mathrm{pc}^3$.

Incidentally, they also showed that the velocity dispersion remains nearly isotropic, and so unless the MW is peculiar, this then strengthens the case for SMBHs in other galaxies, from the arguments in the previous section.

11.3.4 Methods based on gas kinematics

Unlike the situation for stars where anisotropy clouds the interpretation, gas kinematics is much easier to interpret if the gas is in Keplerian motion[13]. There are two caveats, however. There may be non-gravitational forces acting on the gas, magnetic fields for example. Second, there is no *a priori* reason for the gas to be in Keplerian motion. So we need to lucky then.

[12]Recall the need to go to the IR: there is a lot of absorption by dust toward the MW centre but IR-light undergoes far less absorption.

[13]Whereas stars can freely move through one another, gas motion must be ordered not to produce shocks.

Optical emission lines Several examples of nuclear disks around centres of galaxies were found by HST. For example for M87 (the big elliptical in the Virgo cluster), HST measurements found a Keplerian rotation curve in the disk with a velocity of 1000km s^{-1} at a distance of 19pc from the centre[14]. This implies a MDO with mass $\approx 2.4 \times 10^9 M_\odot$. Similar evidence for a MDO now exists for other galaxies as well.

Maser emission Some AGN have luminous maser[15] sources close to the centre. Radio observations of these maser lines show that the masers are in nearly Keplerian rotation, and from the speed one can place tight constraints on the mass density of the MDO, $> 5 \times 10^{12} M_\odot$.

Reverberation mapping Another elegant way to determine the extent of an AGN is 'reverberation mapping'. Many of the emission lines seen from an AGN are due to the light from the central source being reprocessed by the surrounding material. As the central source varies in luminosity, the emission lines vary as well, but with a time-lag that corresponds to the light-travel time between the central source and the line-emitting gas. So by measuring the time-lag, we can estimate the size of the light-emitting region.

11.3.5 Profile of X-ray lines

The Fe Kα line at 6.4keV is a common feature in the X-ray spectra of AGN. It is thought to arise from fluorescence of the X-ray continuum off cold material in the surroundings, possible associated with the accretion disk around the SMBH.

This line is produced very close to the SMBH, and it exhibits Doppler motions approaching relativistic speeds, $\sim 10^5$km $s^{-1} \approx 0.3c$. Now crucially, the line-profile displays asymmetries consistent with a *gravitational redshift*. The photons lose energy as they climb out of the deep potential well near the black hole. In doing so, their wavelengths become longer – a 'gravitational redshift'.

For the AGN galaxy MCG-6-30-15, the best fitting accretion disk has an inner radius of 6 Schwarzschild radii, i.e. *very* close to the event horizon. A

[14]Don't get confused with evidence for dark matter from *flat* rotation curves. Here we are looking close to the centre, whereas for rotation curves we were looking at large distances, $\sim$ 20kpc from the centre.

[15]A *maser* is a laser that operates in the mm regime.

similar analysis has now been performed for other AGN as well. **The shape of the line probably provides the best evidence to date for the existence of a SMBH.** However, other mechanisms for generating the line profile are possible, but implausible. With better data, detailed modelling of the profile even has the potential to determine the spin of the SMBH.

11.4 Discussion

So are you convinced of the existence of SMBH? I think the evidence from stellar motions around the MW centre is truly impressive. The Fe Kα line comes very close to convincing me as well. But I think we really need to see relativistic motions or phenomena originating close to the Schwardschild radius to clinch the matter.

So let me summarise. We have evidence that probably most, if not all, galaxies with masses of order the MW mass or higher, have a SMBH in their centres. It also transpires that the mass of the SMBH seems to correlate with the mass of the galaxy: the more massive the galaxy, the more massive the SMBH. What does that tell us about the formation of the galaxy or of the SMBH? We don't know (yet!).

Also, if most galaxies contain a SMBH, then why are there so many times fewer QSOs than galaxies? Indeed, QSOs are rare, so only a very small fraction of SMBH is active and produces AGN activity at any one time. The best guess is that there is no gas accreting onto those SMBHs: these inactive QSOs are starved for food. Still it may be a bit surprising that so many black holes are so very black – some gas is lost by the stars close to the centre, which somehow must be accreted onto the black hole *without* producing much radiation.

And finally: if a star ventures too close to a SMBH, then it may be torn apart as a result of the tidal forces from the SMBH. This could lead to a very bright flare, which would last for of order months. The realisation that SMBHs may be quite common provides fresh motivation to search for such flares. Upcoming missions designed to detect SNe will find these flares – if they exist.

11.5 Summary

After having studied this lecture, you should be able to

- describe observational properties of AGN.
- describe briefly the current paradigm for energy generation in AGN (mass accretion through a disk onto a SMBH)
- describe three arguments that suggest the presence of SMBHs in galaxies.

Chapter 12

Gravitational lensing

According to the theory of General Relativity, a massive object distorts the space around it. For example, the distortion in space due to the the Sun makes the planets go around it. When light travels through such a distorted space, it also gets deflected.

Therefore, in general, the gravitational field generated by a mass distribution will distort the images you see of background sources, pretty much like the hot air above a road distorts the view behind it. This is in fact a good way to think about this phenomena: it is as if space has an index of refraction which varies in space, because the gravitational field also varies in space. (Whereas it is of course the varying temperature that causes the refractive index to vary above the hot road.) This effect is called 'gravitational lensing', although 'gravitational miraging' might be a better description since the distortion is more 'blurring' (like the hot air) than focusing (like a proper lens).

A very good introduction to the subject is the recent review by Joachim Wambsganss[1]. Scientists are not very good at reconstructing the history of a subject, but Joachim claims that a German physicist in 1804 was the first to suggest that gravitational lensing might be observable, for example in terms of the deflection of star light passing close to the Sun.

There are several possible observable effects of gravitational lensing: (1)

[1]see http://www.livingreviews.org/Articles/Volume1/1998-12wamb

deflection: the image position is displaced from the source position, (2) amplification: the image may be become brighter or fainter, (3) distortion: if the source is extended (for example lensing of a galaxy by a galaxy cluster), then lensing may distort the image, (4) multiple images: in general a single source will be imaged in multiple images. All these effects have been observed.

In general, the deflection and amplification in real systems is usually very small, therefore our view of the Universe is not greatly distorted by gravitational lensing. But occasionally, some objects are lensed quite strongly. We'll start by deriving a very simple – but unfortunately not quite correct – expression for the deflection angle, but also state the correct result. Then we'll be armed with enough understanding to look at some applications of lensing, and you'll realise the tremendous potential of this technique, on a really wide variety of fronts, from constraining the nature of dark matter, via measuring the masses of galaxy clusters, to detecting planets and constraining models of stars.

12.1 The lens equation

12.1.1 Bending of light

Suppose a particle with mass m and velocity v flies with impact parameter b past an object with mass M. The gravitational pull of M will deflect m, and give it a component $v_\perp$ perpendicular to its initial v. For the case $m = M$, this is in fact the situation we envisaged when deriving the relaxation time in Section 8.1. There we found that (see Eq. 8.3) when the deflection angle α is small,

$$v_\perp = \frac{2GM}{b\,v}\,. \tag{12.1}$$

Convince yourself that this is the correct result also in case $m \neq M$. Finally, since $\tan(\alpha) = v_\perp/v$ and when α is small so that $\tan(\alpha) \approx \alpha$,

$$\alpha = \frac{2GM}{b\,v^2}\,. \tag{12.2}$$

Now here's the trick: the expression for α does not depend on the mass m of the particle being reflected, only on the mass M of the deflector. So we might be tempted to put $m = 0$, $v = c$ and claim this is the angle by which light will be deflected as well.

Applying this to starlight passing close to the sun, hence using $M = M_\odot$, $R = R_\odot = 7 \times 10^5$km, we get $\alpha = 0.83$arcsec – so $\alpha \ll 1$ is certainly satisfied. This was actually the value Einstein found in 1911, though one Johann Soldner already obtained it almost a century earlier.

Curiously, the answer is wrong! Once Einstein formulated his theory of General Relativity, he did the problem again, and found that the deflection angle is *twice* our 'Newtonian' value. So the correct deflection angle for light is

$$\alpha = \frac{4GM}{b\,c^2}\,. \tag{12.3}$$

and it was of course a great vindication of the theory when Eddington verified this during a solar eclipse in 1920. Note that the deflection angle increases with M and decreases with b – the closer you pass to the more massive an object, the bigger is the deflection.

In the case of the Sun, the deflection is so small that you might think it is of little use in Astronomy. This was actually true for a long time, but recently it has really taken off as a new way of looking at the Universe. Before I show you some applications, we need to do a little bit more maths first.

12.1.2 Point-like lens and source

Suppose the light source (S), lensing mass M (L) and observer (O) are exactly aligned, as in Fig.12.1. Also assume they are all point like (i.e. not extended like for example a galaxy). Light rays from S are deflected by L over an angle α toward O. Because of symmetry of the situation, O sees a ring of light around L: an Einstein ring. Of course the figure is not to scale: α should be small if we want to apply the lensing equation.

Introducing the distances between observer and lens D_{OL}, observer and source D_{OS}, and lens and source D_{LS}, we can compute R_E and the angle θ_E under which O sees the ring. Since $\psi + \theta_E = \alpha$ we get, using the lensing equation and the figure,

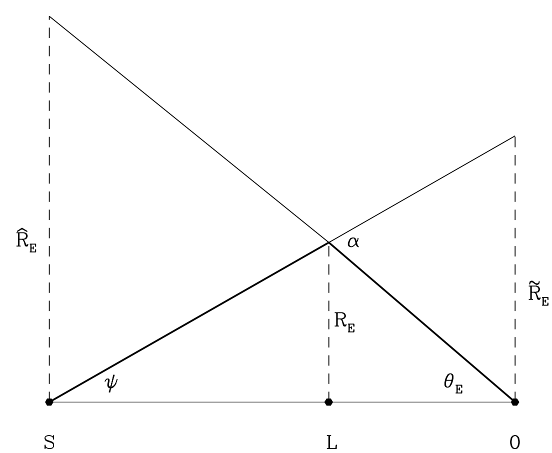

Figure 12.1: The source (S), lens (L) and observer (O) are all aligned. Light from S is deflected by an angle α toward O. Because of symmetry, O sees the source S as ring of radius R_E centred around L.

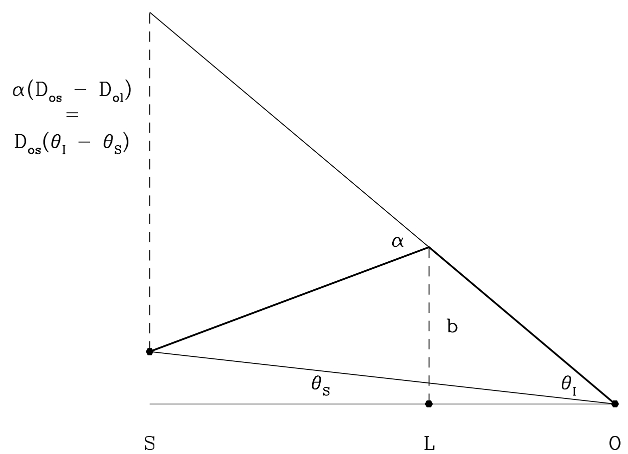

Figure 12.2: The source (S), lens (L) and observer (O) are now misaligned, and O sees two images of S, one of which is depicted here while the other one lies on the other side of L.

$$\begin{aligned}\theta_E &= \frac{R_E}{D_{OL}} \\ \psi &= \frac{R_E}{D_{SL}} \\ \alpha &= \frac{4GM}{R_E\, c^2}\,, \end{aligned} \tag{12.4}$$

and hence

$$R_E^2 = \frac{4GM}{c^2}\,\frac{D_{OL}(D_{OS}-D_{OL})}{D_{OS}}\,. \tag{12.5}$$

If O, L and S are not aligned, as in Fig.12.2, then O does not see a ring. Denote the angular position of S from the line OL by θ_S, and the angle between S and its image I by θ_I. Then from the figure you see that $\alpha(D_{OS}-D_{OL}) = D_{OS}(\theta_I - \theta_S)$. A bit of juggling and using the lensing equation gets you to:

$$\theta_I^2 - \theta_I\theta_S = \theta_E^2\,. \tag{12.6}$$

This is a quadratic equation for θ_I for a given θ_S and so there will be *two* images, at positions $\theta_{I,\pm}$.

Often, the change in position will be very small and not observable. But besides being lensed, an image will also be *magnified* (or de-magnified). So when the displacement is too small to see two images, you'll only see one but magnified by total magnification $A = A_+ + A_-$ of each image separately. Interestingly, for a small fraction of peculiar alignments, A may diverge, and so potentially you can get very large amplifications.

12.1.3 More realistic lensing

In the previous section we assumed that both lens and source where just point masses. When the lens is extended, you can get multiple images (3,5,7,...), depending on the impact parameter and the mass distribution of the lens. If the *source* is extended as well, then different parts of the source may be deflected and amplified in various ways, and so the image may be highly deformed. We are now in a position to look for applications. To wet your appetite, let me show you two beautiful examples.

Figure 12.3 shows an Einstein ring. The lens is probably an elliptical galaxy, and the source a more distant spiral. Because of the near perfect alignment, the spiral has been imaged into a ring.

Figure 12.4 shows how a nearby spiral galaxy has lensed a very distant background quasar and produced multiple images of it, in this case five.

12.2 Micro-lensing

If[2] both source and lens are stars, then we can use our results from the point-like models. We expect very small deflections, but potentially large magnifications. We only get large magnifications for chance alignments.

So as foreground stars move in front of background stars – for example a star in the Milky Way disk moves in front of a star in the bulge – it can occasionally amplify the bulge star by a large amount. However, since the alignment has to be very precise to get a large amplification, the bulge star will not be amplified for a long time, before the disk star moves away from the ideal lensing alignment.

Because the alignment has to be so good, you need to look at millions of stars to have a chance of seeing significant amplification. So people observed millions of stars toward the LMC and toward the bulge, to find stars that brightened by a lot over a few days. This phenomena is called *micro-lensing* (because the deflection angle is so small, $\sim 10^{-6}$arcsec).

There is one crucial aspect of lensing that we haven't paid attention to yet: the deflection and amplification is *independent of the wavelength of the light*, as you can see from Eq. (12.3). This is important if you want to detect micro-lensing, because the brightness of many stars varies anyway, because they are variable stars. But in general, the change in brightness for a variable star depends on the colour you observe it in. But the change in brightness due to micro-lensing is colour independent. This makes it possible to distinguish variable stars from lensed stars. In addition, the light curve – luminosity as function of time – is very easily computed in terms of M and the relative velocity of L and S. Combined, it allows one to confidently detect micro-lensing, and distinguish it from the case where the star is intrinsically varying. Figure 12.5 shows an example of micro-lensing of a star in the LMC,

[2] I took this part, including the figures, from W Evans' 2003 review, see astro-ph/0304252

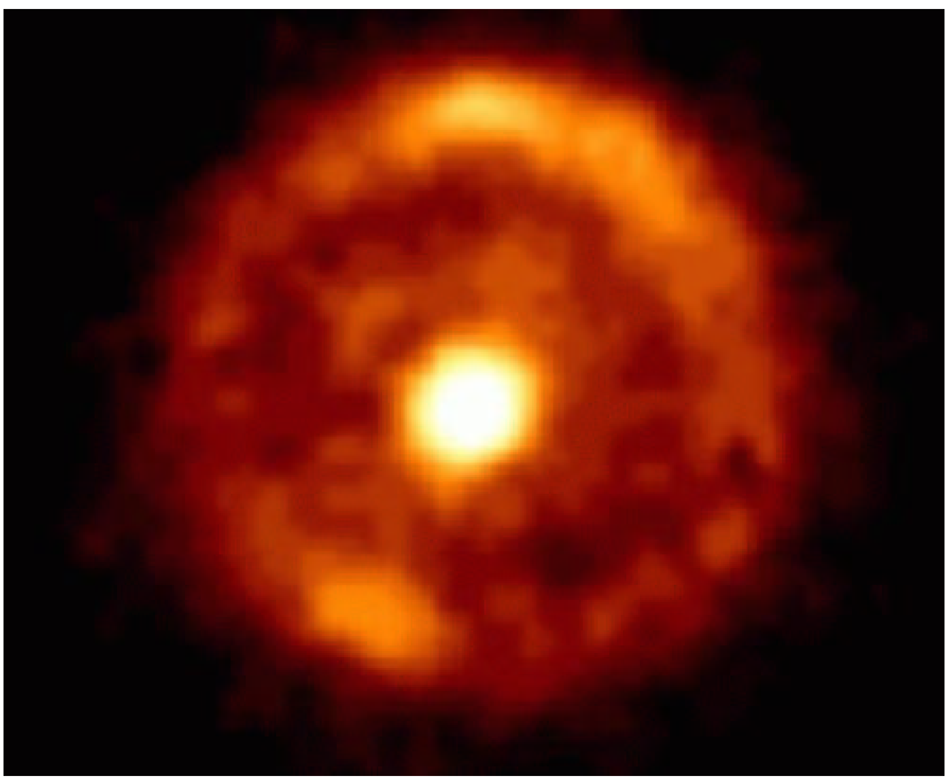

Figure 12.3: Hubble Space Telescope image of an Einstein ring, where a fortuitous alignment has caused an intervening elliptical galaxy (the central spot) to lens a background spiral galaxy into a ring.

Figure 12.4: Hubble Space Telescope image of the ‘Einstein cross’, where a distant quasar has been lensed into multiple images by an intervening spiral galaxy.

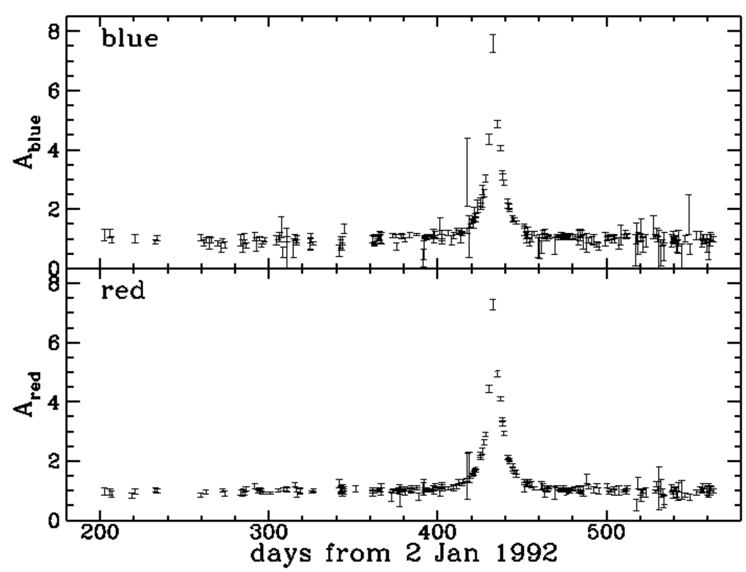

Figure 12.5: Luminosity of a star in the LMC in the blue (top) and red (bottom) as a function of time. Lensing causes it to brighten around day 420.

which brightens considerably around day 420, equally much in the blue as in the red.

Although we so far only considered lensing by stars, in fact any massive object will produce lensing. This has been exploited by the **MACHO collaboration**[3] to constrain the nature of the dark matter in the halo of the Milky Way. Indeed, suppose dark matter consisted of some type of massive objects, like black holes, or Jupiter-like objects, or some type of **MA**ssive **C**ompact **H**alo **O**bject in general. Since we know that the outer parts of the Milky Way halo is dominated by dark matter, there should be many such objects between the Sun and the LMC. If they exist, then we can detect them by their lensing signal.

To detect such MACHO's, the collaboration imaged millions of stars in the LMC for four years, and indeed found several lensing events like the one pictured in Fig.12.5. However, it is now believed that in each case, the lens is actually a normal star in the LMC, and not some exotic kind of matter. As a consequence, this experiment has now ruled-out a wide range of dark matter candidates. If the dark matter is some type of elementary particle, it will not lens LMC stars (well, it will, but $M \to 0!$), so that is still an option.

The future of micro-lensing is very bright. Suppose for example that the lens is a binary star – then the light curve of the lensed object may be much more complicated, as the source may be lensed by each component of the binary in turn. Such events have already been detected toward the bulge. This is also a very promising way for detecting *planets* around distant stars, and planned satellite missions will undoubtedly detect planets in this way.

12.3 Strong lensing

The effect of lensing increases with the mass of the lens. So can we use galaxy clusters with masses $\sim 10^{15} M_\odot$ as lens? Yes we can, and we do get large deflections of several tens of arc seconds up to several arc minutes.

Figure 12.6 is an HST picture of a rich cluster of galaxies – you can see many massive elliptical galaxies in the cluster, as you'd expect. Because of

[3] http://wwwmacho.mcmaster.ca/

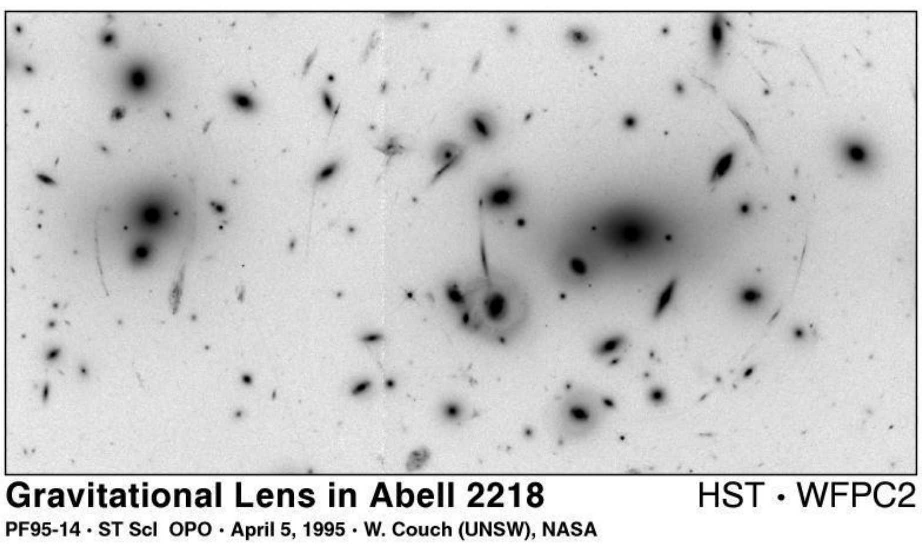

Figure 12.6: Hubble Space Telescope image of the rich cluster of galaxies Abell 2218. The large galaxies in the image are all massive elliptical galaxies in this galaxy cluster. Tens of radial arcs are images of background galaxies that have been lensed and distorted by the dark matter in the cluster.

the superb image quality of HST, you also see tens of radial arcs, centred on the cluster centre. These are images of galaxies that lie far behind the cluster, that have been lensed by the dark matter in the cluster. Clearly, their images have been distorted a lot by the lensing process.

Several of these lensed galaxies have multiple images, i.e. you see the same galaxy at various positions in the cluster. In addition, some images of galaxies may have been magnified by a lot, sometimes as much as a factor of 30. Because of this, the cluster really acts as a telescope, allowing you to study faint and distant galaxies, using the dark matter in the cluster to increase the brightness of the galaxy light. Astronomers are now deliberately targeting foreground clusters in order to study better the faint galaxies behind it, that would not be bright enough to observe in the absence of lensing.

The positions and magnifications of these lensed galaxies can also be used to infer the distribution of matter in the cluster, since it is that which causes the lensing. This is a very important application of lensing. Recall how we inferred the presence of dark matter in clusters so far. Assuming virial equilibrium, we estimated the mass in the cluster from the velocities of the galaxies. Another way was to assume that the hot gas in the cluster was in hydrostatic equilibrium, and obtain the mass from the X-ray observations of the hot gas. In either case, the computed masses were far higher than what we could account for in either stars or gas – hence our claim that galaxy clusters contained a lot of dark matter.

But we had to assume something – virial equilibrium for the galaxies, hydrostatic equilibrium for the gas. The great thing about gravitational lensing is that all we care about is the amount of mass present: we do not need to assume any type of equilibrium, since the velocity of the mass is irrelevant. The lensing mass and X-ray masses are in reasonably good agreement for the still relatively few clusters studied so far, and so this is again very convincing evidence for the presence of dark matter.

12.4 Weak lensing

The arc lets seen in clusters arise because the image of the background galaxy is strongly deformed. Equation (12.3) shows that as we look at galaxies further from the projected centre of the cluster – and hence as the impact parameter b increases – the distortion will become smaller and smaller. Eventually, we won't be able to detect the deformation for a single galaxy any more, but we could look at many galaxies in a given small patch of sky, and try to determine whether perhaps they are all slightly elongated in the same direction, as would be the case if there were a lot of mass nearby. This is called *weak lensing*: you try to identify the presence of a large mass in a given direction, from the fact that galaxies nearby in projection, tend to be elongated in a same direction. Astronomers are now searching through stacks of images trying to find whether may be there are some really massive objects that have no galaxies associated with them. Another use of weak lensing is just try to map directly the distribution of matter in the Universe, irrespective of whether it emits light as well.

12.5 Summary

After having studied this lecture, you should be able to

- derive the Newtonian lensing equation (12.2)
- explain why alignment produces an Einstein ring, and derive its radius.
- explain how micro-lensing works and has been used to search for massive compact halo objects in the Milky Way halo
- explain how gravitational lensing has been used to estimate the mass of galaxy clusters, and hence infer that they contain dark matter.

Contents