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

2007-2008

Aim

The *Hubble Deep Field* shows how even a tiny patch of sky contains thousands of galaxies of different sizes and shapes, lighting-up the Universe with a dazzling display of colours. When did they form? Which physical processes shaped them? How do they evolve? Which types are there, and whence the huge variety? How does the Milky Way galaxy fit in?

12 lectures are too few to answer any of these fundamental questions in any detail. In addition, galaxy formation and evolution require a cosmological setting, see the later lecture courses on cosmology (L3) and Galaxy Formation (L4). The more modest aim of these lectures is to give an overview of the properties of galaxies (galaxy types, properties of spiral galaxies in general and the Milky Way galaxy in particular, properties of elliptical galaxies), look at some of the more spectacular phenomena (quasars, gravitational lensing), and investigate the extent to which all this can be described with simple physics. These lectures will *use* the physics you've learnt in other lectures (in particular classical mechanics, electricity and magnetism, quantum mechanics, and of course the earlier parts Observational Techniques and Stars) to galaxies, rather than teach new physics.

The beauty of galaxies is for most people enough to warrant their study, but of course astronomy and cosmology enable one to study physics in situations which cannot be realised in a laboratory. Examples of fundamental physics made possible by astronomy and cosmology are tests of general relativity (Mercury's orbit, dynamics of pulsars), investigations whether the fine-structure constant evolves (from quasar absorption spectra), the discovery of dark matter (from the motions of galaxies), and more recently the discovery of dark energy (from the expansion of the Universe). Several of these fundamental discoveries rely on a basic understanding of galaxies, the building blocks of the Universe.

Learning outcomes:

- ability to classify galaxies and state defining properties of spiral, elliptical and irregular galaxies
- explain observational basis for discovery of the Milky Way
- explain observational basis of the modern view of the Milky Way
- apply classical mechanics to dynamics of galactic disks, including arguments for the presence of dark matter
- apply classical mechanics and other basic physics to the properties of the interstellar medium in spirals (properties of gas and dust, 21-cm radiation, concept and application of Jeans mass)
- apply mechanics and hydrostatics to describe elliptical galaxies and clusters of galaxies
- understanding of galaxy scaling relations and statistical properties of galaxies
- observational manifestations of active galactic nuclei, and the connection to central supermassive black holes
- basics of gravitational lensing, and its applications

Additional information
There are many excellent sites with images of galaxies, and explanations of some of the issues discussed below. The latest version of *Google Earth* has the option to look at the night sky, with images of galaxies taken from the *Sloan Digital Sky Survey* and the *Hubble Space Telescope*.

Most of the images shown during the lectures come from one of

- 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/

Additional explanations can be found in

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

(These are referred to as CO, ZG, BM, and BT in these notes) The first two are general and basic texts on astronomy, which, together with the web pages, should be your first source for more information. [3] is a much more advanced text on galaxies, and [4] is a superb and detailed discussion of the dynamics of galaxies.
http://burro.astr.cwru.edu/Academics/Astr222/index.html

is a truly excellent web site where you'll find additional material on what is discussed in these lectures, and several images in these notes originate from there.

Summary

Chapters 1-6: spiral galaxies

Galaxies come in a range of colours and sizes. In **spiral** galaxies such as the Milky Way, most stars are in a thin disk. The disk of the Milky Way has a radius of about 15 kpc, and contains $\sim 6 \times 10^{10} M_\odot$ of stars. The disk stars rotate around the centre of the galaxy on nearly circular orbits, with rotation speed $v_c \sim 220$ km s^{-1} almost independent of radius r (unlike the motion of planets in the solar system where $v_c \propto 1/r^{1/2}$). Such *flat rotation curves* require a very extended mass distribution, much more extended than the *observed* light distribution, indicating the presence of large amounts of invisible *dark matter*. Curiously, the spiral arms do not rotate with the same speed as the stars, hence stars may move in and out of spiral arms.

Stars are born in the large molecular gas clouds that are concentrated in spiral arms. Massive short-lived stars light-up their natal gas with their ionising radiation, and such 'HII' regions of ionised hydrogen follow the arms as beads on a string. The blue light from these massive and hot stars affects the colour of the whole galaxy, one reason why spirals are blue. Remember that the luminosity L of a star is a strong function of its mass M, with approximately $L \propto M^3$ on the main sequence, hence a single $100 M_\odot$ star outshines nearly 10^6 solar-mass stars: the colour of a galaxy may be affected by how many massive (hence young) stars it harbours.

The stars in the central region of the Milky Way (and most spirals) are in a nearly spherical *bulge*, a stellar system separate from the spiral disk, with mass $\approx 10^{10} M_\odot$. Dust in the interstellar medium of the disk prevents us from seeing the Milky Way's bulge on the night sky. Disk and bulge are embedded in a much more extended yet very low-density nearly spherical *halo*, consisting mostly of dark matter. The stellar halo contains a sprinkling of 'halo' stars, but also globular clusters, very dense groups of $10^5 - 10^6$

stars of which the Milky Way has ~ 150. The total mass of the Milky Way is $\sim 10^{12} M_\odot$, but stars and gas make-up only $\sim 7 \times 10^{10} M_\odot$ with the rest in dark matter.

Chapters 7-8: elliptical galaxies

Elliptical galaxies are roundish objects which appear yellow. The redder colour as compared to spiral galaxies is due to two reasons: firstly the stars in elliptical galaxies tend to have higher metallicity than spiral galaxies of comparable mass, and the metals in the stellar atmospheres redden the stars. Secondly, elliptical galaxies have no current star formation, and hence they do not contain the massive and short-lived stars that produce the blue light in spirals.

The stars move at high velocities, and it is these random motions, not rotation, that provides the support against gravity. This is similar to how the velocities of atoms in a gas generate a pressure, $p \sim \rho v^2 \sim \rho T$. However, unlike the atoms in a gas, the stellar velocities are in general *no*t isotropic, meaning that also the pressure is not isotropic. The result is that, whereas stars are round, elliptical galaxies are in general flattened, *i.e.* elliptical in shape.

The observed stellar motions, but also the hot X-ray emitting gas detected in them, both suggest the presence of dark matter in elliptical galaxies.

Chapters 9-10: galaxy statistics

Roughly half of all stars in the Universe are in spirals, and half are in ellipticals. Spiral galaxies tend to live in region of low galaxy density. The Milky Way is part of a 'local group' of galaxies, containing in addition to Andromeda a swarm of smaller galaxies. Ellipticals in contrast tend to cluster in groups of tens to hundreds of galaxies, called clusters of galaxies. Such clusters contain large amounts of hot, X-ray emitting gas. Together with the large observed speeds of the galaxies, this hot gas provides evidence that a large fraction of the cluster's mass is invisible dark matter.

Chapter 11: Active Galactic Nuclei

A small fraction of galaxies contains an active nucleus, which may generate as much or even more energy than all the galaxy's stars combined. The observational manifestations of such active galactic nuclei, or AGN, are quite diverse, with energy being emitted from visible light in quasars, to immensely power full radio lobes, to X-rays and even gamma-rays. Some fraction of energy may even be released in the form of a powerful jet.

The energy source is thought to be accretion onto a supermassive black hole, with masses from 10^6 to as high as $10^9 M_\odot$. It is thought that most, if not all, massive galaxies contain such a supermassive black hole in their centres. The evidence is particularly convincing for the presence of a $\sim 10^6 M_\odot$ black hole in the centre of the Milky Way.

Clearly, if most galaxies contain the engine but have little or no observational AGN manifestations, it must imply that the majority of black holes is dormant.

Chapter 12: Gravitational lensing

A gravitational field bends light. In general this is only a small effect and our view of the distant universe is not significantly deformed by the gravitational lensing caused by the intervening potentials generated by stars, galaxies or clusters of galaxies. The phenomenon has been used to test Einstein's theory of relativity, to constrain the nature of dark matter in the Milky Way, to determine the masses of galaxies and clusters, and even to enable astronomers to use clusters of galaxies as truly giant telescopes enabling the detailed study of distant galaxies.

Chapter 1

Introduction

1.1 Historical perspective

Galaxies are extended on the night sky as can be seen even without using a telescope[1]. *Andromeda* is similar in size to the Milky Way and extends over several degrees, whereas the *Large Magellanic Cloud* or LMC, visible from the Southern hemisphere on a clear night, extends over six degrees (diameter of full moon is 1/2 a degree). The Sagittarius dwarf galaxy, a small galaxy gravitationally bound to the Milky Way, extends over a large fraction of the sky.

The Milky Way galaxy is a spiral galaxy, in which most of the stars lie in a thin ($\sim 1/2$kpc) disk of radius ~ 15 kpc, with the Sun at a distance ~ 8kpc from the centre. The Milky Way's disk can be seen as a faint trail of light on the night sky, while the other visible stars are not obviously part of a disk simply because they are nearby.

Other extra-solar objects which appear extended on the night sky include planetary nebulae, supernova remnants, and star clusters. Before intergalactic distances were first measured in the 1920s it was not realised that galaxies were a separate class, and both Messier's catalogue (1780, in which Andromeda is M 31) and Dryers (1988) *New General Catalogue* (in which it is NGC 224) make no distinction between galaxies and these other 'nebulae'. Emmanuel Kant was one of the first to suggest that galaxies were other Milky Ways. The study of galaxies therefore only started in the 20th century,

[1]The disks of stars on the sky can only be resolved with special techniques, and even then only for very nearby stars.

and we now know that the observable Universe contains $\sim 10^9$ galaxies more massive than the Milky Way. Surveys of galaxies now regularly catalogue properties of $10^4 - 10^6$ galaxies.

1.2 Galaxy classification

The light from normal galaxies is produced predominantly by stars. Luminosities of stars vary widely: recall from the earlier lectures that along the main sequence luminosity L depends on mass M as $L \propto M^\alpha$, with $\alpha \approx 3$, hence a single $100M_\odot$ star outshines nearly 10^6 solar-mass stars. For given M, L depends strongly on the phase of stellar evolution, with giant and asymptotic giant branch stars *much* more luminous than the earlier main sequence phase. Therefore the observed luminosity of a galaxy will be dominated by its giant stars.

1.2.1 Observables

- *Luminosity* L Total luminosity of all stars combined.
- *Spectrum* The spectrum of a galaxy is the combined spectrum of all of its stars, weighted by luminosity.
- *Colour* A measure of the fraction of light emitted in long versus short wavelengths. *True colour* images of galaxies[2] are made by combining photographs of the galaxy taken through standard broad-band filters (see observational techniques), and mapping these to RGB colours. Since massive stars are young, hot and blue, galaxies which contain massive stars will tend to be bluer than those without them.
- *Extent* The angle the galaxy extends on the sky.
- *Flux* Just as for stars, the flux F of a galaxy with luminosity L at distance d is $F = L/4\pi d^2$, and is usually expressed using a system of magnitudes. Whereas L is an intrinsic property of the galaxy, the flux also depends on the distance to the observer.

[2] Sometimes narrow-band filters are used to stress the presence of particular emission lines, for example the Hα emission line produced in star-forming regions.

- *Surface brightness and intensity* Approximate a galaxy as a slab of stars, with surface density σ (in stars per unit area), each of identical luminosity L. The total luminosity $d\mathcal{L}$ of an area dS of this galaxy is $d\mathcal{L} = \sigma L\, dS$. The *intensity* I is the luminosity per unit area, therefore $I = d\mathcal{L}/dS = \sigma L$ for a slab. It is an intrinsic quantity, *i.e.* it does not depend on the distance to the observer. The flux dF an observer at distance d receives from this surface area is

$$\begin{aligned} dF &= \frac{d\mathcal{L}}{4\pi d^2} \\ &= \frac{\sigma L dS}{4\pi d^2} \\ &= \frac{I}{4\pi}\, d\Omega\,. \end{aligned} \tag{1.1}$$

Here, $d\Omega = dS/d^2$ is the solid angle the surface area dS extends on the sky. The quantity $dF/d\Omega = I/4\pi$, the flux received per unit solid angle, is called the *surface brightness*,

$$\text{Surface brightness} = \frac{dF}{d\Omega} = \frac{I}{4\pi}\,. \tag{1.2}$$

Note that it is independent of distance[3]. Since σ decreases with radius r from the centre, surface brightness is higher in the centre than in the outskirts. Surface brightness is usually expressed in magnitudes per square arc seconds but this is not correct. What is meant is that one has converted the flux dF, measured in a solid angle of 1 square arc seconds, into magnitudes.

1.2.2 Galaxy types

Images of galaxies immediately show there are two types, called *elliptical* (E, also called early type, or spheroidal) and *spiral* S, also called disk type, or late type. Large galaxies that do not fit into either category have usually undergone a recent violent collision, and most small galaxies are of type *irregular*.

Defining characteristics for E and S galaxies are

[3]This is only true for nearby galaxies with redshift $z \ll 1$. When z is not $\ll 1$, surface brightness dims with redshift $\propto 1/(1+z)^4$.

	Elliptical	Spiral
Shape	________	________
Colour	________	________
Stars	________	________
ISM	________	________
Dynamics	________	________
Environment	________	________

Ellipticals Isophotes (lines of constant surface brightness) of Es are smooth and elliptical and are further classified as En, where $n = 10(1 - b/a)$, where a and b are the major and minor axis of the isophotes, respectively. A round elliptical ($a = b$) is E0, whereas the most flattened ellipticals, E7, have $b = 0.3a$. There is little evidence for rotation in elliptical galaxies (except may be small ellipticals), so their flattening is not due to angular momentum. The intrinsic (as opposed to projected) shapes of Es are thought to be triaxial, with iso-density contours $a > b > c$.

Spirals The very thin flattened disk of spirals galaxies is due to rotation, and the stars in the disk are on nearly circular orbits around the galaxy's centre. Gas collects in Giant Molecular Clouds in the spiral arms of disks, where some fraction collapses into new stars. The massive, newly formed stars ionise their natal gas, and such HII regions of ionised hydrogen follow the arms as beads on a string. Once the cloud is dispersed, the blue light of these stars contributes to making the whole spiral appear blue, in contrast to the yellow/red light emitted by the older stars in Es. Spirals also contain dust which causes dark bands across the disk as the dust obscures background stars. The presence of dust prevents us from seeing the Milky Way's central bulge in visible light. The *bulge* is the central spheroidal stellar system, with many properties in common with elliptical galaxies. Spirals are further divides as Sa to Sc, where along the sequence the ratio of bulge-to-disk luminosity decreases, and the spiral arms become more loosely wound.

Some spirals also contain a *bar*, an almost rectangular stellar system in the disk, sticking-out of the bulge. These are designated as SB. So for example, an SBc galaxy is a barred (B) spiral (S), with loosely wound spiral arms and a small bulge (c), where an Sa has no bar, and a big bulge. The Milky Way is between types SBb and SBc, Andromeda is type Sb.

Hubble's classification

Hubble used the above classification ordering galaxies in a tuning fork diagram called the Hubble Sequence (Fig. 1.1). The commonly used nomenclature of early types (for Es) and late types (for Ss) comes from the mistaken belief that Es evolve in Ss.

The range in physical scales of Es is huge, from masses as little[4] 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.

cDs and S0s are unusual types of Es and Ss, respectively. cDs are very large ellipticals, found in the centres of clusters, with a faint but very large outer halo of stars[5]. S0s (or SB0 when barred) or lenticulars are the divide in Hubble's sequence between E and S, they have disks without gas or dust, and no recent star formation.

Most small galaxies like *e.g.* one of our nearest neighbours, the Small Magellanic Cloud or SMC, have no well defined disk, nor a spheroidal distribution of stars, and hence are neither of type E nor S, they are classified as type *Irregular*. The two main types of galaxies are then elliptical and spiral, with most small galaxies being irregulars.

1.3 Summary

After having studied this lecture, you should be able to

- Describe the Hubble Sequence
- Describe the main galaxy types, Es and Ss, and list five defining characteristics
- 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

[4] It has been suggested that globular clusters are small ellipticals and should represent the low-mass end.

[5] cD refers to properties of the spectrum, but think of it as standing for **c**entral **D**ominant.

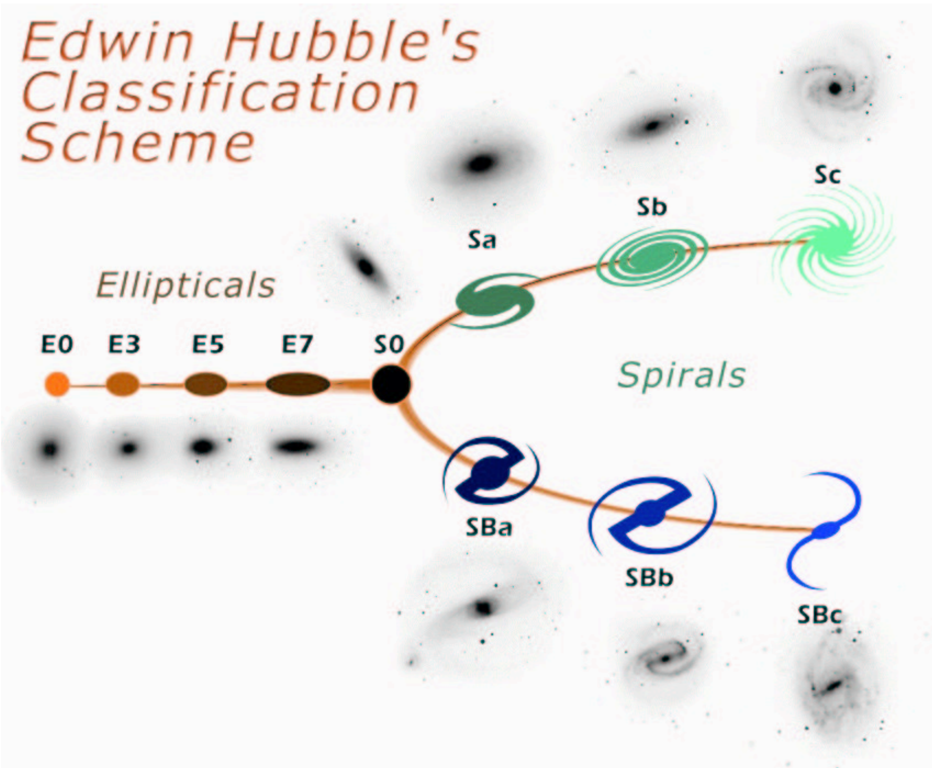

Figure 1.1: The Hubble sequence.

Chapter 2

The discovery of the Milky Way and of other galaxies

Before distances to galaxies were measured, their nature and that of the Milky Way itself, was unclear. One possibility was that nebulae where stages of stellar evolution, such as *e.g.* proto-planetary systems. Observations gave contradictory results, with star counts suggesting the Sun was near the centre of the Milky Way, while globular clusters suggested it was not. The reason for the confusion is interesting and discussed below.

2.1 The main observables

Star clusters

- *Globular clusters* (GCs) are small ($\sim$1pc) dense concentrations of $10^5 - 10^6$ stars of low-mass, old stars. Most galaxies contain 10s to 1000s of GCs, spherically distributed around their centres; the MW galaxy has about 150 GCs.

- *Open clusters* contain typically 10^4 stars, are much less dense than GCs, and are restricted to the galactic plane. They are gravitationally bound systems left-over after star formation has ceased in a Giant Molecular Cloud, and the remaining gas has been dispersed. They are of crucial importance for testing theories of stellar evolution: since their stars are thought to be coeval, any differences in properties should be a consequence of differences in initial mass.

Star counts

Star counts can probe the number density of stars in a stellar system, as a function of position. William Herschel and his sister counted the numbers of stars as function of their magnitude in many (700!!) directions in the sky. Kapteyn and his collaborators used photographic plates for counting stars.

Suppose for simplicity that all stars have the same luminosity, L. The flux of a star, $F = L/4\pi r^2$, then only depends on distance r, and $dF/F = -2dr/r$. If the number density of stars at distance r in direction $\hat\Omega$ is $n(r, \hat\Omega)$, the number of stars with flux between F and $F + dF$ in that direction, is

$$\begin{aligned} dN(F, \hat\Omega) &= n(r, \hat\Omega) r^2 dr\, d\Omega \\ &= \underline{\hspace{10cm}} dF\, d\Omega\,. \end{aligned} \quad (2.1)$$

Therefore counting $dN(F)$ allows one to infer the density $n(r)$.

Parallax

Nearby stars appear to move with respect to more distant stars as Earth moves around the Sun. The extent of the excursion θ depends on the Earth-Sun distance and the distance to the star. Measuring θ gives a *parallax* distance to the star. A star at 1pc distance has by definition a yearly parallax of 2 arcsec. The name derives from **par**allax of one **sec**ond of arc.

Standard candles

Standard candles are objects with a known absolute property, for example a known size, or known luminosity. If the intrinsic size D is know, the distance can be determined by measuring angular extent. If the luminosity is known, distance can be determined by measuring flux.

Cepheid variables are an important example. Henrietta Leavitt studied variable stars in the Magellanic Cloud in 1912. She found a relation between the period P of the periodic 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. Calibrating this relation makes Cepheids standard candles. In addition, Cepheids have a very characteristic saw-tooth shaped $m(t)$, making it easy to recognise them. Leavitt's discovery was therefore of great importance.

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 on to ever greater distances. *Hubble Space Telescope* recently measured Cepheids in the Virgo cluster (at a distance of $\sim$ 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 Discovery of the structure of the Milky Way

Star counts

Jacobus Kapteyn confirmed in the beginning of the 20th century using photographic plates the star count results first obtained by Herschel. 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. Although the star counts are correct, the interpretation to the shape of the MW and the position of the Sun are wrong.

The conversion from F to r, for given L, assumes $F \propto 1/r^2$, which neglects possible absorption of the light on route. But how can one test whether absorption is important? A plausible source of absorption is light scattering off atoms or molecules along the line of sight (Rayleigh scattering). The strength of scattering is colour dependent (stronger for shorter wavelengths): scattering of Sun light off molecules in the atmosphere makes the sky appear blue. Therefore if this were important, distant (hence fainter) stars should appear redder than more nearby stars. Kapteyn did measure the predicted reddening, but the amount was too small to be important.

However the reddening is not due to scattering off atoms, but due to scattering off *dust* (Tyndall scattering). The colour dependence of dust scattering is much less, so the small amount of reddening actually implies that absorption is quite important.

Globular clusters

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 implied a much bigger MW than Kapteyn's, and relegated the Sun to the MW outskirts. But who was right?

Other nebulae

Another famous Dutch astronomer, *van Maanen*, observed galaxies over several years, and decided he could see them move with respect to the stars. This must mean they are relatively nearby, certainly within the MW.

But another astronomer Slipher measured large (1000s km s^{-1}) velocities for some Nebulae, and finds evidence for rotation. He also claims the light is produced by stars, not by gas. This suggests the nebulae are galaxies, and outside of the MW, in conflict with van Maanen'sand Kapteyn's MW picture.

Hubble's discovery

Hubble used the 100-inch Hooker Telescope on Mount Wilson in 1919 to identify Cepheid variables in other nebulae, including Andromeda. He inferred[1] a distance of $\sim$ 0.3 Mpc, much larger than even Shapleys size for the MW, and final prove that the nebulae were other galaxies outside of the Milky Way. Hubble went on to discover that the more distant galaxies move away from us at ever increasing speed. It is difficult to underestimate the importance of this discovery. In one great swoop, the size of the Universe increased dramatically. The Sun got relegated to the outskirts of the MW, and the MW itself was found to be just one out of billions of other galaxies. And the Universe was found to be expanding, making Einstein's attempts to build a static cosmological model out of his theory of relatively irrelevant.

Epilogue

Trumpler found in the 1930s what had been wrong with Kapteyn's interpretation, by studying MW Open Clusters. He assumed the size of Open

[1]The modern value is $\sim$ 0.7Mpc.

Clusters to be a standard candle, and hence assumed you could infer their distance from their angular size. Using this distance indicator, he found that the stars in the more distant clusters (as inferred from cluster size) were invariable much fainter than stars in more nearby clusters. He concluded that the light from distant stars was attenuated much more than expected from scattering off atoms. It had to be absorption by dust, and Kapteyn's neglect of this lead to his error in interpretation.

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 is 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.

2.3 Absorption and reddening

Consider a light ray of intensity I traversing a space containing very large dust grains (bricks). The bricks will absorb a fraction of the incoming light, and the intensity of the ray will decrease as

$$\frac{dI}{dr} = -A\,I\,, \tag{2.2}$$

which expresses the fact that each distance dr will absorb a constant *fraction* $dI/I = -A\,dr$ of the light. The constant A depends on the number density of bricks, and their size.

The solution to this equation is

$$I(r) = \underline{\hspace{8cm}} \tag{2.3}$$

In terms of magnitudes, $\Delta m = -2.5\log(I/I_0) = \hat{A}\,r$, where the relation between $\hat{A}$ and A is left to you. $\hat{A}$ therefore has units of magnitude per unit length. Absorption therefore changes the usual relation $(m-M) = 5\log(r) - 5$ between apparent and absolute magnitude to $(m-M) = 5\log(r) - 5 + \hat{A}\,r$.

If the size of the particles is of order of the wavelength λ of the light, then the value of A will be wavelength dependent. This leads to reddening of the star, since smaller wavelengths (blue light) will be absorbed more strongly than longer wavelengths (red light).

If we apply this reasoning to light in the B versus V band, for example, we obtain

$$\begin{aligned}
(m-M)_{\rm B} &= 5\log(r) - 5 + A_{\rm B}\,r \\
(m-M)_{\rm V} &= 5\log(r) - 5 + A_{\rm V}\,r \\
E_{\rm B-V} &\equiv (m_{\rm B} - m_{\rm V}) - (M_{\rm B} - M_{\rm V}) = (A_{\rm B} - A_{\rm V})\,r\,.
\end{aligned} \tag{2.4}$$

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,

$$E_{\rm B-V} = ({\rm B-V})_{\rm observed} - ({\rm B-V})_{\rm intrinsic}\,. \tag{2.5}$$

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

$$E_{\rm B-V} \approx \frac{1}{3} A_{\rm V}\,r\,. \tag{2.6}$$

This is a crucial result. Reddening and hence $E_{\rm B-V}$, is easy to measure, and so if we do this for stars of known distance, we find[2] $A_{\rm V} \approx 1{\rm mag\,kpc^{-1}}$. If we now measure $E_{\rm B-V}$ for another star of *known* colour (from stellar evolutionary models say) we can estimate r.

[2]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 what are Open and Globular clusters
- Explain and apply the relation between star counts and the density of stars
- Explain what a standard candle is.
- Explain why Cepheids are good standard candles.
- Explain and apply parallax measurements
- 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 and apply the effect of scattering by dust on the apparent magnitude and colour of distant stars.

Chapter 3

The modern view of the Milky Way

Our current picture of the Milky Way is based on a observations using a range of different techniques, often developed for military purposes (*e.g.* radar).

3.1 New technologies

3.1.1 Radio-astronomy

Radio-waves are EM-radiation with long wavelength λ and therefore are not susceptible to absorption by interstellar dust. Therefore they can probe the dense, dusty regions where stars are formed. Earth's atmosphere is nearly transparent for 1 cm $< \lambda <$ 10 m. Radio telescopes usually consist of a (large) single dish (for example at Jodrell bank; the Phjysics department has a small dish on its roof) or an *interferometer* with many inter-connected dishes, for example the *Giant Meter Radio-telescope* in India. Radio-observatories separated by 100-1000s of kilometers sometimes combine their signal to obtain very high angular resolution[1] Radio-waves are usually produced by one of:

1. Roto-vibrational transitions in molecules

2. Thermal radiation from (cold) dust

[1]Recall that the angular diffraction limit depends on telescope diameter D and observed wavelength λ as $\theta \propto \lambda/D$. A radio-interferometer with $D = 10^3$ km and $\lambda = 1$ mm has $\theta \approx 10^{-9}$, whereas an optical telescope with $D \sim 10$ m and $\lambda \sim 5{\times}10^{-5}$ m has $\theta \approx 5{\times}10^{-8}$.

3. Synchrotron radiation from electrons moving in a magnetic field
4. Hyperfine transitions, *e.g.* the 21-cm line in hydrogen (see Ch 4)

Roto-vibrational transitions correspond to the rotational or vibrational transitions in molecules. For example energy can be stored in the vibration of the C and O atoms in a CO molecule, whereby the distance between C and O varies. A quantum transition whereby the amount of vibration in the molecule decreases is associated with the emission of a photon. A diatomic molecule such as CO can also store energy in rotation along one of the two axes perpendicular to the C-O molecular bond. A decrease in the amount of rotation again results in the emission of a photon. The associated energies ΔE are in general much less than of *electronic transitions* hence the associated wavelengths $\lambda = hc/\Delta E$ are longer (IR or radio-waves, as opposed to optical or UV-radiation).

Thermal radiation Dust heated by nearby stars cools by radiating radio/infrared waves. If the temperature are low, the blackbody may peak in micro/mm wavelengths[2]

Synchrotron radiation Electrons moving in a magnetic field may emit radiation at radio-wavelengths depending on their speed (just as in a terrestrial synchrotron). Astronomical examples include supernovae remnants.

3.1.2 Infrared astronomy

IR photons are not significantly 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 IR observations can look deep inside star forming regions. The DIRBE instrument on the COBE (Cosmic Background Explorer) satellite provided us with one of the best views of the the Milky Way, because it could see through the dust clouds that obscure large parts of the MW in the optical. Unfortunately earth's atmosphere absorbs most IR radiation, except in some narrow bands. IR observations therefore require balloon, rocket or

[2] Recall Wien's displacement law, relating the peak emission wavelength λ and temperature T as $\lambda =$________________.

satellites, or are limited to narrow regions of EM radiation in between the atmosphere's absorption bands.

3.1.3 Star counts

The Hipparcos satellite used diffraction limited observations above the atmosphere to obtain superbly accurate positions of stars on the night sky. By repeating measurements over several years, the satellite measured parallaxes of many 1000s of stars, and in addition proper motions for some stars (*i.e.* the velocity of some stars in the plane of the sky). Parallax distances are the crucial first step to the distance-ladder: parallax distances to Cepheids are required to calibrate the period-luminsity relation.

GAIA[3], the successor to HIPPARCOS, is due for launch in ~ 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 modest 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. It will revolutionize the study of the MW.

3.2 The components of the Milky Way

The techniques above have shaped our view of the MW: it consist of three well-defined separate stellar systems: *disk*, *bulge* and *halo*. A summary of the properties of these is in Tables 3.1 and 3.2.

3.2.1 Disk

The disk is a round, thin distribution of stars. The Sun is part of the MWs disk.

- luminosity $L \sim 2 \times 10^{10} L_\odot$ in the B-band, ≈ 70 per cent of the MWs total B-band luminosity.

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

- radius[4] $R \sim 15$kpc.
- Define the thickness t of the disk as the ratio $t \equiv \rho/\sigma$ of the volume density of stars, ρ, at the galactic plane, and the surface density σ. The thickness t depends on the type of star, and is ~ 200pc for young stars, ~ 700pc for stars like the Sun.

The disk also contains gas and dust (see lecture 4 on the interstellar medium). 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, with speeds ~ 220km s^{-1}. The oldest disk White Dwarfs are $\sim 10 - 12 \times 10^9$ years old, but these could have formed *before* the disk.

The density distribution of stars in disk in cylindrical coordinates (R, ϕ, z) can be written as

$$n(R, \phi, z) \propto \exp\left(\frac{-R}{R_h}\right) \exp\left(\frac{-|z|}{z_h}\right), \tag{3.1}$$

i.e. independent of ϕ since it is cylindrically symmetric, and falling exponentially both in radius R and height z above the disk, with *scale-length* $R_h \approx 3.5$ kpc and *scale-height* $z_h \approx 0.3$ kpc.

Note that there isn't really an edge to the disk, either in radius or height. 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!

The thick disk About 4 per cent of the MW's stars belong to a thicker disk, aligned with the (thin) disk, but with a larger scale height of $z_h \approx 1$kpc. Stars in this thick disk differ from the thin disk stars discussed above both in composition (having a lower metal content) and kinematically.

3.2.2 Bulge

The bulge is a spheroidal distribution of stars in the centre of the MW (and of most spirals).

- central spheroidal stellar system with radius of ~ 1kpc

[4]The distance from Sun to the galactic centre has recently been determined with remarkable accuracy to be $R_0 = 7.94 \pm 0.42$ kpc, from the observed motions of stars around the galactic centre (see lecture 11 for more details).

- luminosity ≈ 30 per cent of total MW luminosity
- luminosity profile is a de Vaucouleurs or '$r^{1/4}$' profile, defined as

$$I(r) = I_e \exp[-7.67(r/r_e)^{1/4} - 1]\,, \tag{3.2}$$

with *effective radius* $r_e \approx 0.7$kpc.

The bulge stars are generally older and more metal poor than disk stars, suggesting the bulge formed before the disk, and before stellar evolution had managed to pollute the gas from which stars form with metals. In contrast to the disk, there is little or no star formation in the bulge.

The MW's bulge is elongated with axis ratio 5:3, with strong evidence for a bar. Whereas the disk stars are rotating in ordered fashion around the MW centre, the bulge has little net rotation, but the stars have large random velocities. All these properties: no star formation, large random stellar motions, spheroidal geometry, are reminiscent of the properties of an elliptical galaxy: it is as if there is a small elliptical galaxy at the centre of each spiral.

The bulge is bright but obscured in the visible due to dust: we need IR observations to make it visible.

3.2.3 Halo

A very small fraction (≤ 1 per cent) of the MWs stars are contained in a large, spheroidal, extremely tenuous stellar system called the halo, mostly (99%) made-up out of single 'field' stars with a sprinkling of Globular Clusters. The halo does not appear to rotate, and the halo stars have very low metallicities, typically 1/10 to 1/100 times the solar metallicity. When a halo star occasionally plunges through the disk, it is spotted because of its very high velocity, which is partly due to its intrinsic high random velocity, partly due to the fact the the Sun (and other disk stars) races with 220 km s^{-1} around the MW centre, but the halo star does not partake in this rotation. The halo contains ≈ 150 Globular Clusters (GCs), although there is some evidence that younger GCs may be associated with the disk.

The density of halo stars, and of GCs, falls spherically as

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

and extremely distant field stars have been detected out to more than $\approx$ 50 kpc.

3.2.4 The dark matter halo

See lecture 5 for evidence that as much as 90 per cent of the mass in the MW is invisible, and consists of some unknown type of matter.

3.3 Metallicity of stars, and stellar populations

Nuclear-synthesis during the first three minutes after the Big Bang produced mostly Hydrogen and Helium, and trace-amounts of other elements such as Li and B[5]. By mass, a fraction $X \approx 0.76$ was Hydrogen, with most of the remaining mass fraction $Y \approx 1-X \sim 0.24$ in Helium. All other elements were synthesised in stars and flung into space either due to winds (in AGB stars), during a Planetary Nebula phase, or during supernova explosions[6]. These elements are usually (but erroneously) referred to as 'metals' in astronomy, and their mass fraction denoted as $Z \equiv 1 - X - Y$.

A star formed out of gas already enriched in metals by (a) previous generation(s) of stars, will have a higher metal fraction Z than stars formed from more pristine gas. Since most of the stellar burning converts Hydrogen into Helium, such a star will have also tend to have $X < 0.76$ and hence $Y > 0.24$. For example the Sun has $X \sim 0.7$, $Y \sim 0.28$ leaving a total fraction of $Z = 1-X-Y \approx 0.02$ in 'metals'. The composition of a star therefore contains a wealth of information on the properties of earlier generations of stars. It is quite an awesome realisation that the metals in the Sun (and also in you and me) were produced by 1000s of stars and SNe that have long since perished.

Metallicity is often expressed relative to that of the Sun on a logarithmic scale, denoted (for example for Fe) as

$$[\mathrm{Fe/H}] \equiv \underline{\hspace{6cm}}\,, \tag{3.4}$$

that is (the logarithm of) the ratio of Fe-to-H mass, divided by that ratio for the Sun. A star with [Fe/H]=0 has the same Fe abundance as the Sun, a star with [Fe/H]=-1 is ten times more Fe poor, a star with [Fe/H]=1 ten

[5]See L3 and L4 lectures.

[6]Elements heavier than ^{56}Fe are endothermic, meaning energy is required to synthesise them, as opposed to elements that release energy during synthesis, and are almost exclusively produced during SNe explosions.

Table 3.1: Disk parameters

	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	$\exp(-z/0.16\text{kpc})$	$\exp(-z/0.325\text{kpc})$	$\exp(-z/1.4\text{kpc})$
[Fe/H]	> 0.1	-0.5 – 0.3	-1.6 – -0.4
Age (Gyr)	0 – 17	< 12	14 – 17

times more Fe rich.

Disk stars have $-0.5 \leq$ [Fe/H] $\leq +0.3$, with a clear trend of increasing [Fe/H] toward the centre of the disk, hence the MWs material has been processed more vigourously in its interior than towards its outskirts. The disk population of relatively metal rich stars, still undergoing star formation is called *population I* stars. The true nature of the thick disk stars is not completely clear, but thick disk stars tend to be more metal poor, [Fe/H]≈ -0.6

Bulge stars have a wide range of $-3 \leq$ [Fe/H] ≤ 0.3, with no significant ongoing star formation. This population of relatively metal rich stars without star formation is called *population II*[7].

Halo stars have much lower abundances $-3 <$ [Fe/H]< -1. A population with such extremely low abundance, composed of old stars, is called *extreme population II.*

3.3.1 Galactic Coordinates

The position of an object on the sky as seen from the Sun can be characterised by two angles (see Fig. 3.1) : (1) *galactic latitude*: the angle b above the Milky Way plane and (2) *galactic longitude*: the angle l between the direction Sun-Galactic centre, and the projection Sun-star onto the Milky Way disk.

[7] The stellar population of elliptical galaxies is similar

Table 3.2: Spheroid parameters

	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_{B} $(M_\odot/L_\odot)$	3	~ 1	-
Diameter (kpc)	2	100	> 200
Distribution	bar	$r^{-3.5}$	$(a^2 + r^2)^{-1}$
[Fe/H]	-3 – 0.3	-4.5 – -0.5	
Age (Gyr)	10 – 17	14 – 17	17

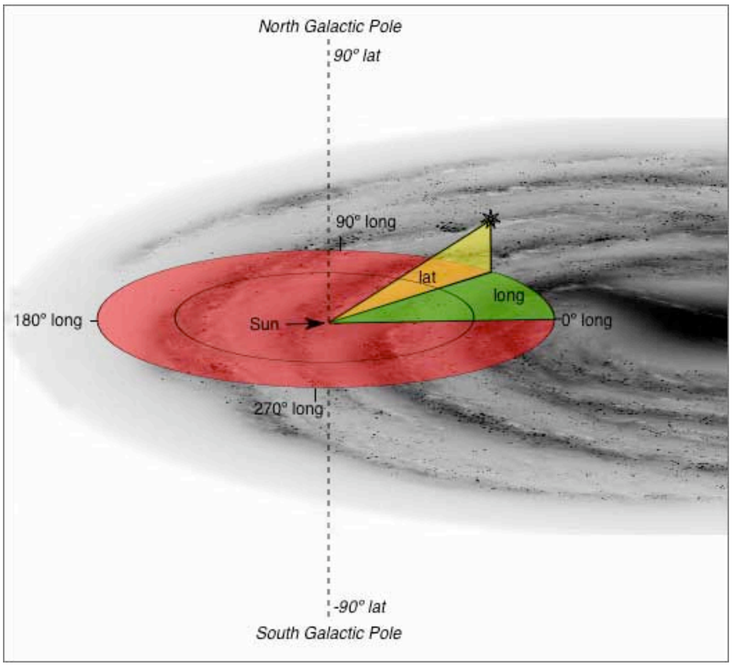

Figure 3.1: Diagram illustrating galactic longitude l and galactic latitude b. From *thinkastronomy.com*.

3.4 Summary

After having studied this lecture, you should be able to

- Explain which processes generate observable radio and IR radiation, and why this radiation was important to clarify the MW structure.
- 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 three main stellar components of the MW, and give two characteristic properties of each
- Explain what is meant by the metallicity of a star, and how it is expressed
- Explain how the Galactic Coordinate system (l, b) is defined

Chapter 4

The Interstellar Medium

The interstellar medium (ISM) is the *stuff* between the stars. It is a rich but complex physical system, with gas collecting in large molecular clouds (*Giant Molecular Clouds*), which form stars that then destroy the clouds, polluting the gas with metals. Stellar winds and super nova explosions stir the gas, any may even expell some gas from the MW into the surrounding intergalactic medium. Dust, magnetic fields and cosmic rays[1] play an important but poorly understood rôle. The L4 course on the ISM treats some of the issues in more detail. This chapter concentrates on the composition of the ISM, how the different components can be observed, how stars and gas interact, and the important concept of Jeans mass, relevant for how stars form in clouds.

4.1 Interstellar dust

Interstellar dust is composed of metals (*e.g.* Si), affects the chemistry of the ISM, and also the propagation of light. Dust particles interact with light both through *scattering* and *absorption*. Sunlight passing through the Earth's atmosphere scatters off molecules (mostly N and O molecules), with blue light scattered more strongly than red light. Away from the Sun, we

[1]Cosmic rays are energetic particles (photons, protons, nuclei of heavier elements), with energies up to orders of magnitude higher than what can presently be achieved in labs such as SLAC or Cern. With Profs Rochester and Wolfendale, Durham has a rich history in cosmic ray physics, and continues to do so with its involvement in the HESS telescope in Namibia.

Figure 4.1: Scattering of Sunlight in the Earth's atmosphere

see the scattered Sun light on the sky, which is hence much bluer than the Sun itself: this is why the sky is blue. Conversely, looking at the Sun itself, we see a larger fraction of red light: the Sun reddens at sunset or sunrise. Sunlight can also scatter on dust (for example following a volcanic eruption or a dessert storm): this makes the sun appear even redder at sunset/rise. Figure 4.1 illustrates this phenomena:

The scattered light is also *polarised*. Sun glasses are sometimes polarised to block sunlight scattered from the Ocean, or from road surfaces. Light may also be absorbed by clouds our dust, darkening the day. The same process of scattering and absorption occur in the ISM: distant stars appear redder, the light detected from reflection nebulae is blue and polarized, all due to the effects of dust grains.

Scattering: changes the direction of the incoming photon, but not its energy. EM radiation with wavelength much longer than the size of a grain scatters must less efficiently, hence red light, and in particular IR and radio-waves, is hardly affected by dust.

Absorption: a photon interacting with a dust grain may break a molecular bond, or simply heat the dust grain. This destroys the photon, which is lost to the observer. The grain may cool by emitting EM radiation, with black-body spectrum corresponding to its temperature: the typical energies of this cooling radiation is much less than those of the incoming photon. The net effect is that the dust grain has converted short wavelength (blue) photons to long wavelength (red or IR) ones. Dust in the surroundings of star forming regions for example reprocesses the blue light emitted by the hot stars into IR radiation. This obscures the star forming region in the visible, but makes it glow in the IR.

Although in general the amount of scattering/absorption increases with decreasing wavelength λ, some λs correspond to resonance transitions in the molecules or dust grains, allowing determination of their composition[2].

4.2 Interstellar gas

Gas in the ISM can be in molecular, neutral, or ionised form. What determines which phase the gas is in? How are the different phases observed? The processes discussed below are illustrated in Fig. 4.2.

Notation Astronomers use the (confusing) notation XIII to denote the *doubly* ionised state of element X, for example CIII$\equiv$$C^{2+}$, HI is neutral Hydrogen, HII ionised Hydrogen, and OVI is five-times ionised O.

4.2.1 Collisional processes

Kinetic theory of gases The typical velocity v of particles with mass m in a gas with temperature T is $v^2 \sim kT/m$. A collision between such particles will have typical kinetic energy $E \sim mv^2 \sim kT$, and will be elastic (energy conserving) when E is much less than any excitation energies of the particle. Rotational and vibrational excitation energies are small, and so even at relatively low temperatures $T \sim 100$ K, the kinetic and roto-vibrational energy levels are in equipartition ('thermal equilibrium'). Because of such 'internal degrees of freedom' a diatomic gas (*e.g.* H_2) does not behave the same as a mono-atomic gas (*e.g.* HI)[3].

[2]A typical example is the detection of the Si-O resonance in dust grains, showing they contain both Si and O.

[3]Recall that a diatomic gas has adiabatic exponent $\gamma = 7/5$ in contrast to the ideal (mono-atomic gas) case $\gamma = 5/3$.

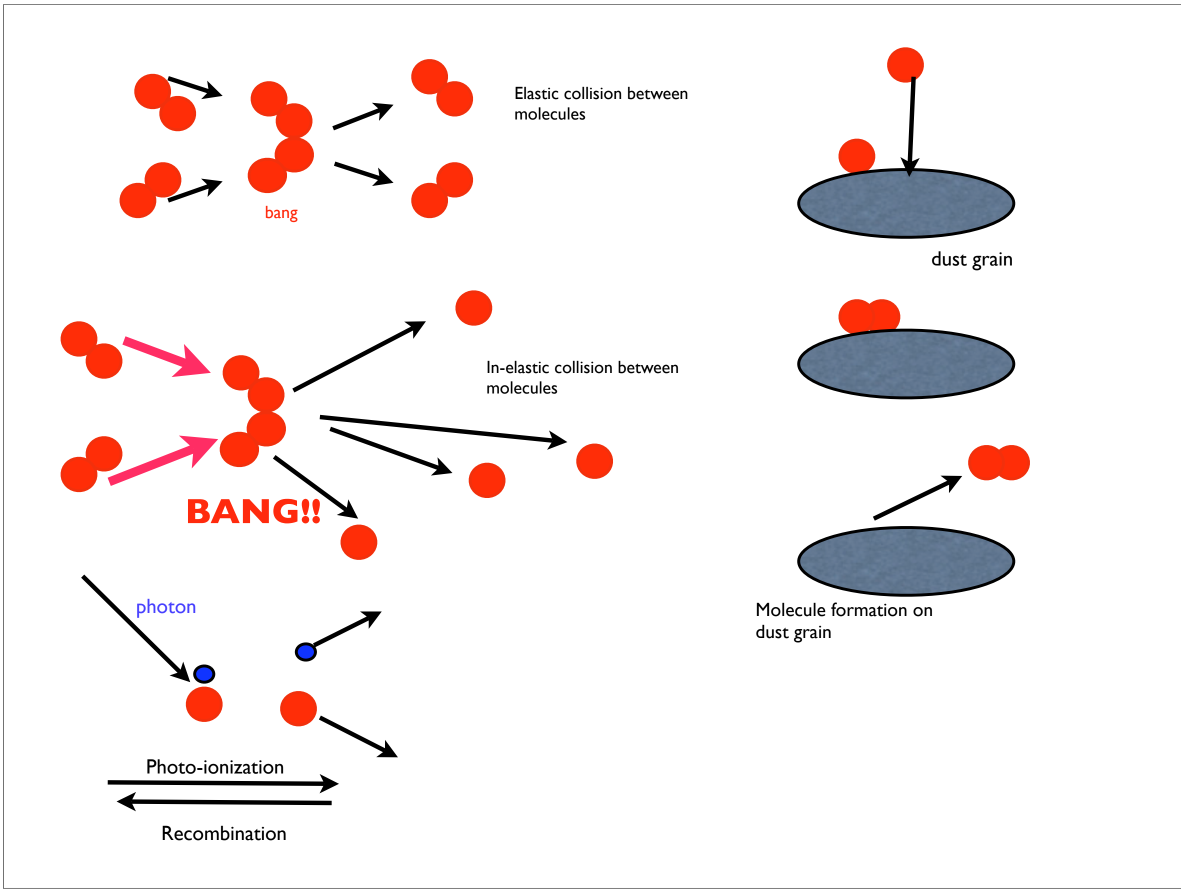

Figure 4.2: Important physical processes in the ISM.

At higher T, the kinetic energy may be high enough for collisions to break the molecular bond in the particles: collisions are no longer elastic, but destroy molecular bonds. The molecular binding energy in H_2 is ~ 11eV[4], so gas at temperatures of $T \geq 10^3$K will cease to be molecular. At even higher T, E is enough to ionise the atoms. For example the binding energy of the Hydrogen atom is 13.6eV, hence for $T \sim 10^4$K, Hydrogen will tend to be ionised. This suggests that **increasing T leads to initially molecular gas to become first neutral, then ionised.**

How about the reverse processes? Ionised gas can become neutral through recombinations. This process produces a photon which could escape from the gas, taking energy with it. This energy loss means the gas is cooling: it is called *radiative cooling*. Neutral gas can become molecular through gas-phase reactions, but in the ISM mostly because of reactions on dust grains.

The ISM has regions where the gas is cold, warm , and hot . Because we expect the *pressure* in these regions to be similar (since otherwise gas would flow from high to low pressure regions, equilibrating the pressure), the cold regions tend to be dense and molecular (called molecular clouds), the warm regions tend to be neutral, whereas the hot regions are at low density and contain ionised gas.

4.2.2 Photo-ionisation and HII regions

When a photon impinging on an atom has enough energy, it may ionise the atom in a process called *photo-ionisation*. The ionisation energy for HI is 13.6eV, and the corresponding wavelength of the photon is in the UV. The ionised Hydrogen nucleus may catch a free electron and recombine: it is the inverse process. On its way to the $n = 1$ ground-state, the electron will pass through one or more allowed electronic energy levels, emitting photons with characteristic energies 13.6eV $(1/n_1^2 - 1/n_2^2)$ for a transition $n_2 \rightarrow n_1$. The series of emission lines which end in the ground state ($n_1 = 1$) are called the *Lyman-series*, with Lyman-α corresponding to $n_2 = 2$, Lyman-β to $n_2 = 3$, *etc.* Transitions with $n_1 = 2$ are called the Balmer series, with Balmer-α or Hα corresponding to $n_2 = 3$, Hβ to $n_2 = 4$, *etc.* Fig 4.3 illustrates these

[4]An energy of 1eV corresponds to a temperature T=1eV/$k_B \approx 10^4$K.

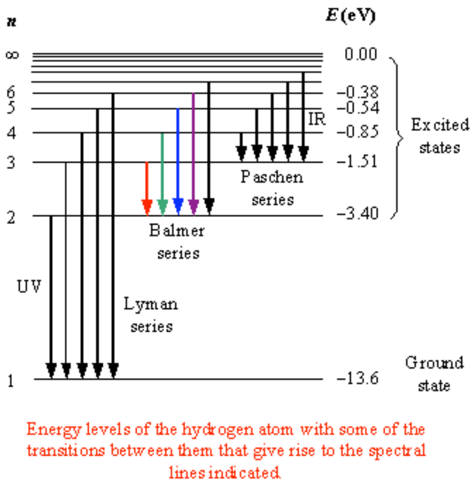

Figure 4.3: Electronic transitions in the Hydrogen atom, and nomenclature of the corresponding emission lines. Figure taken from the internet encyclopedia.

transitions.

Only massive stars are hot enough to produce Hydrogen ionising photons in abundance. Since massive stars have such short lifetimes, they are almost exclusively found in the star forming region where they were born. As the hot stars ionise the gas, most of Hydrogen gas is ionised and the region is therefore called an **HII region**. Recombinations make HII regions emit the very characteristic pattern of recombination lines discussed above. It is the Hα emission lines which is used to trace HII regions in spiral galaxies[5]. Such HII regions follow spiral arms as beads on a string: this is because the spiral structure sweeps gas into dense clouds, which can then form stars, which subsequently ionise the gas making it radiate in recombination lines and be observable. All this stellar feedback destroys the cloud, the massive stars die, and so there are no HII regions nor massive stars, in between spiral arms. The Rosetta Nebula and Orion are well known HII regions in the Milky Way.

[5]Other lines are too far into the UV to be easily observable from the ground

4.2.3 Strömgren spheres

Suppose a hot star starts at time $t = 0$ to emit ionising photons at a rate $\dot{N}$ into a gas cloud, which initially contains neutral Hydrogen gas with uniform density n. The radius R of the sphere within which the gas is ionised will increase in time. We can find $R(t)$ by requiring that the number of photons emitted by the source in time t, $\dot{N}\,t$, equals the number N of (initially) neutral Hydrogen atoms within $R(t)$, $N = (4\pi/3)\,n\,R^3(t)$, hence

$$\dot{N}\,t = \frac{4\pi}{3}\,n\,R^3(t)\,. \tag{4.1}$$

This simply expresses the fact that the source needs to emit one ionising photon per neutral atom. Take the time derivative to obtain the speed of the *ionisation front*,

$$\frac{dR}{dt} = \underline{\hspace{8cm}}\,. \tag{4.2}$$

Initially when R is small the speed of the front is high, then decreases $\propto 1/R^2$. Equation (4.2) predicts the speed becomes arbitrarily large for small R, but this cannot be right: it must be limited by the speed of light, so our treatment cannot be applied at very early times

As the ionisation front runs into the cloud, some of the Hydrogen ions behind the front will start to recombine. Each neutral atom within the HII region will absorb an ionising photon, which then can no longer ionise a neutral atom outside of the HII region: recombinations will slow the ionisation front. Eventually, there may be so many recombinations occuring within the HII region that the speed of the front decrease to zero: the front has reached its **Strömgren radius** R_s. The rate $dn_{\rm HII}/dt$ at which Hydrogen ions recombine, is given by

$$\frac{\dot{n}_{\rm HII}}{dt} = -\alpha n_e n_{\rm HII}\,. \tag{4.3}$$

Figure 4.2 illustrates the recombination of a Hydrogen ion with an electron. Since both a proton and an electron should be present for a recombination, the recombination rate is proportional to the product of the densities $n_e n_{\rm HII}$ of electrons and protons, respectively. The proportionality constant is denoted α[6]

When Hydrogen gas is highly ionised, $n_e = n_{\rm HII} \approx n$, and the Strömgren radius R_s is found by equating the rate of recombinations in a sphere of radius R, $\dot{N}_r \equiv dn_{\rm HII}/dt \times (4\pi/3)\, R^3$, with the rate at which the source produces ionising photons, $\dot{N}$:

$$\begin{aligned} \dot{N} &= \frac{dn_{\rm HII}}{dt}\frac{4\pi}{3}R_s^3 \\ R_s &= \underline{\hspace{6cm}}\,. \end{aligned} \tag{4.4}$$

Using typical values $n_{\rm H} = 5 \times 10^3 {\rm cm}^{-3}$ for the density of a cloud, $\dot{N} = 10^{49}{\rm s}^{-1}$ for the ionisation rate of a massive star, and using $\alpha \approx 3.1 \times 10^{-13}$ cm^3 s^{-1} (valid for temperatures $\sim 10^4$ K in HII regions) gives a Strömgren radius $R_S \approx 0.21$pc.

4.2.4 21-cm radiation

The Lyman and Balmer series discussed earlier corresponded to transitions of the electron between different *orbital* energy levels, but there are other possible transitions for an electron in an atom, which result from its **spin**. Spin is a purely quantum mechanical degree of freedom, with some similarities to angular momentum[7]. A spinning electric charge generates a dipole moment, and spin is also associated with a magnetic dipole. However, the electron's magnetic dipole $\mathbf{S}$ is *quantised*: along any axis, the dipole is either up or down: $\mathbf{S} = \pm 1/2$. The electron has a series of energy levels resulting from the interaction of this dipole moment with its orbital angular momentum, $\mathbf{L}$, and with the proton spin, $\mathbf{S_p}$. This is similar to the interaction between two bar magnets: when they are aligned, they will repel each other, since like-poles repel. When they are anti-alingned, they will attract each other.

fine-structure lines result from coupling between $\mathbf{S}$ and $\mathbf{L}$: for given $\mathbf{L}$, the electron's energy is slightly higher when $\mathbf{S}$ and $\mathbf{L}$ are parallel, than when they are anti-parallel.

[6]Determine the units of α as an exercise.
[7]See later lectures on quantum mechanics

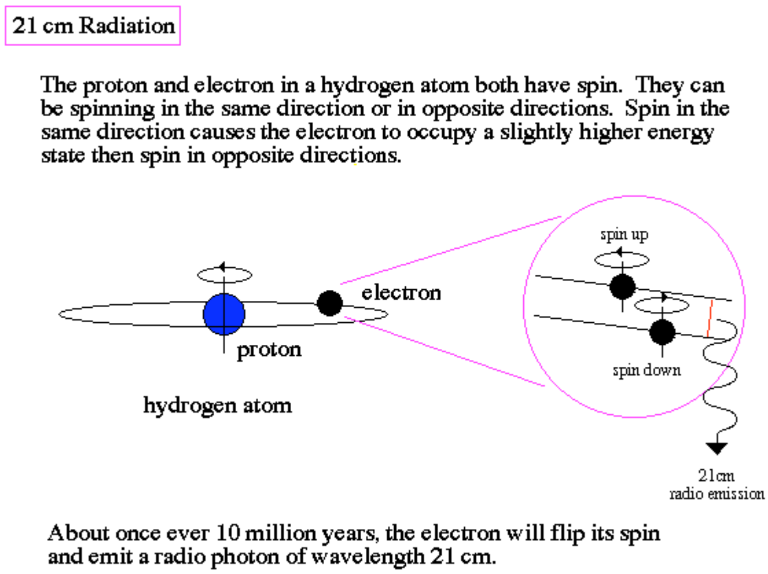

Figure 4.4: Emission of 21-cm radiation in Hydrogen due to a hyperfine transition, taken from Schombert.

hyperfine structure lines result from coupling between the electron spin $\mathbf{S}$ and the *proton* spin $\mathbf{S_p}$: the energy of the electron is higher for $\mathbf{S}$ and $\mathbf{S_p}$ parallel than when they are anti-parallel, see Fig.4.4.

The wavelength of the hyperfine transition in the Hydrogen atom is $\lambda \approx 21.1$ cm and it is therefore called the **Hydrogen 21-cm line**. The transition probability is very low, which is why it is very hard to observe it 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, and **this line is of great observational importance** because it can be used to detect neutral gas. The long wavelength also means it is not affected by dust[8].

An external magnetic field shifts the energy levels of the electron in a Hydrogen atom, the 'Zeeman effect', and the 21-cm radiation can be used to

[8] The radio-dish on top of the physics department can detect 21-cm radiation from neutral gas in the MW

estimate the strength of the magnetic field in the ISM.

4.2.5 Other radio-wavelengths

The 21-cm line cannot be used to study gas in dense clouds, because the gas will tend to be molecular instead of atomic. But radio-telescopes can be used to detect roto-vibrational transitions of these molecules, enabling the study of molecular clouds.

The MW contains molecular clouds with a wide range in masses, up to *Giant Molecular Clouds*, (GMCs) with masses up to $10^7 M_\odot$. These enormous complexes of gas and dust are almost exclusively found in spiral arms, and are the sites of star formation in the MW: most, if not all, stars are thought to form in GMCs.

4.2.6 The Jeans mass

The masses of GMCs, $\sim 10^6 M_\odot$, are very much higher than those of the stars that form in them: why?

The **Jeans** mass M_J in a gas with uniform density, ρ, and temperature, T, is the characteristic mass for which the thermal energy, K, and gravitational energy, U, in a sphere are in virial equilibrium, $2K = U$. When $M > M_J$, gravity dominates, and the sphere will tend to collapse; when $M < M_J$, pressure dominates, and the sphere is stable to collapse.

The thermal energy K in a sphere with mass M is

$$\begin{aligned} K &= Mu \\ u &= \frac{\mathrm{k_B} T}{(\gamma - 1)\,\mu \mathrm{m_H}}\,, \end{aligned} \tag{4.5}$$

where $\gamma = 5/3$ is the adiabatic exponent, $\mathrm{k_B}$ Boltzmann's constant, and $\mu\mathrm{m_H}$ the mean molecular weight per particle. The gravitational potential energy U of the sphere is

$$
\begin{aligned}
U &= \frac{3}{5}\frac{GM^2}{R} \\
M &= \frac{4\pi}{3}\rho R^3 \,.
\end{aligned}
\tag{4.6}
$$

In virial equilibrium, $2K = U$, and the mass of the sphere is the Jeans mass[9]:

$$
M_J = \left(\frac{10\mathrm{k_B}T}{3(\gamma-1)\mu \mathrm{m_H} G}\right)^{3/2}\left(\frac{3}{4\pi\rho}\right)^{1/2} \,. \tag{4.7}
$$

Fragmentation Consider the fate of a cloud with mass $M = M_J$ that starts to collapse. In general, both T and ρ will change, and hence M_J will change as well. If the gas behaves *adiabatically*, $\rho \propto T^{3/2}$, and hence $M_J \propto \rho^{1/2}$. Hence M_J *increases* at the cloud starts to collapse, pressure will increase faster than gravity, and the cloud will bounce back. However if the gas is *isothermally*, T =constant, then $M_J \propto 1/\rho^{1/2}$, and the Jeans mass will *decrease*. Now smaller spheres will have $M = M_J$ and hence the cloud may fragment.

[9]You may find expressions which differ by factors of order unity in other texts

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.
- Compute the speed of an ionisation front.
- 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

The velocity of a test mass m in circular motion around a mass M at distance R is

$$\frac{V_c^2}{R} = \frac{GM(< R)}{R^2}\,, \tag{5.1}$$

independent of m along as $m << M$. Mass M need not be a point mass: Newton's theorem guarantees this equation is also correct for an extended spherical mass distribution, as long as M is the mass enclosed[1] by the orbit R: mass outside R does not contribute to the gravitational force. This equation describes for example the motion of planets in the solar system, with M

[1]We'll use the notation $M(< R)$ for the mass enclosed in a sphere of radius R.

the mass of the Sun.

This equation also describes the circular motion of stars around the MW's centre: Newton's theorem also applies for motion in a disk[2] In the outskirts of the disk, for example at the position[3] $R_\odot$ of the Sun, we observe that most of MW's light is at $R \leq R_\odot$. This suggests that also most of the mass is within $R_\odot$, hence $M(< R) \approx M_{\rm MW}$, the total mass of the MW. For $M(< R) = M_{\rm MW}$ constant, Eq. (5.1) gives $V_c \propto R^{-1/2}$. The function $V_c(R)$ (circular velocity as function of radius) is called the **rotation curve**, and we expect the disk to be in *differential rotation*, with rotation curve $V_c \propto R^{-1/2}$. We can test this assertion by studying the motion of stars in the solar neighbourhood.

5.1.2 Oort's constants

Assume all disk stars are on circular orbits, with circular velocity $V_c(R)$ for a star at distance R from the MW centre. The observer (at radius R_0) is moving with circular velocity $V_{c,0} \equiv V_c(R = R_0)$ measures the following radial, V_r and tangential, V_t of a star at distance d and radius R, which has galactic coordinates $(b = 0, l)$ (see Fig.5.1)

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

In the indicated right-angled triangle

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

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

[2]Demonstrate this as an exercise.

[3]$R_\odot \sim 8$kpc denote here the distance of the Sun to the centre of the MW.

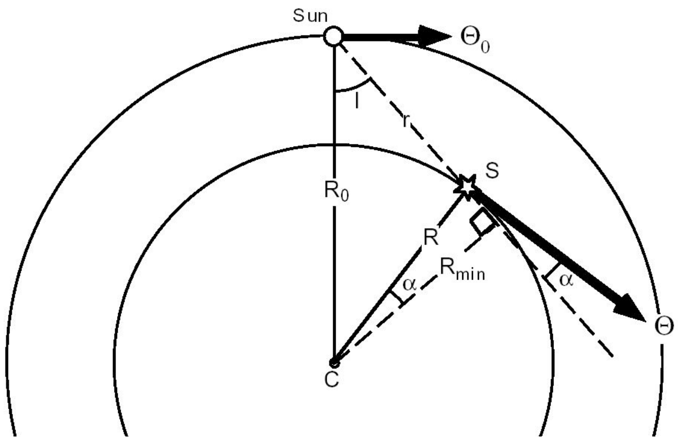

Figure 5.1: (Taken from http://www.csupomona.edu/˜jis/1999/kong.pdf) The observer is moving on a circular orbit with velocity Θ_0 and radius R_0, and observes a star at distance r, and galactic longitude l, on circular orbit with velocity Θ and radius R. In the text, r is denoted as d and the circular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

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

where $\Omega_0 \equiv V_{c,0}/R_0$ is the *angular velocity* of the observer, and $\Omega \equiv V_c/R$ the angular velocity of the star. For *nearby stars*, $R \approx R_0$ and $\Omega(R) \approx \Omega_0 + (d\Omega/dR)(R - R_0)$. If we now define **Oort's constants** A and B as

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

then

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

For example a star toward the galactic centre (or anti-centre; $l = 0$ and $l = 180^o$ respectively), has $V_r = 0$ and $V_t = (A + B)d$.

Jan Oort[4] measured (V_r, V_t) for stars as function of l and d, and inferred $A \approx 14.4 \pm 1.2 \mathrm{km\,s^{-1}\,kpc^{-1}}$ and $B = -12.0 \pm 2.8 \mathrm{km\,s^{-1}\,kpc^{-1}}$.

Using the results from section 5.1.1 we find that for a Keplerian disk, $V_c \equiv V_{c,0}(R_0/R)^{1/2}$, hence $dV_c/dR = -(1/2)V_c/R$. At the position of the Sun, $V_{c,0} \approx 220 \mathrm{km\ s^{-1}}$, with $R_0 \approx 8$kpc, hence we expect the values of Oort's constants to be

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

Our expected values are inconsistent with Oort's measured values. Assuming we measured V_c/R correctly, the measured and expected values for dV_c/dR are:

[4]Note that these are values appropriate for the Sun.

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

As expected, the Keplerian circular velocity drops $\propto 1/R^{1/2}$, hence $\frac{dV_c}{dR}| < 0$, but the *measured* value is consistent with zero: **the Mily Way's rotation curve is flat, *i.e.* $V_c(R) \approx$ constant**.

Of course in a real galaxy stars are not *exactly* on circular orbits, and each star has a small *peculiar* velocity with respect to the perfect circular motion $V_c(R)$. A standard of rest that moves on an exact circular orbit is called the 'local standard of rest'[5]: the speed of the Sun is ~ 16 km s^{-1} with respect to its local standard of rest.

Oort also discovered a small number of stars with very large deviations from the expectation given by Eqs. (5.6), which he called *high velocity stars*. He correctly identified these with stars belonging to the halo: the high velocity is because the halo does not rotate, whereas the disk, and the Sun with it, rotates with a speed ~ 220 km s^{-1}.

5.1.3 Rotation curves measured from HI 21-cm emission

The HI 21-cm emission line can be used to measure the velocity of the HI from its Doppler shift. This can be used to determine the rotation curve of the HI gas[6], if it were possible to determine R. Eqs. (5.2) show that V_r has a *maximum*

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

for any l with $|l| < \pi/2$, which occurs for $\alpha = 0$, at which point $d = R_0\cos(l)$. Therefore

$$\frac{dV_{r,\rm max}(l)}{dl} = \frac{dV_c(R)}{dR}\frac{dR}{dl} - V_{c,0}\,\cos(l) \qquad (5.10)$$

[5]Note this is not an inertial system

[6]You can attempt this in L4, using the radio dish on the physics building

hence

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

The 21-cm analysis confirmed Oort's measurements: the MW's rotation curve near the Sun is essentially flat.

It is much easier though to measure rotation curves in other spiral galaxies, by simply measuring the Doppler shift as function of distance to the centre. Such measurements are not available for tens of spirals, and they all show flat rotation curves in their outskirts. The MW is definitely not unusual in this respect.

5.2 Rotation curves and dark matter

A flat rotation curve, instead of a $1/R^{1/2}$ falling rotation curve, implies the mass is not all concentrated within the solar circle, but is more extended. To find the shape of the density distribution that gives rise to a flat rotation curve, take the derivative[7] with respect to R of

$$V_c^2\,R = GM\,, \tag{5.12}$$

for V_c is constant:

$$\begin{aligned} V_c^2 &= G\frac{dM}{dR} \\ \rho(R) &= \frac{V_c^2}{4\pi G R^2}\,. \end{aligned} \tag{5.13}$$

Hence a spherical distribution of mass, with $\rho(R) \propto 1/R^2$, gives rise to a flat rotation curve. But the observed light distribution in the MW is very different from this. This suggests three equally astonishing alternatives,

1. The mass-to-light ratio of stars in spirals conspires such that, although the light is very much centrally concentrated, the mass in stars is not. But stellar populations do not seem to vary significantly between the centre and the outskirts. However there may be unseen gas for example in the outskirts of the Milky Way providing the required $\rho(R) \propto 1/R^2$ density.

[7]Show that $dM/dR = 4\pi\rho(R)\,R^2$ in a spherically symmetric density distribution.

2. The Milky Way contains invisible matter, which does not emit, nor absorb light.

3. Gravity does not behave as $1/R^2$ on galactic scales. If this were true, then our reasoning above is simply not valid. 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.

The currently favoured option is (2), namely that the MW, and other galaxies, contains invisible dark matter. More evidence for this later, including the fact that this matter cannot be baryonic[8] in nature.

5.3 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*[9] 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

[8]Baryons are subatomic particles made out of three quarks, such as protons and neutrons. The dark matter has to be composed of something else, hence cannot be in the form of faint stars, planets, rocks, or past exam papers.

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

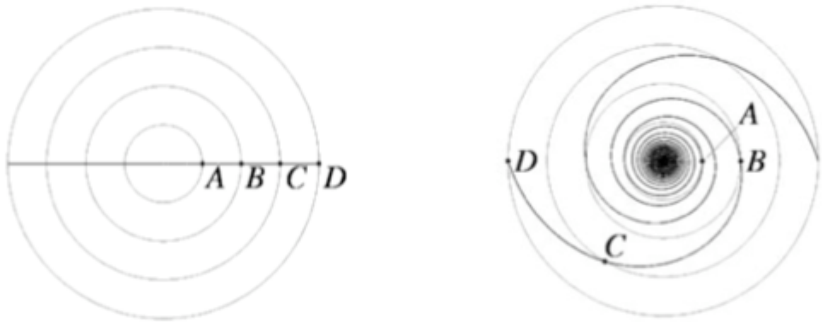

Figure 5.2: A spiral pattern made out of stars will rapidly wind-up in a disk undergoing differential rotation.

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 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.[10] 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.

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

5.4 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.

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, $v_e = 2^{1/2} \, V_c \approx 311$km s^{-1} (Using $V_c = 220$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 $\Phi =$ 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

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

where M_{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

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

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

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

The last three equations can be combined to eliminate $R_{\max}$, G and M_{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 $V_c = 220$km s^{-1} out to $R_\star$, then from $V_c^2 = GM/R_\star$ we find

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

If, as is more likely, the rotation speed eventually drops below 220km 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. Unlike spirals, they have a large range in sizes, and rotation is generally unimportant. Their interstellar medium is very different from that of spirals, consisting of very tenuous, hot, X-ray emitting gas. Observations of this gas, and stellar dynamics, both suggest the presence of significant amounts of dark matter.

7.1 Luminosity profile

Elliptical galaxies have smooth surface brightness (SB) profiles, which in projection are ellipsoidal in shape (Fig.7.1). Note that the isophotes near the centre are well defined, becoming more noisy in the outskirts.

Average intensities as function of radius for these galaxies are shown in Fig. (7.2). The drawn line which fits the data well, is the de Vaucouleurs or '$r^{1/4}$' fit introduced previously in Eq. (3.2) for the MW's bulge:

$$I(r) = I_e \exp[-7.67(r/r_e)^{1/4} - 1]\,. \tag{7.1}$$

Here, $I(r)$ is the intensity at projected distance r from the centre. When surface brightness, $\mu = -2.5\log(I) + \text{const}$, is plotted as function of $r^{1/4}$ this profile is a straight line as in Figure (7.3). The intensity profiles of most elliptical galaxies can be fit with just the two parameters I_e and r_e.

Some of the giant elliptical galaxies that are invariably found in the centres of big clusters of galaxies have an extended stellar halo not fit well by the

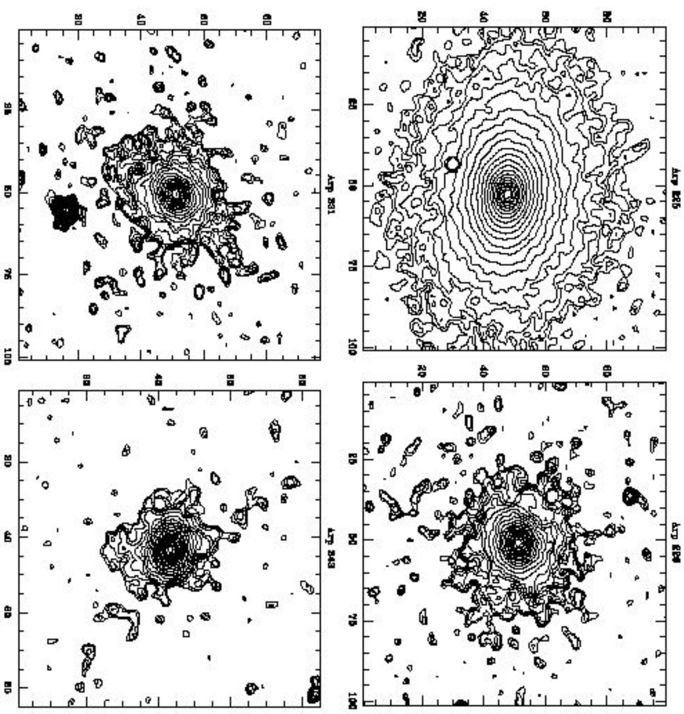

Figure 7.1: (Taken from astro-ph/0206097) Lines of constant surface brightness (isophotes) in the K-band for four elliptical galaxies. In successive isophotes, the surface brightness increases by 0.25 magnitudes.

arXiv:astro-ph/0206097 v1 6 Jun 2002

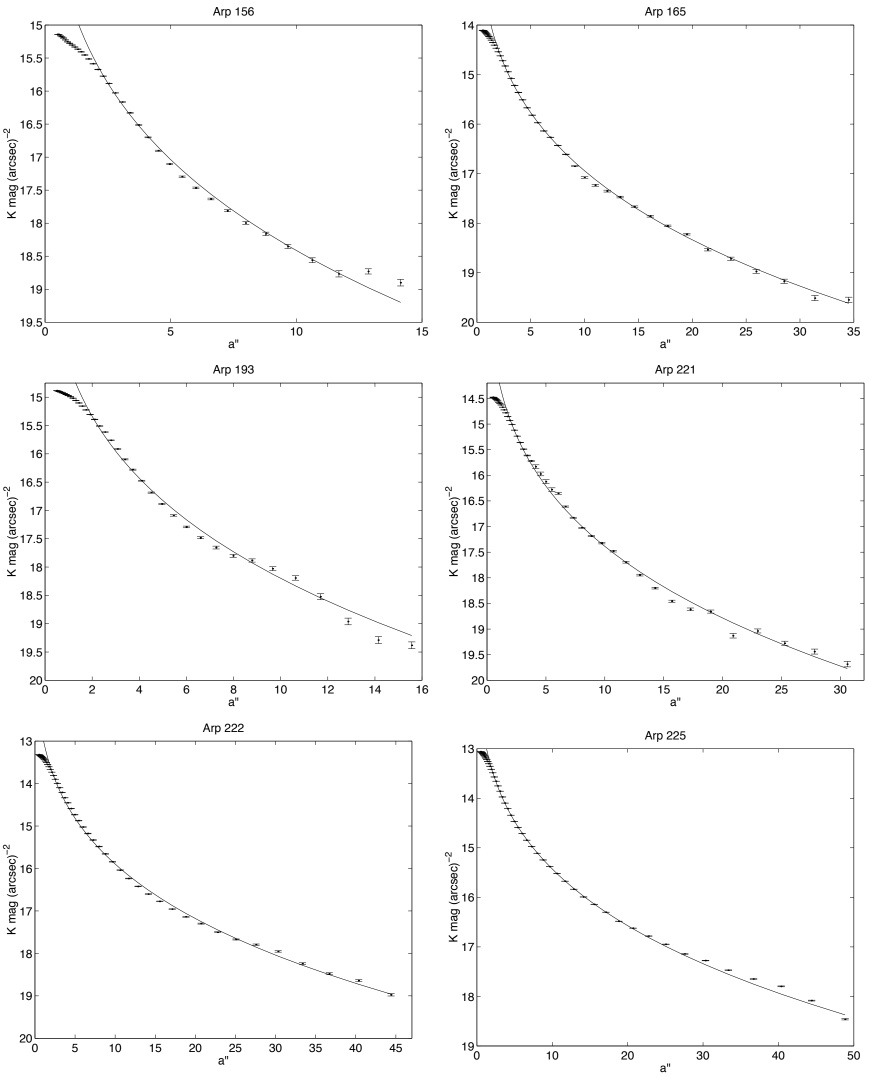

Figure 7.2: (Taken from astro-ph/0206097) Surface brightness as function of distance to the centre for the galaxies from Fig. 7.1. The drawn line is the best $r^{1/4}$ fit.

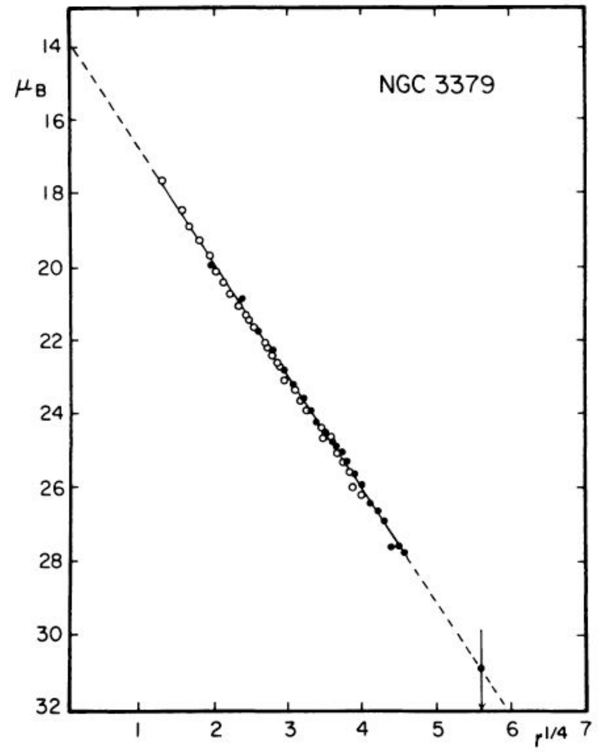

FIG. 2.—Mean E-W luminosity profile of NGC 3379 derived from McDonald photoelectric data. ●, Pe 4 data with 90 cm reflector; ○, Pe 1 data (M + P) with 2 m reflector. Note close agreement with $r^{1/4}$ law.

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.

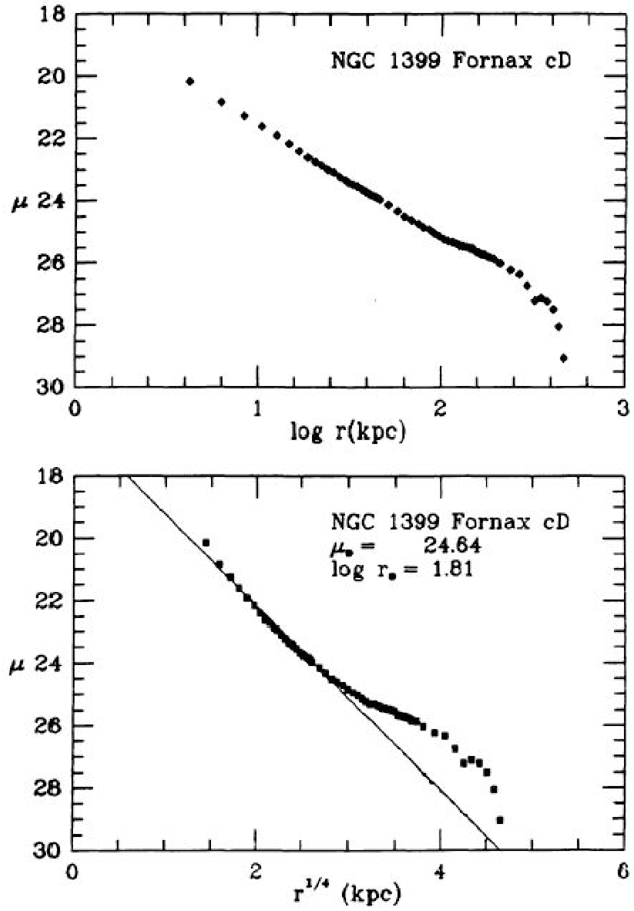

Figure 7.4: SB profile for NGC 1399 from Schombert (filled symbols; ApJS, 1989). The outer parts of the profile do not follow the $r^{1/4}$ fit (bottom panel), but are nearly a straight line in a SB-log(r) (i.e. a log − log) plot (top panel), meaning the profile is close to a power-law.

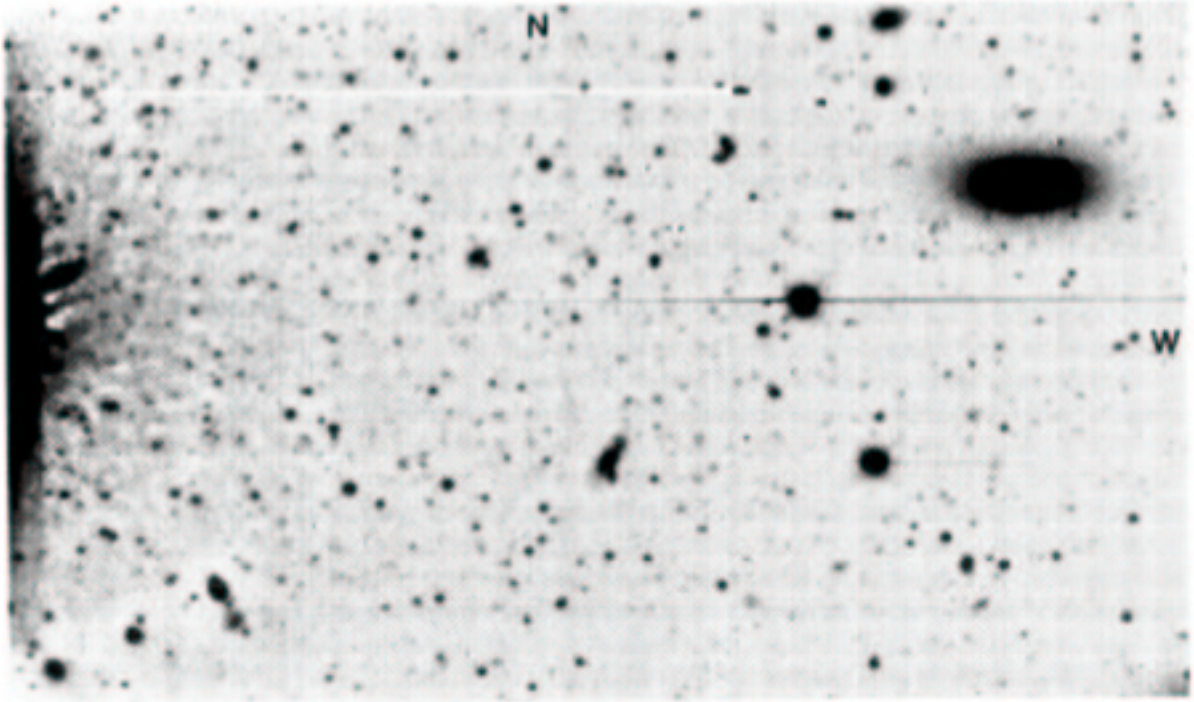

Figure 7.5: Image of NGC 1399 with smooth, best fit $r^{1/4}$ profile subtracted (Bridges et al., AJ 101, 469 (1991). The centre of NGC 1399 is at the left, the object toward the top right is another galaxy in the galaxy cluster. The other extended objects in the image are other galaxies. Clearly seen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

de Vaucouleurs profile. Such galaxies are called cD galaxies. An example is NGC 1399, the cD galaxy in the nearby Fornax cluster of galaxies (Fig. 7.4). The extent of such a halo can be truly enormous and be traced to around a Mpc, making the size of NGC 1399 on the night sky about the size of the moon, even though Fornax is at a distance of ~ 20 Mpc.

Ellipticals 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 surface brightness objects, indistinguishable on this plate from foreground stars, which are in fact GCs in the halo of the galaxy.

What is the origin of the $r^{1/4}$ profile? Numerical simulations of of collisions between two spiral galaxies produce objects with surface brightness profiles resembling the $r^{1/4}$ profile. Since elliptical galaxies are found in dense environments where such collisions are frequent, it seems plausible that elliptical galaxies form from collisions between spirals galaxies. However the stellar populations of ellipticals and spirals differ: ellipticals contain mostly old and metal rich stars (see below), whereas spirals contain younger and more metal poor stars. So ellipticals may still have formed from collisions, but not collisions between *todays* spiral galaxies.

7.2 Stellar populations and ISM

The redder colours of stars in elliptical galaxies is due to

1. the absence of star formation: the blue light in spirals is partly due to young and massive stars, but these massive stars are not present in ellipticals

2. the higher metallicity of stars in ellipticals. The presence of heavy elements ('metals') in stars leads to redder colours, both because of changes in the stellar structure, and because the metals in the stellar atmosphere redden the emitted light.

The fact that a stellar population is blue either because it is young (and the blue light is emitted by massive and hence young stars), or because it is metal poor, or both, is called the *age-metallicity degeneracy*: the colour of the population is not enough to tell a young but metal rich population, from an old metal poor one. Metallicity measurements of stars in elliptical galaxies have shown that both effects, absence of star formation, and high metallicities, cause the redder colours of ellipticals.

Dust and gas Whereas the ISM in spirals consists of dust and cold gas, these are generally absent in ellipticals. The beautiful Sombrero galaxy (Fig. 7.6) is an exception: a recent merger with a smaller system has left a strong dust lane strung across this elliptical. Such recent mergers may be quite common: a number of ellipticals have faint rings of stars around them (see for example Fig. 7.7), probably a result of such 'galactic cannibalism'.

7.3 X-rays

X-rays are absorbed by the Earth's atmosphere hence are observed from rockets or satellites. They are focussed onto a detector by placing a set of nearly parallel plates in the beam: the X-rays ricochet off these plates onto the detector[1].

The emission of X-rays from elliptical galaxies is due to two unrelated sources: hot gas in their interstellar medium (thermal bremsstrahlung), and binary stars. Binary stars may emit X-rays due to mass transfer (see your notes on stars). Subtracting such point sources from the X-ray image of the galaxy reveals extended X-ray emission from *thermal bremsstrahlung*.

Thermal bremsstrahlung[2] is the process whereby high energy radiation (X-rays) is produced in a plasma of ions and electrons, due to the electromagnetic deflection of a high-speed (thermal) hot electrons passing close to a positively charged ion. The deflection of the electron means it is accelerated, and an accelerating charge emits electromagnetic dipole radiation: this is the radiation that we observe[3] . The plasma is highly ionised because it is so

[1]See additional images on the web-page or http://chandra.harvard.edu/xray_astro/.

[2]German for 'braking radiation'.

[3]Note that the detected radiation is **not** blackbody radiation from the hot gas: the spectrum of thermal bremsstrahlung is not the same as that of a black body. Also other ions undergo electromagnetic deflections, but since they are much more massive than the electrons, their contribution to the emission is negligible.

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

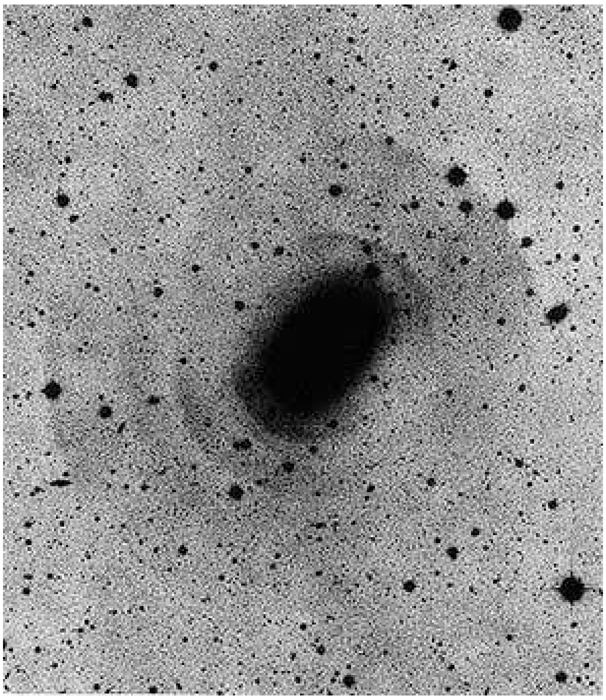

Figure 7.7: Image of NGC 3923 taken by David Malin. Several faint shells of stars appear as ripples in the outer parts of the galaxy, probably resulting from the merger of the central galaxy with a much smaller system. (see http://www.ast.cam.ac.uk/AAO/images/general/ngc3923.html)

hot, and the electrons are not bound to any particular ion. For this reason, thermal bremsstrahlung is sometimes called 'free-free' radiation.

The *emissivity* ϵ (the amount of energy radiated per unit volume) is

$$\epsilon \propto n_e \, n_i \, T^{1/2} \,, \tag{7.2}$$

where n_e and n_i are the electron and ion number densities respectively, and T is the temperature of the plasma.

Thermal bremsstrahlung produces a power-law spectrum up to some cut-off frequency, but the observed X-ray spectra of elliptical galaxies also show *absorption lines*, where electronic transitions in some highly ionised elements present in the hot plasma absorb the thermal bremsstrahlung. This makes it possible to measure *abundances* of elements in the ISM.

It is often possible to detect several different transitions of Iron. The relative strength of these transitions depends only on the temperature T of the gas, and not on the abundance of Iron: this makes it possible to measure T from the X-ray spectrum.

7.4 Evidence for dark matter from X-rays

It is possible to constrain the density distribution of the elliptical galaxy by assuming the hot X-ray emitting gas is in *hydrostatic equilibrium*, much in the same way as one can derive the equations for stellar structure. Gravity pulls the gas inward, whereas the pressure of the gas pushes it outward. In hydrostatic equilibrium these two forces balance. We can determine the temperature T of the gas from its X-ray spectrum, its density ρ from the emissivity, $\epsilon \propto n_e \, n_i \, T^{1/2}$ and hence infer its pressure, p, as function of radius. Hydrostatic equilibrium then allows us to measure the gravitational potential and hence the total density distribution. Assume for simplicity that the galaxy is spherical.

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 F_g on the shell is due to this total enclosed mass

$$F_g(r) = \frac{GM(< r)M_s}{r^2}\,, \tag{7.3}$$

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 pressures:

$$F_p(r) = 4\pi r^2(p(r) - p(r + dr)) = -4\pi r^2 \frac{dp}{dr}\, dr\,, \tag{7.4}$$

where the pressure was expanded in a Taylor series, assuming $dr << r$. Expressing hydrostatic equilibrium, $F_g = F_p$, yields

$$\frac{GM(< r)}{r^2} 4\pi r^2 \rho_{\rm gas}(r)dr = -4\pi r^2 \frac{dp}{dr} dr\,, \tag{7.5}$$

hence

$$\frac{GM(< r)}{r^2} = -\frac{1}{\rho_{\rm gas}(r)} \frac{dp}{dr}\,, \tag{7.6}$$

or in terms of the gravitational potential Φ

$$\nabla\Phi = -\frac{\nabla p}{\rho_{\rm gas}}\,. \tag{7.7}$$

This should look familiar from hydrostatic equilibrium in stars!

The X-ray observations can be used to determine the right-hand-side: the X-ray spectrum yields T, and the emissivity ρ. Application of Eq. (7.6) then yields $M(< r)$.

The temperature T is observed to remain approximately constant, in which case we can obtain an approximate expression for the density profile. The pressure p in an ideal gas depends on T and ρ as

$$p = \frac{\mathrm{k_B}\, T}{\mu\, m_{\rm p}}\, \rho_{\rm gas}\,, \tag{7.8}$$

where $\mu\, m_{\rm p}$ is the mean molecular weight per particle. Substituting this into Eq. (7.6) and assuming T to be constant yields

$$GM(<r) = -\frac{{\rm k_B}T}{\mu\, m_{\rm p}} r^2 \frac{d\ln\rho_{\rm gas}}{dr}\,. \tag{7.9}$$

Finally taking the derivative of both sides with respect to r, gives

$$G4\pi r^2 \rho_{\rm tot} = -\frac{{\rm k_B}T}{\mu\, m_{\rm p}}\frac{d}{dr}(r^2\frac{d\ln\rho_{\rm gas}}{dr})\,, \tag{7.10}$$

where $\rho_{\rm tot}$ is the *total mass density*, *i.e.* due to stars, gas and dark matter.

If the total density and gas density were proportional, $\rho_{\rm gas} \propto \rho_{\rm total}$, we could solve Eq. (7.9) by trying a solution of the form $\rho_{\rm total} = \rho_0\,(r_0/r)^\beta$. Substitution shows this is indeed a solution provided $\beta = 2$, and yields

$$\rho_{\rm tot}(r) = \frac{{\rm k_B}\,T}{2\pi\, G\mu m_{\rm p}}\, r^{-2}\,. \tag{7.11}$$

Even if total and gas density do not trace each other it is possible to infer $\rho_{\rm total}$ from the X-ray data. We can then compare the inferred total density $\rho_{\rm total}$ with that from stars, $\rho_\star$, and gas, ρ_{rmgas} combined: this analysis demonstrates that elliptical galaxies also contain dark matter, since $\rho_{\rm total} > \rho_\star + \rho_{\rm gas}$.

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 ellipticals and shells of stars around them are thought to be evidence for galactic cannibalism.
- contrast the stellar population and ISM of ellipticals with those of spirals
- describe how X-rays are detected, and explain the process with which X-rays are produced in the hot gas in ellipticals
- explain how X-ray observations can be used to infer the gravitational potential and derive the equation for hydrostatic equilibrium, relating gas properties to the total density

Chapter 8

Elliptical galaxies. II

Stars in a spiral disk are on nearly circular orbits. Although some of the smaller elliptical galaxies sometimes rotate as well, the angular momentum of their stars is not sufficient to balance gravity. It is the **large random motions** of stars in an elliptical which balances its gravity. This is in fact similar to hydrostatic equilibrium in a star, where it is the pressure, $p \propto \rho T$, which balances gravity.

The random velocities v of the gas atoms generate the temperature, $T \propto v^2$. Similarly the high velocities of stars in an elliptical generate something akin to pressure, $p \propto \rho v^2$, where ρ is the density of stars, and v^2 their velocity dispersion. However there is a significant difference: the velocities of gas particles are isotropic, $\langle v_x^2 \rangle = \langle v_y^2 \rangle = \langle v_z^2 \rangle$, whereas in general those of stars in an elliptical are not. This leads to a significant difference: stars are spherical[1] but elliptical galaxies are not. This chapter explores why this is.

The Jeans's equations are the stellar dynamical analogue of the gas equations, and they can be used to relate stellar density and velocity distribution. We will not derive these equations here, but simply make them plausible. But first we need to show that a description in terms of fluid variables makes sense for a stellar system, by introducing the concept of *relaxation time*.

[1]Except for effects due to rotation and tidal or magnetic fields.

8.1 Relaxation in stellar systems

A star moving through a stellar system feels the gravitational pull of 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, $E = v^2/2 + \Phi$, are conserved along the stellar orbit.

Now consider the case where the smooth stellar density ρ is made-up of stars each with mass m. When the test star comes close to another star, it will be deflected from its orbit, in the process exchanging both energy and angular momentum[2]. Therefore in this case neither E nor L are conserved. But which description applies to stars in an elliptical galaxy?

Clearly, if the test star has many close encounters with other stars, its energy E will start to differ considerably from its initial value. *The relaxation time T_E is a measure of how long the star remembers its initial energy.* If T_E is very long, the energy of each star in (nearly) conserved, and the fluid description applies. Conversely if T_E is short, the energy of the star changes rapidly, and the second description applies: systems with short T_E cannot be described with fluid equations.

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, 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,

[2]Such 'gravitational slingshots' are exploited in interplanetary travels, for example the Voyager space craft gained enough speed to escape the solar system by passing very close to other planets.

the derivation here is slightly different from BT (p. 188).

Assume that the change in velocity $\delta\mathbf{v}$ as the test star passes another star is small, and therefore approximate the orbit as a straight line, traversed with constant velocity v. The force perpendicular to the orbit is

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

assuming 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 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?

A real stellar system has finite size, therefore the impact parameter should be limited to the radius R of the system: this avoids the divergence at infinity. Our derivation so far assumed the changes in velocity of the star during a single encounter to be small, but this is not true for $b \to 0$ which causes the divergence at small b. Therefore we want to replace the lower limit of integration by $b_{\rm min}$, which is that impact parameter for which the *change* in velocity is comparable to the *initial* velocity, $b_{\rm min} = Gm/v^2$ Making these somewhat suspicious changes we obtain

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

When the stellar system is in **virial equilibrium** its kinetic energy $K \sim N\, m\, v^2/2$ is half its the potential energy, $U \sim G\,(N\, m)^2/R$, hence

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

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

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

The first line combines Eq.(8.6) with our finding $R/b_{\min} = N$ and the definition of the crossing time. The second step assumes virial equilibrium, Eq. (8.7).

We now define the *relaxation time* $T_{\rm relax}$, as the time it takes for encounters to change the velocity of a test star by order of itself, hence

$$\frac{d\delta \mathbf{v}_\perp^2}{dt} T_{\rm relax} \equiv v^2 \,. \tag{8.9}$$

Combing the last two equations provides a relation between the relaxation time and the crossing time of a stellar system:

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

Systems of the same mass and size, have the same crossing time $t_{\rm cr} \propto R^{3/2}/M^{1/2}$ However their relaxation time depends on the *number* of stars, with $T_{\rm relax} \propto N/\ln(N)$. Put differently, the effect of encounters decreases with increasing particle numbers as $N/\ln(N)$.

A[3] typical globular cluster has $N = 10^5$ stars, radius $R = 10$pc, and velocity dispersion $v = 10$km s^{-1}, and hence $t_{\rm cr} \approx 10^6$ years but $T_{\rm relax} \approx 10^9$ years. The relaxation time of such a system is therefore much less than its age, and encounters between stars have significantly changed the stellar orbits. It would be wrong to describe globular clusters with fluid equations.

However, a big elliptical galaxy may have $N = 10^{11}$ stars, radius $R = 20$kpc, and velocity dispersion $v = 200$km s^{-1}, therefore $t_{\rm cr} \approx 10^8$yr, and $T_{\rm relax} \sim 10^{19}$yr. The relaxation time is **much** longer than the age of the Universe ($\sim 10^{10}$yr), and encounters between stars have had negligible effect on stellar orbits. Such a system is called *collisionless* and can be described using a set of fluid-like equations called the Jeans equations.

8.2 Jeans equations

The Jeans equations are the stellar dynamical analogues of fluid equations. Here we only consider the stellar dynamical analogue of the equation for

[3] A handy approximate relation is $1{\rm km\,s^{-1}} \approx 1{\rm pc\,Myr^{-1}}$.

hydrostatic equilibrium in an ideal gas

$$\begin{aligned} \nabla\phi &= -\frac{1}{\rho}\nabla p \\ p &= n\,\mathrm{k_B}T\,, \end{aligned} \tag{8.11}$$

where ϕ is the gravitational potential, ρ the density, T the temperature, $n = \rho/m$ the particle density, and p the pressure. Temperature and velocity dispersion of the particles are related by[4]

$$\frac{1}{2}mv^2 = \frac{3}{2}\,\mathrm{k_B}T\,, \tag{8.12}$$

hence $p = \rho v^2/3$. Therefore the equation for hydrostatic equilibrium in an ideal gas can be written as

$$\nabla\phi = -\frac{1}{\rho}\,\nabla\,\frac{\rho v^2}{3}\,, \tag{8.13}$$

or in Cartesian Coordinates

$$\begin{aligned} \frac{\partial\phi}{\partial x} &= -\frac{1}{\rho}\frac{\partial}{\partial x}\,\rho v_x^2 \\ \frac{\partial\phi}{\partial y} &= -\frac{1}{\rho}\frac{\partial}{\partial y}\,\rho v_y^2 \\ \frac{\partial\phi}{\partial z} &= -\frac{1}{\rho}\frac{\partial}{\partial z}\,\rho v_z^2\,. \end{aligned} \tag{8.14}$$

Collisions between atoms in a gas guarantee that velocities are isotropic, $\langle v_x^2\rangle = \langle v_y^2\rangle = \langle v_z^2\rangle$, therefore the right-hand-sides of these equations are the same, hence also $\frac{\partial\phi}{\partial x} = \frac{\partial\phi}{\partial y} = \frac{\partial\phi}{\partial z}$ and the density distribution for the *star* is spherical.

However, encounters between stars in an elliptical galaxy are so infrequent that this is not true, $\langle v_x^2\rangle \neq \langle v_y^2\rangle \neq \langle v_z^2\rangle$, and the right-hand-sides are **not** equal, and the density distribution can be ellipsoidal: it need not be spherical. It is the anisotropy of velocities of the stars in an elliptical galaxy

[4]These relations are derive from the kinetic gas theory.

that allow it to be ellipsoidal in shape[5].

In conclusion: the behaviour of collisionless stellar systems is similar to that of self-gravitating gas spheres (stars). 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 gravity. 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 (circular) motions[6]

In a gas the particle velocities are isotropic and hence stars are spherical. In contrast in collisionless stellar systems velocities are not isotropic and the system is elliptical.

[5]Careful readers may have noticed that the semi-major axes of the ellipsoid are along our chosen Cartesian axes. In general this is of course not the case, and the correct equations are not Eq. (8.14). Fortunately it is always possible to rotate the system of axes so that they do turn into this set.

[6]Compare for example the large random velocities of shoppers in a shopping mall during the sales (hot system) versus the streaming motion of commuters leaving Durham station (cold system). The speeds of people may be the same in both cases, but the commuters all move in the *same* direction, hence their relative velocities are much smaller than those of the shoppers.

8.3 Summary

After having studied this lecture, you should be able to

- interpret the equation for the stellar relaxation time, and discuss its relevance for being able to describe a stellar system as collisionless (a galaxy) or collisional (a globular cluster)
- 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 why the ellipticity of stellar systems is related to the underlying orbits of the stars.

Chapter 9

Groups and clusters of galaxies

Galaxies are not distributed homogeneously in the Universe. Galaxies such as the Milky Way huddle together in small groups (the Local Group), voids are very large regions of space where galaxies are very rare, and *clusters* of galaxies are regions where the galaxy density is very high. Galaxy 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

The Milky Way, Andromeda, and some 40 small irregular galaxies within 2Mpc or so from the MW, are part of a gravitationally bound system, called the Local Group (see Sect. 6.2). Most spiral galaxies are found in such loose groups.

Galaxy clusters are gravitationally bound systems of 10s-100s of elliptical galaxies and 1000s of smaller dwarf galaxies. Fornax and Virgo are two clusters of galaxies within $\sim 20\,\mathrm{Mpc}$ of the Milky Way. The galaxy M 87, which has a super massive black hole (see Sect 11) with mass $\sim 10^9\,M_\odot$ lurking in its centre, is the central object in the nearby *Virgo* cluster of galaxies. The Local Group, including the Milky Way, is falling toward Virgo. The large cD galaxy NGC 1399 (see Sect.7.1) is at the centre of the *Fornax* cluster of galaxies. The *Coma* cluster of galaxies, at a distance of 90Mpc, has a mass as large as $10^{15}M_\odot$. Clusters such as Coma contain the most massive galaxies known.

George Abell and collaborators eye balled photographic plates in the 1950s to identify galaxy clusters as over dense concentrations of galaxies. He discovered some 4500 clusters, ranking them from poor to rich, according to the number of galaxies they had. Distant clusters are still identified with the number Abell allocated.

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 these systems are thought to be in dynamical equilibrium.

9.1.1 Evidence for dark matter in clusters from galaxy motions

The Swiss astronomer Fritz Zwicky measured Doppler velocities of galaxies to infer the dynamical mass of galaxy clusters in the 1930s. Inferring the mass in stars from the observed luminosity, he decided that clusters must contain dark matter: this was the first time that the need for dark matter arose. Zwicky's mass measurement was based on the following argument.

Assume all cluster galaxies have the same mass m, and that the cluster contains N galaxies. The kinetic energy of the system 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

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

Note that these velocities are measured with respect to the *mean* velocity of the cluster. 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 can be determined from measuring σ and R from

$$M = \rule{10cm}{0.4pt}\,. \tag{9.4}$$

This allowed Zwicky to estimate the dynamical mass, M, of the cluster, from the observed velocities and size of the system.

Note that the velocity inferred from the Doppler shift, $\lambda = \lambda_0\,(1 + v_r/c)$, is only the *radial component* of the velocity, but the kinetic energy employs the *total* velocity, $v^2 = v_r^2 + v_t^2$, the sum of radial and tangential components. Since we cannot measure v_t, Zwicky assumed[2] that $v_t^2 = 2v_r^2$, since we (the observers) are presumably not in a special position.

Zwicky estimated the masses of the galaxies as inferred from their luminosities, $M_L = \Gamma\,L$, where $\Gamma_L \sim 3\,M_\odot/L_\odot$ is the stellar mass-to-light ratio, and compared the result with the dynamical mass, M. He found that, not only was the dynamical mass much larger, $M \gg M_L$, 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: $K + U > 0$. 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, hence introducing *dark matter*.

[2]The extra factor 2 arises because the tangential velocity has two Cartesian components. To see this, choose a Cartesian system of axes with z along the radial direction, then $v_r^2 = v_z^2$, whereas the tangential velocity $v_t^2 = v_x^2 + v_y^2$. Assuming isotropy, $v_x^2 = v_y^2 = v_z^2$, hence $v_t^2 = 2v_r^2$.

9.2 Evidence for dark matter from X-rays observations

The hot gas[3] in clusters emits X-rays due to *thermal bremsstrahlung*, which can be used to determine their mass assuming the hot gas is in hydrostatic equilibrium. We did the same for elliptical galaxies in Section 7.3. The (total) mass of the cluster in a sphere of radius r, relates to gas density, $\rho_{\rm gas}$, and pressure, p, as (see Eq. 7.6)

$$\frac{GM(<r)}{r^2} = -\frac{1}{\rho_{\rm gas}(r)}\frac{dp}{dr}\,. \tag{9.5}$$

As for ellipticals, the X-ray data can be used to measure the right hand side, allowing a determination of the enclosed mass, $M(<r)$. Comparing this to the mass in stars and gas, $M_\star = \Gamma\, L$, and $M_{\rm gas}$ inferred from the X-ray emissivity, yields $M \gg M_\star + M_{\rm gas}$: X-rays strongly indicate the presence of dark matter in clusters.

What is the origin of the gas, and why is it so hot? A high-mass cluster will suck gas in from its surroundings due to its large gravitational pull. The accreting gas slams into the gas already there, and the rapid compression of the gas converts kinetic energy into thermal energy in an *an accretion shock*.

Assume a parcel of gas starts at infinity has velocity $v = 0$, and is accreted by a cluster with mass M. It slams into the gas already present at distance R (the radius of the cluster). Expressing energy conservation yields

$$E = 0 = \frac{1}{2}v^2 - \frac{G\,M}{R}\,, \tag{9.6}$$

since the energy of the parcel at infinity is zero. When the gas converts its kinetic energy $(1/2)\,v^2$, into thermal energy, it gets heated to temperature T given by

$$\frac{3}{2}\frac{{\rm k_B}T}{\mu\, m_p} = \frac{1}{2}\,v^2\,. \tag{9.7}$$

[3]Note that the emission is extended, and not just due to the elliptical galaxies themselves: most of the hot gas is between the galaxies, not associated with any particular galaxy.

Here, μ is the mean molecular weight of the gas. These equations relate the cluster's mass and radius to the **virial temperature** T of the gas:

$$\frac{G\,M}{R} = \underline{\hspace{12cm}}\,. \tag{9.8}$$

Clusters have such high values of T that all the hydrogen atoms are fully ionised, hence[4] $\mu = 1/2$.

9.3 Metallicity of the X-ray emitting gas.

The X-ray spectra of clusters consists of a power-law component due to thermal bremsstrahlung, with additional emission lines mostly from inner-shell transitions of the Fe ion, see Figure 9.1. Recall that the ratio of emission lines can be used to determine T, whereas the emissivity $\epsilon \propto \rho_{\rm gas}^2 T$ then determine $\rho_{\rm gas}$. The first surprise is that most of the baryonic mass in the cluster is in the X-ray gas: $M_\star \ll M_{\rm gas}$. Although we identified galaxy clusters as regions with a high density of galaxies, most of the baryons in the cluster have not actually collapsed to form stars.

Once T and $\rho_{\rm gas}$ are known, the strength of the lines allows one to determine the *amount* of Fe in this intra-cluster medium. Comparison with the total amount of gas gives a mean metallicity $M_{\rm Fe}/M_{\rm gas} \approx 1/3$ of the solar value[5].

The high metallicity of the cluster *gas*, and the enormous amount of metals outside of galaxies, is an even bigger surprise. Note that metal abundance is higher than that of *stars* in most globular clusters and dwarf galaxies. Presumably these metals were produced in stars, and subsequently ejected from their parent galaxy into the intra-cluster gas. This demonstrates that galaxies are far from closed boxes: a large fraction of the products of stellar evolution are blown out of the galaxy. Current models suggest that this is due to the combined effect of many super nova explosions.

[4]A more accurate expression for μ takes into account the presence of Helium and other elements, giving $\mu \approx 0.6$.

[5]Recall that the Sun contains a mass fraction of 0.02 of metals.

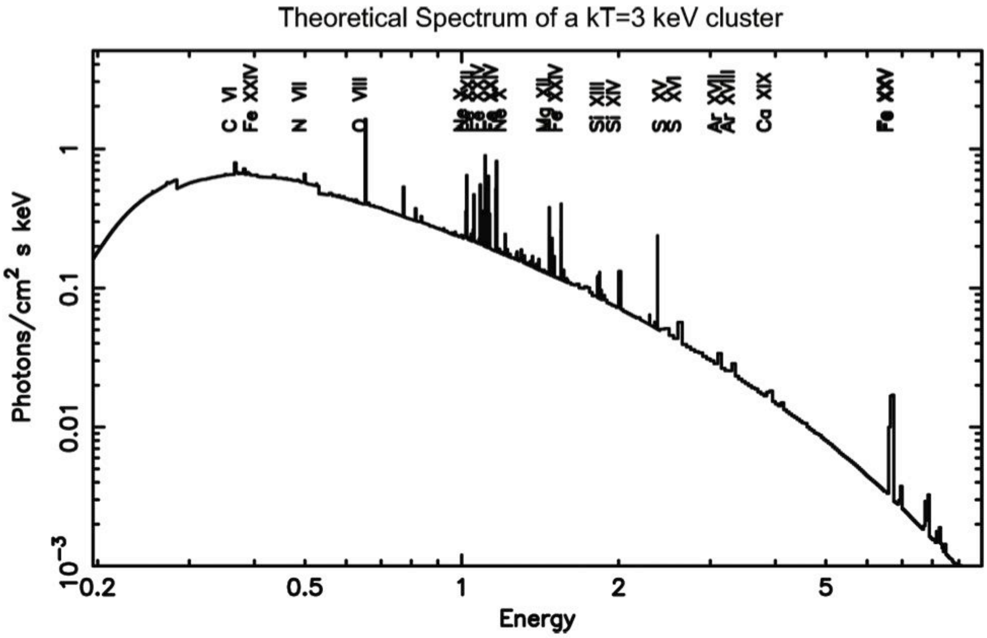

Figure 9.1: Theoretical emission spectrum of a hot plasma. Emission lines due to a range of elements (indicated) appear on top of a smooth continuum caused by thermal bremsstrahlung.

9.4 The dark matter density of the Universe

Clusters can be used to determine the dark matter density of the Universe as a whole. Assume the Universe starts-out smooth, with a (nearly) constant ratio ω of dark matter to baryons:

$$\omega = \frac{\rho_{\rm dm}}{\rho_{\rm b}}\,. \tag{9.9}$$

The forming cluster grows through accretion of dark matter and baryons, and so the density of both components increases. As the potential well of the cluster deepens, it is probably a good approximation to assume that eventually neither dark matter, nor baryons, can ever escape from the cluster, meaning the ratio of dark matter to baryons, *in the cluster as a whole*, $M_{\rm dm}/M_{\rm b}$, equals the primordial value:

$$\omega \approx \frac{M_{\rm dm}}{M_{\rm b}}\,. \tag{9.10}$$

Therefore we can determine ω by measuring the dark matter mass of a cluster, $M_{\rm dm} = M - M_{\rm b}$, and the baryonic mass, $M_{\rm b}$. Recall that we could determine M from either galaxy motions (Eq. 9.4) or X-ray observations, with $M_{\rm b} = M_\star + M_{\rm gas}$ from combining the stellar mass (from the observed stellar luminosity) and the gas mass from the X-ray emissivity. Doing so for a range of clusters yields $\omega \approx 6$.

An estimate for the mean baryon density, $\rho_{\rm b}$, follows from the *abundance* of Deuterium[6] relative to normal Hydrogen, $\rho_{\rm D}/\rho_{\rm H}$. Deuterium is produced during nucleosynthesis in the first three minutes after the Big Bang, and destroyed during nuclear burning in stars. Therefore any Deuterium found[7] is primordial and left over from Big Bang. Crucially, *the Deuterium abundance depends on the baryon density* (Figure 9.2). The inferred mean baryon density is $\rho_{\rm b} \approx 2.2 \times 10^{-7}$ hydrogen atoms per cm^{-3}. This is approximately 30 orders of magnitude lower than the density of the paper that these notes are printed on (or the density of the screen if you read them on your computer).

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

[7]Deuterium abundance measurements use quasar spectra (see Chapter 11 for a discussion of quasars) to probe nearly primordial gas in between the galaxies.

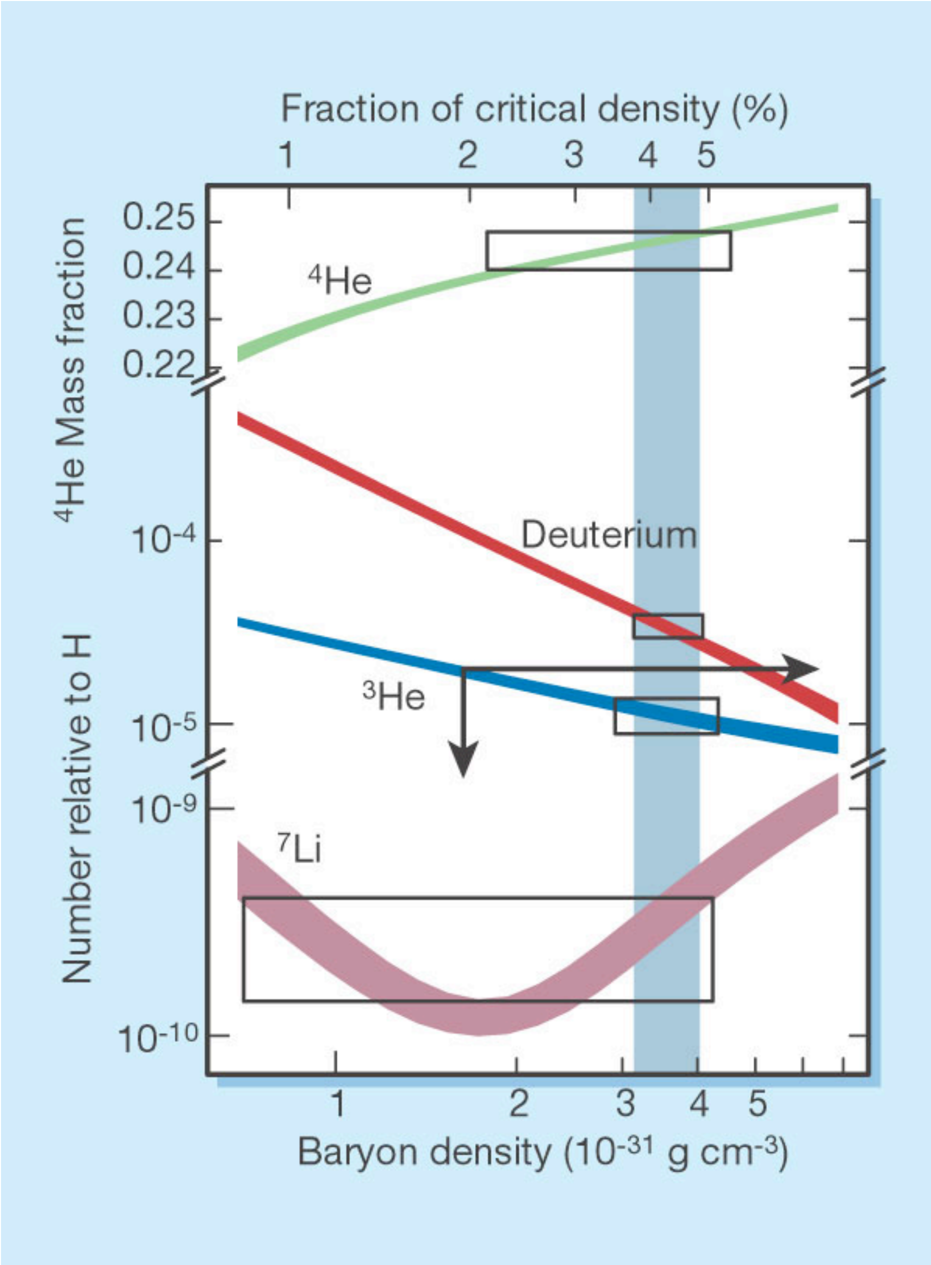

Figure 9.2: The production of various elements during Big Bang nucleosynthesis as a function of the baryon density. (*Nature* **415**, p. 27, 2002)

Combining the mean baryon density with $\omega = 6$ yields $\rho_{\rm dm} \approx 2.2 \times 10^{-30}$ g cm^{-3} for the mean dark matter density of the Universe.

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 a large fraction of the products of stellar evolution are blown out of galaxies
- 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

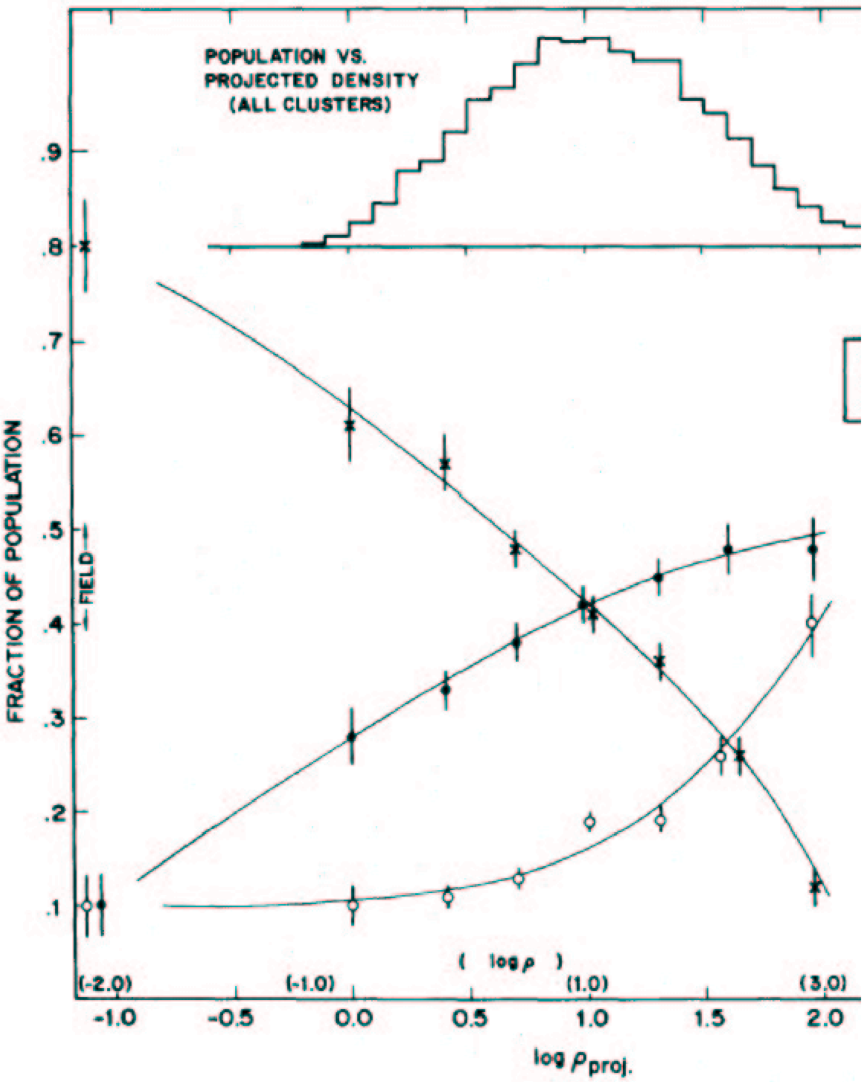

Figure 10.1: The fraction of Elliptical (E), S0, and Spiral + Irregular (S+I) galaxies as function of the logarithm of the projected density, in galaxies Mpc^{-2}. The data shown are for all cluster galaxies in the sample and for the field. Also shown is an estimated scale of true space density in galaxies 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.

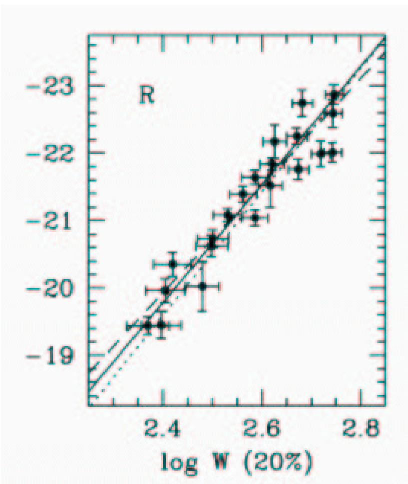

Figure 10.2: R-band absolute magnitude M of spiral galaxies as function of the maximum rotational velocity $W = V_c$ (in km s^{-1}). *Brighter* spiral galaxies (more negative M; recall that magnitude M is related to luminosity L as $M = -2.5\log(L)$+constant) rotate *faster*.

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

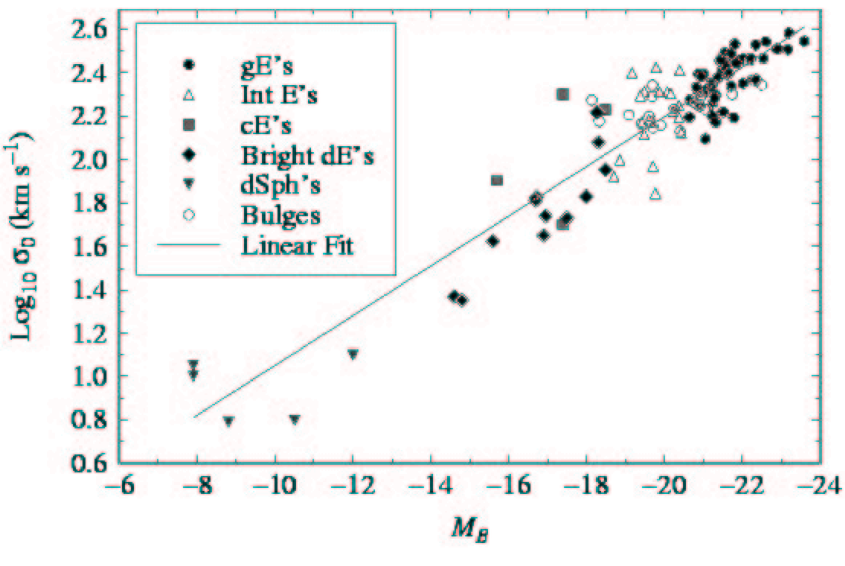

Figure 10.3: B-band absolute magnitude M for a sample of elliptical galaxies plotted versus the velocity dispersion σ (in km s^{-1}). *Brighter* ellipticals galaxies (more negative $M_{\rm B}$; recall that magnitude M is related to luminosity L as $M = -2.5 \log(L)$+constant) have a *larger* velocity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

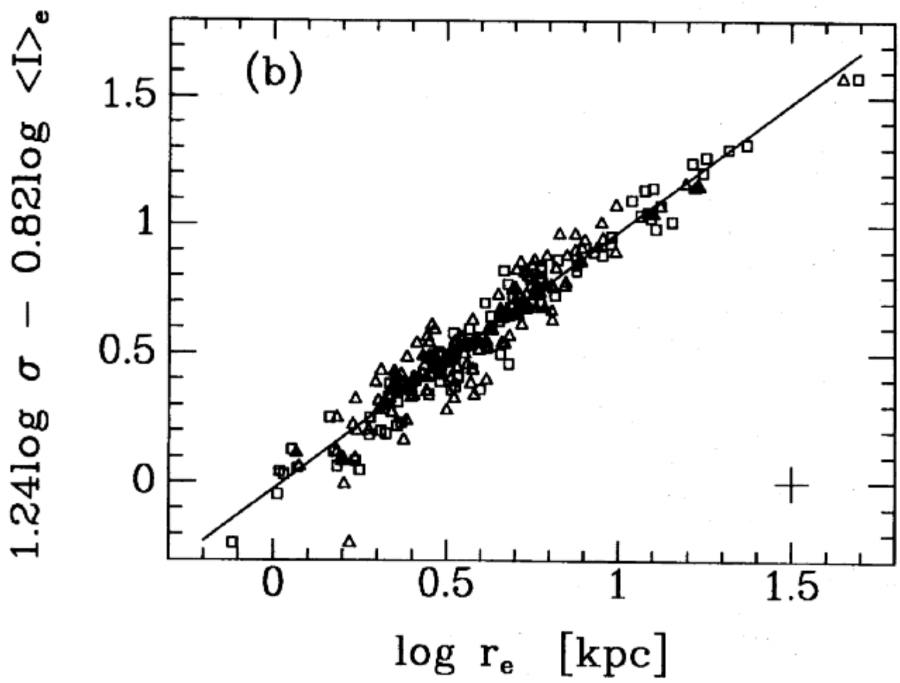

Figure 10.4: Scale radius R_e versus $\sigma^{1.24}\, I_e^{-0.82}$ for the same sample of Es plotted in Figure 10.3. The galaxies follow this relation much better, with much less scatter then the Faber-Jackson relation in Fig.10.3.

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

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 $L_\star \approx 2\times10^{10}L_\odot$ in the B-band. From Table **??**, we find that the total B-band luminosity of the MW is about $1.8\times10^{10}L_\odot$, 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!

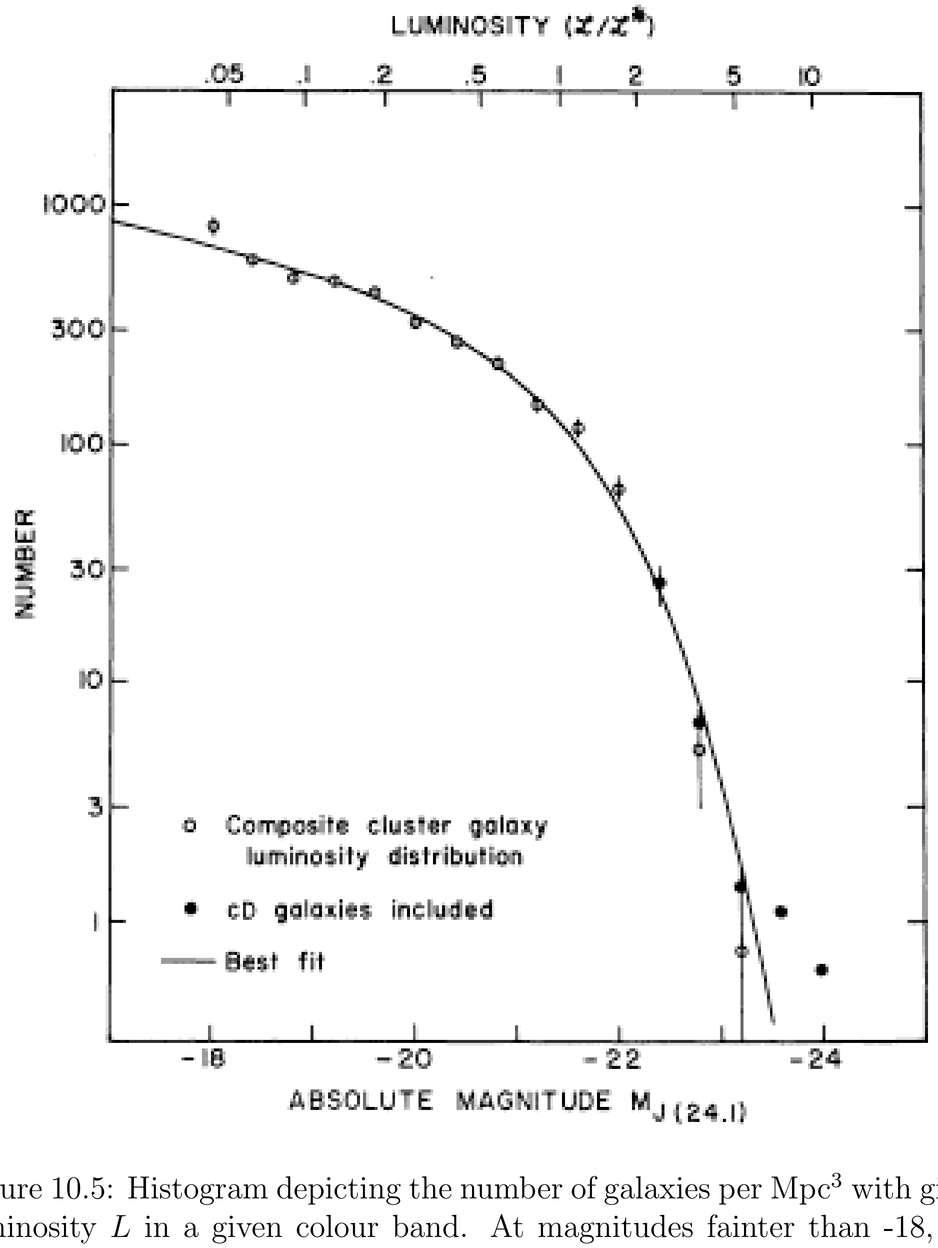

Figure 10.5: Histogram depicting the number of galaxies per Mpc^3 with given luminosity L in a given colour band. At magnitudes fainter than -18, the number increases as a power in L, but for magnitudes brighter than -22, the number of galaxies cuts-of exponential. The drawn line is the best-fit Schechter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

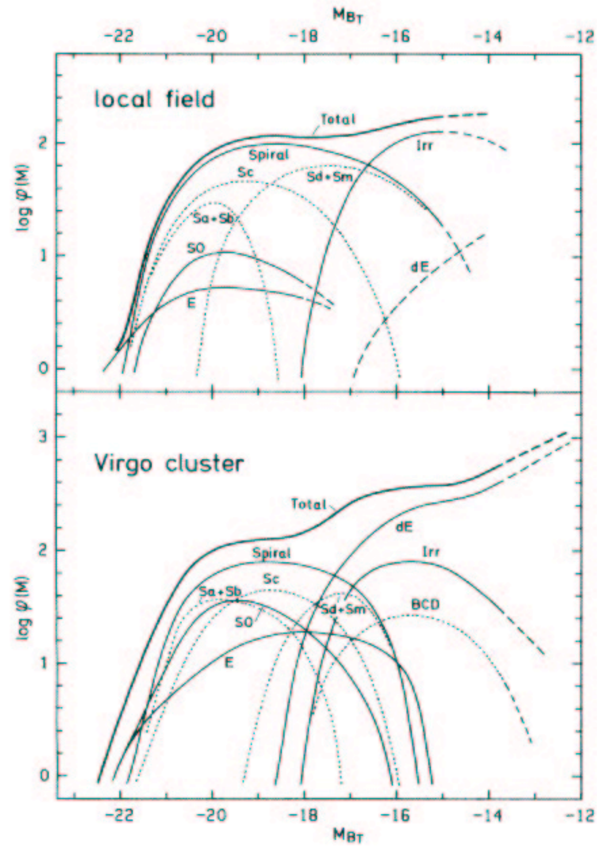

Figure 1 The LF of field galaxies (top) and Virgo cluster members (bottom). The zero point of $\log \varphi(M)$ is arbitrary. The LFs for individual galaxy types are shown. Extrapolations are marked by dashed lines. In addition to the LF of all spirals, the LFs of the subtypes Sa + Sb, Sc, and Sd + Sm 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$(M_{\rm DO}/2{\times}10^8 M_\odot)\,(\sigma/200{\rm km\ s^{-1}})^{-2}\,(D/5{\rm Mpc})$. 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,

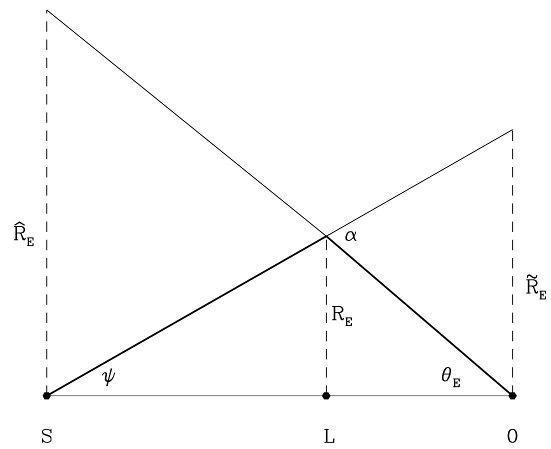

Figure 12.1: The source (S), lens (L) and observer (O) are all aligned. Light from S is deflected by an angle α toward O. Because of symmetry, O sees the source S as ring of radius R_E centred around L.

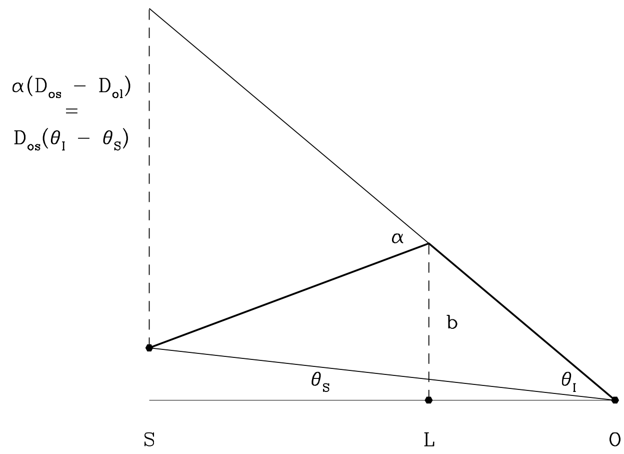

Figure 12.2: The source (S), lens (L) and observer (O) are now misaligned, and O sees two images of S, one of which is depicted here while the other one lies on the other side of L.

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

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

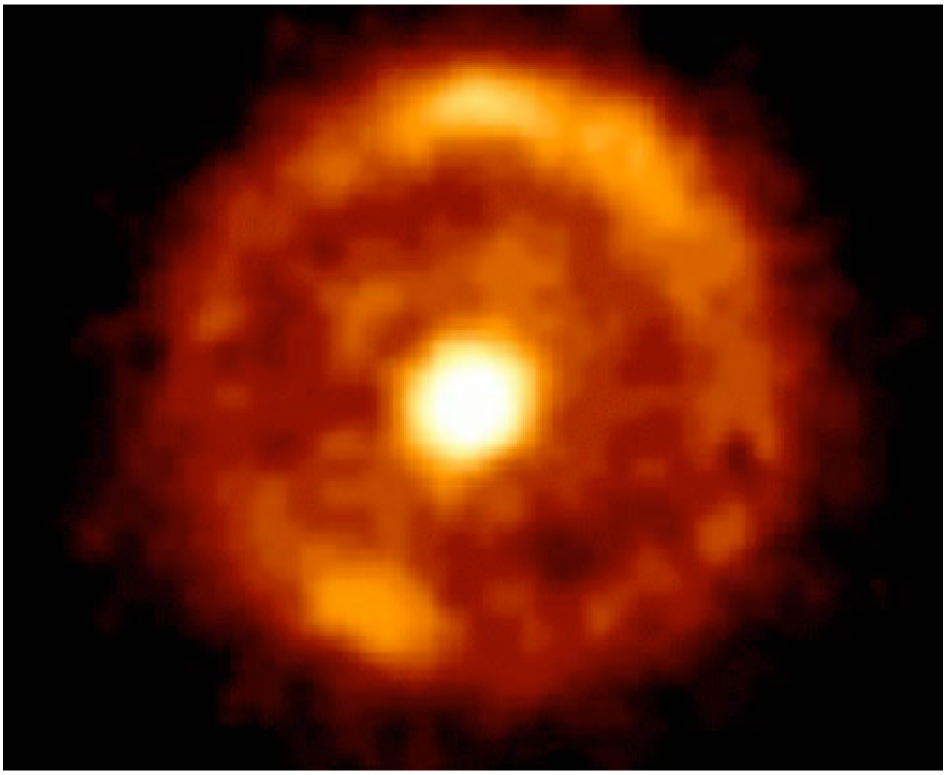

Figure 12.3: Hubble Space Telescope image of an Einstein ring, where a fortuitous alignment has caused an intervening elliptical galaxy (the central spot) to lens a background spiral galaxy into a ring.

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.

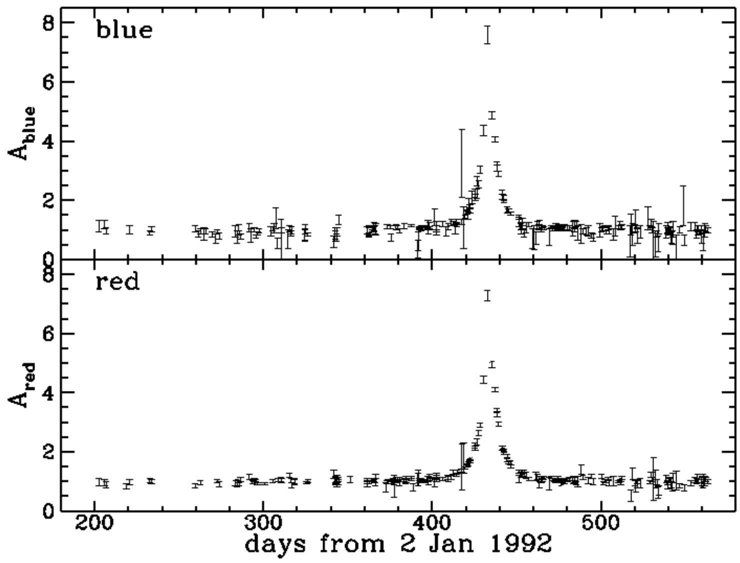

Figure 12.5: Luminosity of a star in the LMC in the blue (top) and red (bottom) as a function of time. Lensing causes it to brighten around day 420.

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/

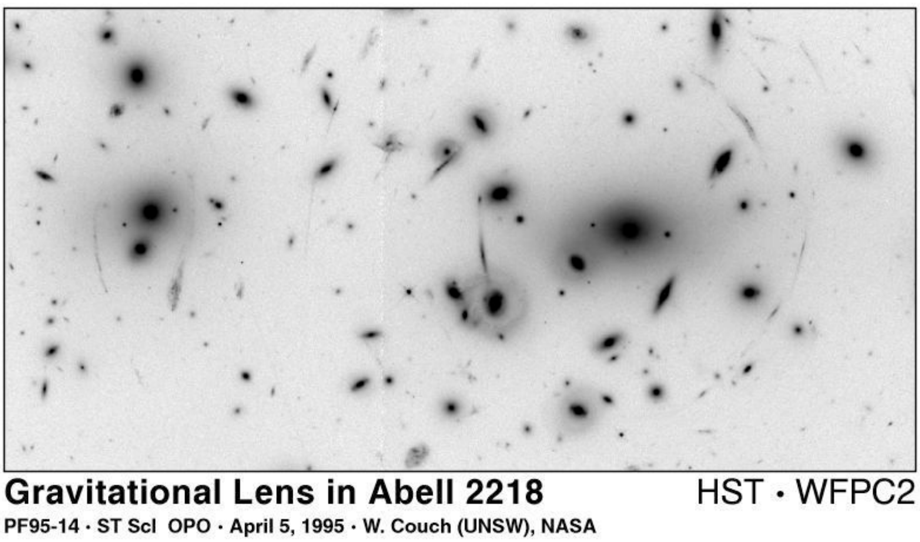

Figure 12.6: Hubble Space Telescope image of the rich cluster of galaxies Abell 2218. The large galaxies in the image are all massive elliptical galaxies in this galaxy cluster. Tens of radial arcs are images of background galaxies that have been lensed and distorted by the dark matter in the cluster.

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