Stars and **Galaxies**

Tom Theuns
Office OCW 207
Institute for Computational Cosmology
Ogden Centre for Fundamental Physics
Durham University
tom.theuns@durham.ac.uk

- ⋆ **Chapter 1: Introduction**
- ⋆ **Chapter 2: The discovery of the Milky Way and of other galaxies**
- ⋆ **Chapter 3: The modern view of the Milky Way**
- ⋆ **Chapter 4: The Interstellar Medium**
- ⋆ **Chapter 5: Dynamics of galactic discs**
- ⋆ **Chapter 6: The dark halo**
- ⋆ **Chapter 7: Elliptical galaxies**
- ⋆ **Chapter 8: Groups and Clusters of galaxies**
- ⋆ **Chapter 9: Galaxy statistics**
- ⋆ **Chapter 10: Active Galactic Nuclei**
- ⋆ **Chapter 11: Gravitational lensing**

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 in 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? What can we learn from them?

11 lectures are too few to answer any of these fundamental questions in 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 learned in other lectures (in particular classical mechanics, electricity and magnetism, quantum mechanics, and of course the earlier parts of this course on 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, growth of structure in the Universe, detection of gravitational waves, study of black holes), investigations whether the fine-structure constant evolves (from quasar absorption spectra), the discovery of dark matter (from the motions of galaxies and from the growth of structure), the discovery of dark energy (from the expansion of the Universe), and even estimating the masses of neutrinos (from the

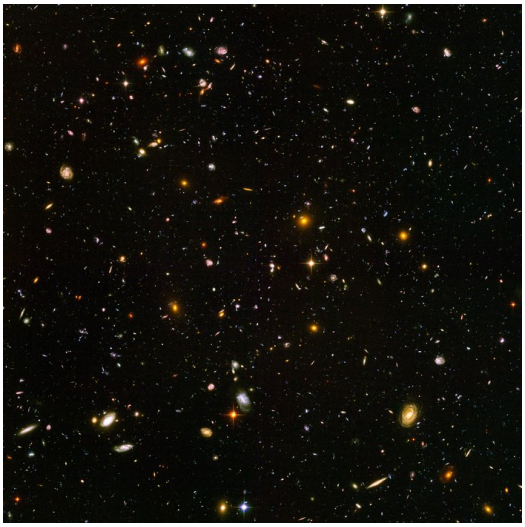

Figure 1: The Hubble Ultra Deep Field

growth of structure). Several of these fundamental discoveries rely on a basic understanding of galaxies, the building blocks of the visible 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 discs, 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/

Much of what is discussed in the notes can be found often with additional explanations and diagrams in several text books. The lectures here will refer very often to Carroll & Ostlie: paragraphs refer to the corresponding section in that reference book (*e.g.* CO 6.2 refers to Section 6.2 of that text book, CO p. 310 to page 310.)

Additional explanations can be found in

1 Carroll & Ostlie, *Modern Astrophysics*, Addison-Wesley, second edition, 2007

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. Chris Mihos maintains an 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.

About your lecturer

I research how galaxies form and evolve using numerical cosmological simulations. The aim of these simulations is to try to understand which are the main physical processes that shape galaxies and how, starting from the nearly uniform Universe that we see in the cosmic microwave background, the different types of galaxies emerge over cosmic time. If you are interested, please take a look at the website of the Eagle simulations. I was one of the main contributors to the simulation code that was used to perform these cosmological hydrodynamical simulations. The simulations themselves were performed on the Curie supercomputer in Paris , part of the European PRACE supercomputing consortium, as well as on the Cosma supercomputer that is part of the Dirac infrastructure, here in Durham.

The Eagle project has resulted in several hundred scientific publications, one of which is the most cited paper in the whole of astronomy in 2015. (I am a little bit proud of that!) The simulation data are available to anyone, and you can play with the results using a database. If you're keen to try this and want some suggestions of what you might want to look into, do not hesitate to ask me. The database has several python examples to get you started.

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 disc. The disc of the Milky Way has a radius of about 15 kpc, and contains $\sim 6 \times 10^{10} M_\odot$ of stars. The disc 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 disc, with mass $\approx 10^{10} M_\odot$. Dust in the interstellar medium of the disc prevents us from seeing the Milky Way's bulge on the night sky. disc 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.

Chapter 7: elliptical galaxies

Elliptical galaxies are roundish objects which appear yellow. The redder colour as compared to spiral galaxies is mostly due to the absence of massive (blue) stars: elliptical galaxies have much lower star formation rates than spirals of the same mass (and massive stars are short lived, so can only be present in a galaxy that is forming stars). We suspect that the low star formation rate is because the central super massive black hole prevents star formation.

The stars move at high velocities but with little ordered motion: 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 (tri-axial in 3 dimension).

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 8-9: 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 10: 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 - starved of fuel.

Chapter 11: 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

CO §25.1 & 25.2

1.1 Historical perspective (*CO §24.1*)

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 its $\sim 200 \times 10^9$ stars lie in a thin ($\sim$ 1/2kpc) disc of radius $\sim$ 15 kpc, with the Sun at a distance $\sim$ 8kpc from the centre. The Milky Way's disc can be seen as a faint trail of light on the night sky, while the other visible stars are not obviously part of a disc 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 M31) and Dryers (1988) *New General Catalogue* (in which it

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

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, and we now know that the observable Universe contains $\sim 10^{11}$ galaxies more massive than the Milky Way. Surveys of galaxies routinely catalogue properties of millions of galaxies, with the planned EUCLID mission aiming at detecting 2 billion galaxies.

1.2 Galaxy classification (*CO §25.1*)

Optical light from normal galaxies is produced predominantly by stars. Luminosities of stars vary widely. The luminosity, L, of a main sequence star depends on its mass, M, approximately as $L \propto M^{\alpha}$, with $\alpha \approx 3$, hence a single $100M_{\odot}$ star outshines $\sim 10^6$ 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 their main sequence progenitors. In conclusion, we expect that the observed luminosity of a galaxy will be dominated by giants and massive main sequence stars.

1.2.1 Galaxy 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 their 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 of screen/projector or printer. 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.

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

- *Flux, F* 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, F additionally depends on the distance to the observer.

- *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 disc type, or

[3]This is only true for nearby galaxies with redshift $z \ll 1$. When z is not $\ll 1$, surface brightness dims with redshift $\propto 1/(1+z)^4$ (see my cosmology lecture notes), but this falls outside the scope of this course on nearby galaxies.

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

	Elliptical	Spiral
Shape	spheroidal	most stars in a disc
Colour	red	blue
Stars	old stars	old and young stars
ISM	little gas or dust	gas and dust
Stellar Dynamics	large random motions (hot)	circular orbits (cold)
Environment	dense (clusters)	low density (groups and field)

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 type 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 disc of spirals galaxies is due to rotation, and the stars in the disc are on nearly circular orbits around the galaxy's centre. Gas collects in Giant Molecular Clouds in the spiral arms of discss, 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 disc 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-disc luminosity decreases, and the spiral arms become more loosely wound.

Some spirals also contain a *bar*, an almost rectangular stellar system in the disc, sticking-out of the bulge. These are designated as SB. So for exam-

ple, 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 a fraction of a kpc to hundreds of kpc. In contrast, spirals tend to be more homogeneous, with masses 10^9–$10^{12} M_\odot$, and disc diameters from 5 to 30 kpc 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 discs 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 disc, 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

[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.

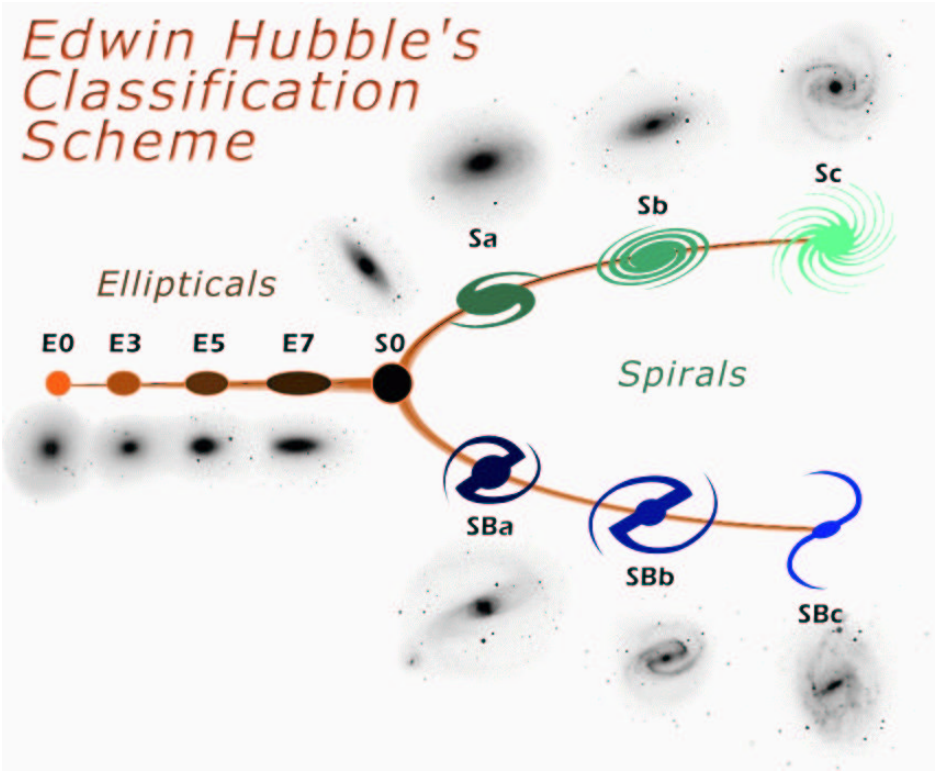

Figure 1.1: The Hubble sequence.

- 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

Chapter 2

The discovery of the Milky Way and of other galaxies

CO §25.1

Before distances to galaxies were measured, their nature and that of the Milky Way itself, was unclear. One suggestion was that nebulae were related to stellar evolution - maybe proto-stellar systems? Observations gave contradictory results, with star counts suggesting the Sun was near the centre of the Milky Way, while globular clusters observations suggested it was not. This chapter discussed the reasons of the confusion and how all was finally resolved.

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. How and where these systems form is not well known, even today.

- *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 (CO p.878)

Star counts can probe the number density of stars in a stellar system, as a function of position. William and Caroline Herschel, brother and sister, counted the numbers of stars as function of their magnitude in ~ 700 directions in the sky. Such counts can tell us about the distribution of stars around the Sun - and hence the shape of the MW galaxy, as follows.

Suppose for simplicity that all stars have the same luminosity, L. The flux of a star, $F = L/4\pi r^2$. Because we assumed L is the same for every star, the flux received from a star only depends on its distance - hence counting the number of stars as a function of flux is equivalent to counting the number of stars as a function of distance.

Let the number density of stars at distance r in direction $\hat{\Omega}$ be $n(r, \hat{\Omega})$. Using the relation between F and r, means that the the number of stars with flux between F and $F + dF$ in that direction, is

$$\begin{aligned} dN(F, \hat{\Omega}) &= n(r, \hat{\Omega}) r^2 dr \, d\Omega \\ &= -\frac{1}{2} n(F, \hat{\Omega}) \, (\frac{L}{4\pi})^{3/2} \, F^{-5/2} \, dF \, d\Omega \, . \end{aligned} \tag{2.1}$$

Therefore counting $dN(F)$ allows one to infer the density $n(r)$. Curiously - the answer the Herschel's found for $n(r, \hat{\Omega})$ was (very) wrong.

Parallax

Nearby stars appear to move with respect to more distant stars as Earth moves around the Sun. The angular 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 'parsec' (pc) derives from **par**allax of one **sec**ond of arc.

Standard candles

Standard candles are objects with a known intrinsic property, for example a known size, or known luminosity. If the luminosity L is know, measuring the flux F yields the distance to the object. If the (physical) size is known, measuring the angular extent yields the distance to the object.

Cepheid variables are an important example. Henrietta Leavitt studied variable stars in the Magellanic Cloud in 1912. She recognised that the flux of some of the variable stars varied with a very characteristic pattern which was periodic in time. She realised that there was a simple relation between the (mean) flux of these stars and this period. Since all these stars are at (nearly) the same distance (namely the distance Earth-LMC), she realised that such 'Cepheid variables' follow a period-luminosity relation. The importance of her discovery for astronomy can hardly be overstated. They are *standard candles* because observationally we can measure P as well as F. The $P - L$ relation yields L from the emasured P, and the combination of L and F yields r, the distance. Parallax measurements to nearby Cepheids are required to calibrate the relation.

Parallax and Cepheid variables are therefore the first two steps in establishing 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. One of the key science goals for the *Hubble Space Telescope* was to detected the periods and fluxes of Cepheid variables in the nearby Virgo cluster (see the later chapter on clusters), in order to get an accurate measurement of the local Hubble constant (even today one of the most accurate measurements of H_0)

2.2 Discovery of the structure of the Milky Way (CO p 875)

Star counts

Jacobus Kapteyn confirmed in the beginning of the 20th century using photographic plates the star count results obtained by peering by eye through a telescope by the Herschels. The conclusion from these counts is that 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 (think of a grey day, when clouds absorb a lot of Sun light). Kapteyn realised this, 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 such 'reddening' yet the amount was too small for dimming due to absorption to strongly affect the interpretation of the star counts.

However the reddening is not due to scattering off atoms, but due to scattering off *dust* (Tyndall scattering). Although this *also* reddens the light (because more blue light is absorbed by dust than red light) the amount of reddening is less than for Rayleigh scattering. Therefore the small amount of reddening detected implies a *large* amount of dimming of distant stars - and hence the interpretation of the star counts was significantly wrong.

Globular clusters

Harlow Shapley estimated the distances to the MW's 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 found 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's and Kapteyn's MW picture.

The conflicting interpretations of data (Kapteyn: small elliptical MW with Sun near the centre, Shapley: very large system of globular clusters, with Sun far from the centre; van Maanen: nebulae are nearby, Slipher: nebulae are far away) culminated in a 'Great Debate' to try to settle the issue once and for all.

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 ~ 0.3 Mpc, much larger than even Shapleys size for the MW. This decisively proved that nebulae were other galaxies outside of the MW, and that the size of the MW is very large (Shapley and Slipher were right!). Hubble went on to discover that the more distant galaxies move away from us, with speed proportional to distance (Hubble's law). It is difficult to overstate the importance of this discovery. In one fell 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 discovered only later, in the 1930s, what was wrong with Kapteyn's interpretation, by studying MW Open Clusters. He assumed the size of Open 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.

[1]The modern value is ~ 0.7Mpc.

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 (CO p.878)

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

$$I(r) = I_0\,\exp(-Ar)\,. \tag{2.3}$$

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 the reader (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

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

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,

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

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

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

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

CO §24.2

3.1 New technologies

3.1.1 Radio astronomy

'Radio-waves' is a generic name for EM radiation with long wavelength λ (somewhat arbitrarily taken to be $\lambda > 1$ m) not susceptible to absorption by interstellar dust. Relevant to galaxies, they can probe the dense, dusty regions where stars form. Earth's atmosphere is nearly transparent for wavelengths 1 cm $< \lambda <$ 10 m. Radio telescopes either consist of a (large) single dish (for example at JODRELL BANK; Ph8 has a small dish on its roof) or an *interferometer* with many inter-connected antennae, for example the *Giant Meter Radio-telescope* (GMRT) in India. Radio-observatories separated by 100-1000s of kilometres sometimes combine their signal to obtain very high angular resolution images[1]. In their wildest dream, astronomers think of building radio-telescopes on the dark side of the Moon (to be shielded from radio signals from Earth) or even in space.

[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 $\theta \approx 5{\times}10^{-8}$.

Just as with optical light, you can think of sources that generate a continuum radio-signal or sources that generate lines (In optical light, the analogy would be a black-body spectrum generating EM-radiation over the whole EM spectrum, as compared to emission or absorption lines created by electronic transitions in atoms or ions - for example the Balmer series in the H I atom). Physical processes that generate radio waves include

1. Roto-vibrational transitions in molecules (lines)
2. Thermal radiation from (cold) dust (continuum)
3. Synchrotron radiation from electrons moving in a magnetic field (continuum)
4. Hyperfine transitions, *e.g.* the 21-cm line in hydrogen (see Ch 4, a line)

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 dust temperature is low, the black-body 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, and AGN discussed in a later chapter.

[2]Recall Wien's displacement law, relating the peak emission wavelength λ and temperature T as $\lambda = b/T$, with $b = 3 \times 10^{-3}$ m K. .

The 21-cm line is discussed in more detail in the next chapter.

Recent new radio-telescopes include LOFAR in the Netherlands, which has stations throughout Europe, the Atacama Large Millimeter Array in Chile (ALMA), and path-finders to the Square Kilometre Array (SKA) such as MEERKAT in South Africa and WALLABY in Australia, as well as the giant FAST telescope in China.

3.1.2 Infrared astronomy

IR photons are not significantly absorbed by dust either. 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 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, the successor to HIPPARCOS, was launched in November 2013 and is currently taking data. Once its mission is completed, it will have measured positions on the sky of billions of stars to astonishing precision. Its measurements are so accurate that it can determine the motion of stars in the plane of the sky, simply by measuring positions as a function of time. It will revolutionize our understanding of the MW. You may want to spend some time exploring the brilliant website of this mission - use the link above.

Fig. 3.1 shows the striking difference in how the MW appears when observed in different wave lengths. This is partly because the source of the EM

radiation is different (e.g. super nova remnants in the X-rays, versus dust in the IR) and partly because some wave lengths are more absorbed than others.

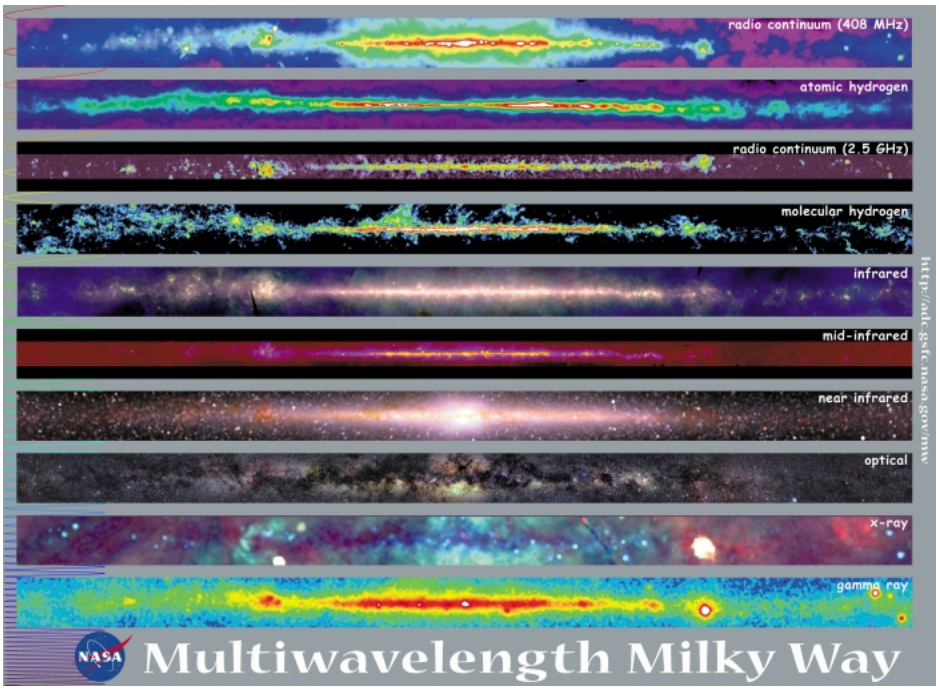

Figure 3.1: NASA's multi-wavelength view of the MW galaxy.

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: *disc*, *bulge* and (stellar) *halo*. A summary of the properties of these is in Table 3.1.

3.2.1 disc

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

- luminosity $L \sim 1.4 \times 10^{10} L_{\odot}$ in the V-band, ≈ 70 per cent of the MWs total V-band luminosity.

- radius[3] $R \sim 15$kpc.
- Define the thickness t of the disc as the ratio $t \equiv \rho/\sigma$ of the volume density of stars, ρ, in 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 disc also contains gas and dust (see lecture 4 on the interstellar medium). On top of the smooth disc are spiral arms, traced by young stars, molecular clouds, and ionised gas. The disc stars are in (nearly) circular motion around the centre, with speeds ~ 220km s^{-1}. The oldest disc White Dwarfs are $\sim 10 - 12 \times 10^9$ years old, but these could have formed *before* the disc.

The density distribution of stars in disc 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 disc, with *scale-length* $R_h \approx 3$ kpc and *scale-height* $z_h \approx 0.3$ kpc.

Note that there isn't really an edge to the disc, 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 disc!

The thick disc About 4 per cent of the MW's stars belong to a thicker disc, aligned with the (thin) disc, but with a larger scale height of $z_h \approx 1$ kpc. Stars in this thick disc differ from the thin disc 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

[3]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 their elemental composition differs from that of disc stars, suggesting that the bulge formed before the disc. Currently, there is little or no star formation in the bulge.

The MW's bulge is not exactly spherical, rather it is ellipsoidal in shape with axis ratio 5:3, and with strong evidence for a bar. Whereas the disc 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 disc galaxy.

Although the bulge is bright, you cannot see it on the night sky, because its visible star light is absorbed by intervening dust. We need IR observations to make it visible.

3.2.3 Stellar halo

A very small fraction (≤ 1 per cent) of the MWs stars are contained in a large, spheroidal, extremely tenuous stellar system called the (stellar) 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 disc, 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 disc stars) races with 220 km s^{-1} around the MW centre, but the halo star does not partake in this rotation. It is thought that the majority of halo stars is debris from the destruction of MW satellites. When such a small satellite galaxy ventures close to the disc, it is destroyed by the galactic tides, and its stars scattered throughout the halo. Evidence for such destruction is seen in the form of 'tidal streams' streams of stars torn out of a tidally disrupting galaxy. See for example this GAIA web-site.

The halo also contains ≈ 150 Globular Clusters (GCs).

The density of halo stars, and of GCs, falls off with distance r from the centre, 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 ≈ 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.

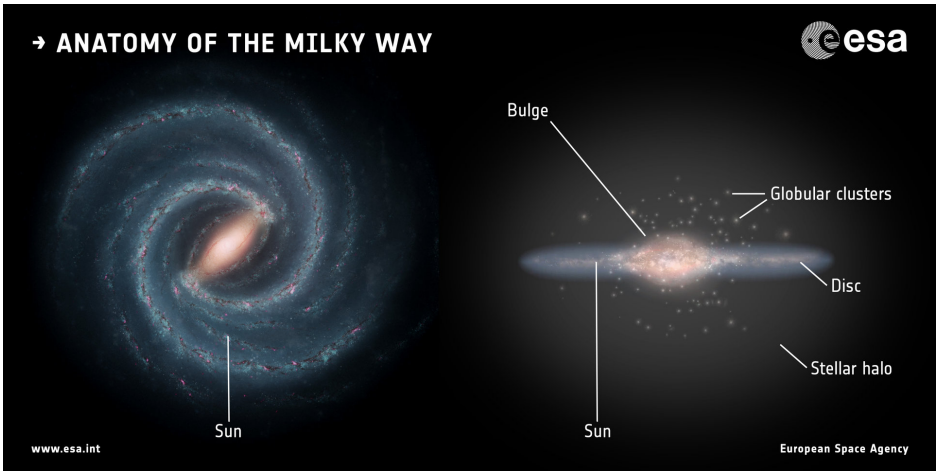

Figure 3.2: ESA's cartoon of the MW galaxy.

3.3 Metallicities of stars and galaxies (CO p.885)

Nucleo-synthesis during the first three minutes after the Big Bang produced mostly Hydrogen and Helium as well as trace amounts of other light elements such as Li and B[4]. By mass, a fraction $X \approx 0.76$ was Hydrogen, with most

[4]See L3 and L4 lectures.

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[5]. These elements are usually (but inaccurately) 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

$$[\mathrm{Fe/H}] \equiv \log_{10} \left[\frac{M_{\mathrm{Fe}}/M_{\mathrm{H}}}{(M_{\mathrm{Fe}}/M_{\mathrm{H}})_{\odot}} \right] , \tag{3.4}$$

that is (the logarithm of) the ratio of Fe-to-H by mass in the object, 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 times more Fe rich.

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

Bulge stars have a wide range of $-3 \leq [\mathrm{Fe/H}] \leq 0.3$, with no significant ongoing star formation. This population of relatively metal rich stars in a

[5]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.

region not undergoing star formation is called *population II*[6].

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*.

Figure 3.3 illustrates the abundance pattern of the Sun.

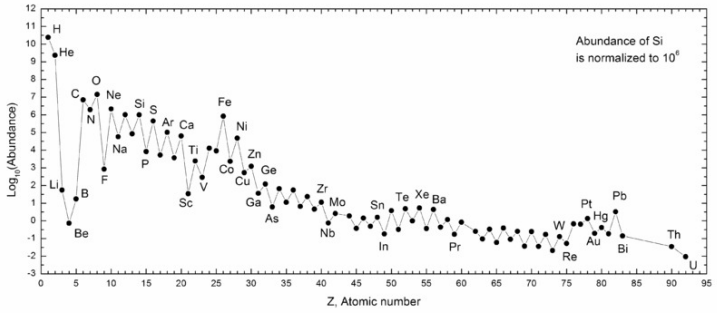

Figure 3.3: Elemental abundances in the Sun reflect the different nucleo-synthetic processes that enriched the gas cloud from which the Sun formed. What is plotted is the relative abundance of various elements, characterized by their atomic number, with respect to Hydrogen. Notice the dominance of H and He - mostly formed during Big Bang nucleo-synthesis, the relatively high fraction of α elements (such as C, O, Ne, Mg etc) whose nucleus is multiple of an α particle (i.e. a He nucleus) synthesized in massive stars, and the general sharp drop in abundance past Fe. Iron is the most bound of all nuclei and fusion of more massive elements does not release energy and hence cannot power stars.

3.3.1 Galactic Coordinates (CO §24.3)

The position of an object on the sky as seen from the Sun can be characterised by two angles (see Fig. 3.4) : (1) *galactic latitude*: the angle b above the Milky

[6]The stellar population of elliptical galaxies is similar

Table 3.1: MW parameters - Binney & Tremain, Binney & Merrifield

Disc scale length, R_d	3 kpc
Disc luminosity	$1.4 \times 10^{10} L_\odot$
Disc mass	$5 \times 10^{10} M_\odot$
Bulge luminosity	$2 \times 10^{9} L_\odot$
Bulge mass	$2 \times 10^{10} M_\odot$

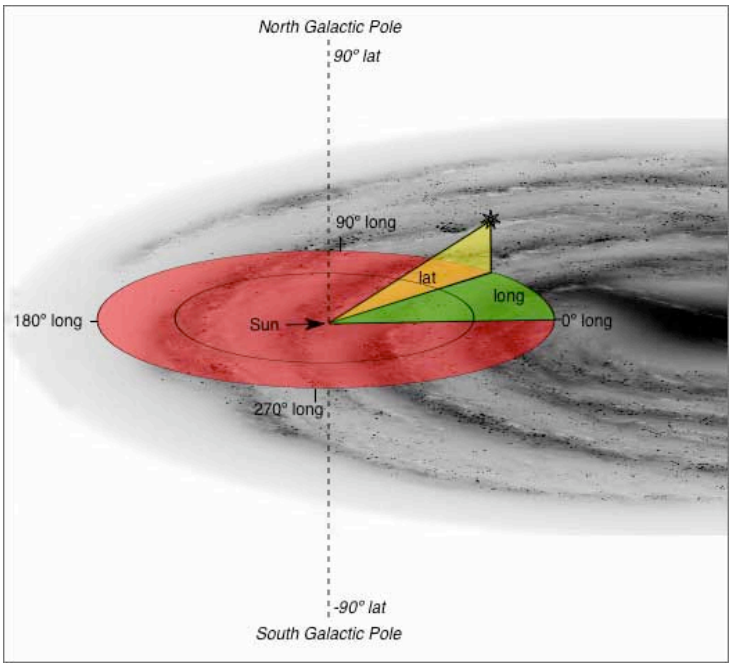

Figure 3.4: Diagram illustrating galactic longitude l and galactic latitude b. From *thinkastronomy.com*.

Way plane and (2) *galactic longitude*: the angle l between the direction Sun-Galactic centre, and the projection Sun-star onto the Milky Way disc.

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 (ISM)

CO §12

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*, GMC's), which form stars that then destroy the clouds, polluting the gas with metals. Stellar winds and super nova explosions stir the gas, and may even expell some gas from the MW into the surrounding circumgalactic medium. Dust, magnetic fields and cosmic rays[1] play an important but poorly understood rôle. This chapter concentrates on the composition of the ISM, explains how the different components can be observed, discusses how stars and gas interact, and concludes with the important concept of Jeans mass, relevant for how stars form in clouds.

4.1 The baryon cycle

Figure 4.1 attempts to illustrate the complex flow of baryons inside the ISM. On the left, some of the ISM gas gets dense enough to form molecular clouds in which stars form. Stellar evolution makes some of these stars lose mass, and at the end of their lives, they may return most of their mass back into the

[1] COSMIC RAYS are energetic particles (photons, electrons, protons, nuclei of heavier elements), with kinetic energies up to orders of magnitude higher than what can presently be achieved in labs such as SLAC or Cern, that bombard Earth. 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.

ISM, for example through a planetary nebula phase (for intermediate mass stars), or through super nova (SN) explosions (for more massive stars). The gas lost by these stars is enriched by their nucleo-synthesis products and may also contain dust. In addition, the SNe inject tremendous amounts of energy into the ISM. How all of this fits together is not terribly well understood.

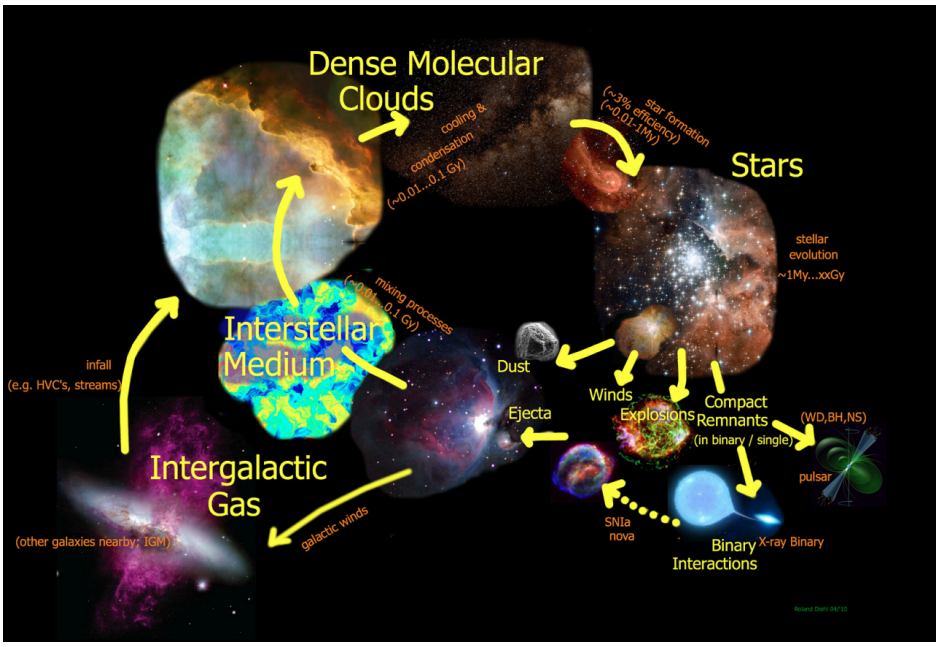

Figure 4.1: Baryon cycle in the ISM, from gas to clouds to stars and back to gas.

4.2 Interstellar dust (CO §12.1)

Interstellar dust consists of small (micron-sized) solid particles, made of C and/or Si, and various ices, as illustrated by Fig. 4.2. The main source of dust is thought to be super novae and AGB stars. The dust grains affect the chemistry of the ISM, its thermodynamics (relation between temperature and density), and also the propagation of light. Quite relevant to life, such tiny particles presumably enable planet formation. Dust particles may *absorb* or *scatter* light, with the scattered light polarized.

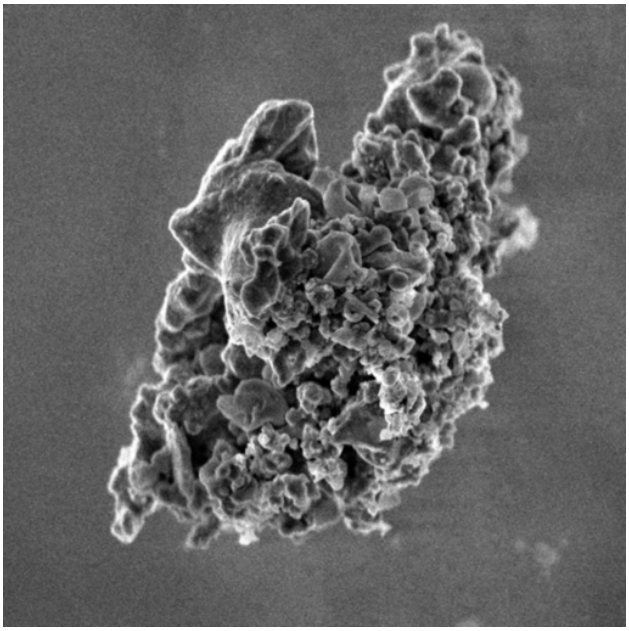

Figure 4.2: A dust grain.

Absorption: a photon may be absorbed by a dust grain, meaning its energy goes into heating the grain. When the dust grain cools down again, it does so by emitting radiation at longer wavelengths. It is thought that nearly 50% of all star light produced in the Universe is reprocessed by dust, converting (mostly) blue and UV light (emitted by stars) into IR light. Notice that the momentum of the photon is absorbed as well, meaning the dust grain gets a little kick when absorbing light. Dust clouds may therefore be accelerated by the radiation pressure exerted by nearby stars. A beautiful example is the 'light erosion' suffered by the so-called PILLARS OF CREATION (the Eagle nebula).

Scattering: a photon may reflect (scatter) off a dust grain, changing its direction but not its energy. The scattered light is polarized.

The amount of scattering and absorption depends on the wavelength of the light (in addition to the properties of the grains), a phenomenon we are familiar with in terms of how the Earth's atmosphere affects Sun light. Sun light gets scattered (mostly by atoms and molecules rather than by dust) with blue light getting scattered more than red light. This makes the sky blue and the Sun appear (slightly) redder, see Fig. 4.3. That scattering

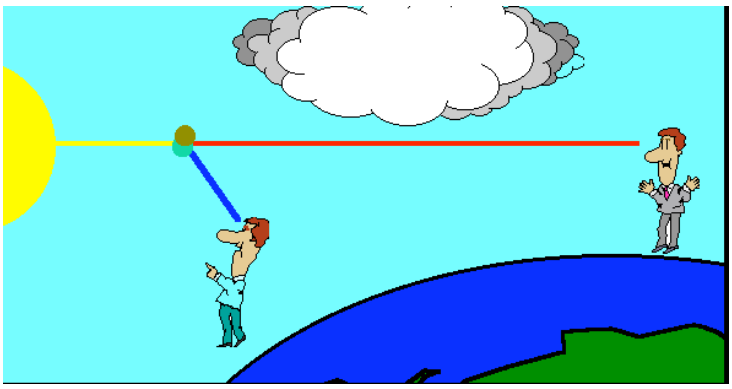

Figure 4.3: Scattering of Sunlight in the Earth's atmosphere, from The Physics and Relativity FAQ.

induces polarization is well known to anybody that has polarized sun glasses to reduce the glare from sunlight scattering off the ocean.

Applied to the ISM: dust will make more distant stars appear *fainter* (because light is absorbed) and *redder* (because blue light is absorbed more strongly). Think back to the dark bands we noticed in images of spiral galaxies: these are dust lanes where the amount of absorption is clearly very large; in the previous lecture we saw how these dust clouds glow in the IR. The large amount of dust in the MWs disc prevents us from seeing the MW's bulge in optical light - but we can detect it in the IR where absorption is much reduced.

Although the amount of scattering/absorption increases with decreasing wavelength λ, on average, some photons have *just* the right wavelength to be in resonance with quantum transitions of the atoms in the dust grain. Such absorption features allow us to determine the composition of grains.

4.3 Interstellar gas (CO §12.1)

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. 4.4.

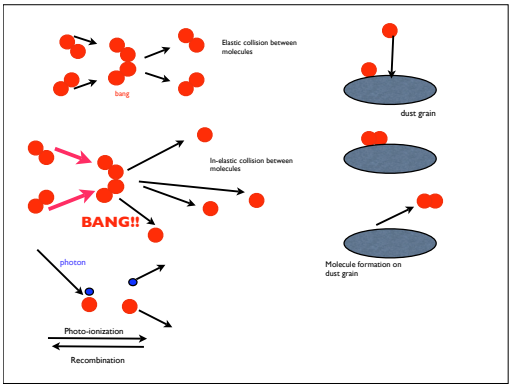

Figure 4.4: Common physical and chemical processes in the ISM.

Notation Astronomers use the (confusing) notation X III to denote the *doubly* ionised state of element X. For example, C III≡C^{2+}, H I is neutral Hydrogen, H II ionised Hydrogen, and O VI is five-times ionised O.

4.3.1 Collisional processes

Reminder L1 course, Kinetic theory of gases, Young & Freedman §18.3 The typical velocity v of particles (atoms, say) with mass m in a gas with temperature T is $v^2 \sim 3kT/m$. A collision between such particles will have typical kinetic energy $E \sim \frac{1}{2}mv^2 \sim 3kT/2$. The velocity distribution of such particles is a Maxwell-Boltzmann distribution, $\mathcal{P}(v) \propto \exp(-mv^2/(2kT))$.

To understand the physics of particle-particle collisions in a gas, let us start with cold molecular gas, which for simplicity consists of H_2 molecules only, and let's concentrate on collisions with electrons, e.

A diatomic molecule such as H_2 has 'internal degrees of freedom'(doF), meaning energy can be stored in rotation (2 doF) and vibration (1 doF, but with twice the amount of energy), in addition to the 3 doF associated with the velocity of the molecule. As with all energy levels in quantum mechanics, the energy of roto-vibrational doF are quantised, meaning you need to transfer sufficient energy to the molecule to make it spin, and even more to make it vibrate.

Consider an electron colliding with H_2 at low speed (hence low T). It cannot excite a rotation in the molecule because the collisions energy is too low to even excite the lowest rotational energy level. Hence the kinetic energy of the collision cannot change, and the collision is *elastic*. The diatomic molecule effectively acts as a mono-atomic molecule, with heat capacity at constant volume, $C_V = 3R/2$. Higher speed electrons (higher T) *can* excite rotations, the collision is no longer elastic, and $C_V = 5R/2$. At even higher T, also vibrations can be excited, and $C_V = 7R/2$. Here, R is the gas constant, see Y&F §18.4, and in particular Fig. 18.19. At even higher T, things get even more interesting.

If kT is sufficiently high, a collision may break the molecular bond,

$$e + H_2 \rightarrow e + 2\,H\,.$$

We can estimate the required temperature by comparing kT to the binding energy $E_{H_2} \approx 4.5$ eV of H_2, $kT \approx 4.5\text{eV} \rightarrow T \approx 5 \times 10^4\text{K}$. This is nearly an order of magnitude higher than measured! The reason is that electrons have a Gaussian distribution of energies at a given T, and the higher energy electrons in the tail of the Gaussian *can* destroy the bond, even if the *average* energy electron can not.

At even higher $T \approx 10^4\text{K}$, a collision may ionise the hydrogen atom,

$$e + \text{H I} \rightarrow 2e + \text{H II}\,.$$

In summary, we expect *cold gas to be molecular*, *warm gas to be atomic*, and *hot gas to be ionised*.

Finally consider regions in the ISM with different temperatures, for example a warm, neutral gas cloud embedded in hot, ionised gas. The pressure in these gases should be comparable. Indeed, suppose the pressure in the cloud were lower, then the cloud will be compressed, increasing its pressure, until the cloud is in pressure balance. Our previous reasoning then further implies that warm clouds are neutral *and dense*, and hot gas is ionised and *tenuous* (of low density). This is indeed what we see in the ISM, were cold, dense molecular clouds are embedded in warmer, neutral gas, itself embedded in hot, ionised and tenuous gas. However there is one more mechanism that is important: photo-ionisation.

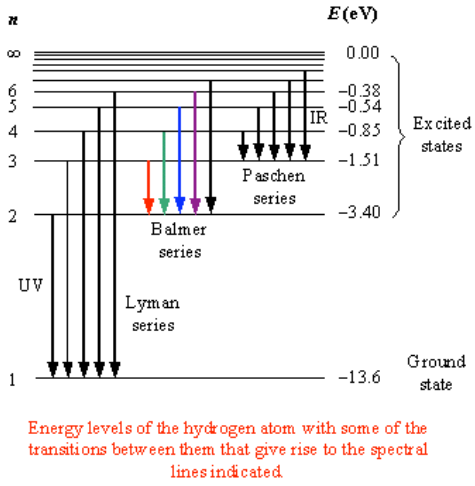

Figure 4.5: Electronic transitions in the Hydrogen atom, and nomenclature of the corresponding emission lines. Figure taken from the internet encyclopledia.

4.3.2 Photo-ionisation and HII regions (CO p.431)

A photon impinging on an atom may ionise the atom in a process called *photo-ionisation*, for example for H I:

$$\gamma + \mathrm{H\ I} \rightarrow e + \mathrm{H\ II}\,.$$

The photon energy has to be higher than $h\nu = 13.6$eV, where 13.6 eV is the binding energy of H I. The wave-length of such a photon is $\lambda = hc/(13.6\ \mathrm{eV}) \leq 911$ Å. Such high-energy photons are only emitted in large numbers by hot stars, in particular by massive MS-stars.

In an ionised gas (a plasma), the reaction can also occur from right to left, when a proton catches an electron, leading to the emission of one or more photons: this is called *recombination*. Recombining hydrogen gas emits a set of characteristic lines, associated with the energy levels of H I, shown in Fig. 4.5. Recombinations in star forming regions cause the characteristic red glow of the H I $n = 3 \rightarrow 2$ Hα line of the Balmer series[2]

Recall that the massive stars that emit ionising radiation are short-lived. The detection of Hα emission from a galaxy is therefore a good handle of its

[2]The other transitions are also detected, but Hα is particularly strong.

star formation rate. Go back to the images of spiral galaxies from chapter I: you can see how many reddish star forming regions follow the spiral structure in disc galaxies, as beads on a string. It is the gas that surrounds the massive stars that produces this recombination radiation, with the image revealing the associated Hα radiation. We describe the ionisation structure of such gas clouds, called H II regions, next.

4.3.3 H II regions and Strömgren spheres

Suppose a hot star forms at time $t = 0$, emitting ionising radiation at a constant rate, $\dot{N}_\gamma$ (in ionising photons per second). Suppose further that the star is at the centre of a spherical cloud, initially atomic, with uniform hydrogen number density n (in hydrogen atoms per unit volume).

An *ionization* front will run into the cloud, with gas at distance $r \leq R$ being ionised. The radius of the front at time t, $R(t)$, follows from requiring each of the $(4\pi/3)R^3 n$ hydrogen atoms (the number of atoms in a sphere of radius R) has interacted with a photon. Since the number of photons emitted in time t is $\dot{N}_\gamma t$, this results in

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

The speed of the front follows from taking the derivative of this equation (taking into account that $\dot{N}_\gamma$ and n are both constant)

$$\dot{R}(t) = \frac{\dot{N}_\gamma}{4\pi R^2\, n} \,.$$

The speed of the front slows down as R increases. Notice also that $\dot{R} \to \infty$ for $R \to 0$: clearly this can't be right: the front can't move faster than c. The fact that is does is a limitation of our model.

When the gas is ionised, it will also recombine as discussed in the previous section. The *recombination rate* - the rate at which the gas recombines per unit volume producing H I, is of the form

$$\frac{\mathrm{d}n_{\mathrm{H\ I}}}{\mathrm{d}t} = \alpha\, n_{\mathrm{H\ II}}\, n_e \approx \alpha n^2 \,.$$

The recombination coefficient, α, is an atomic constant; the last part of the equation assumes that inside the H II region, the gas is very high ionised,

so that[3] $n_{\mathrm{H\ II}} \approx n$. The larger the ionised volume, the higher the total recombination rate in the gas. Eventually this causes the ionisation front to stall (stop increasing), because each recombination consumes a photon, leaving no ionising photons to ionise H I for the first time. This limiting radius is called the *Strömgren* radius, R_S, and its value can be found by equating the total recombination rate within R_S, to the rate $\dot{N}_\gamma$ at which the star emits ionising radiation:

$$\frac{4\pi}{3}\,\alpha\, n^2\, R_s^3 = \dot{N}_\gamma\,.$$

Using typical values of $n = 5 \times 10^3 \mathrm{cm}^{-3}$ for the density of a cloud, $\dot{N}_\gamma = 10^{49}\mathrm{s}^{-1}$ 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.

4.3.4 21-cm radiation (CO p. 405)

The electronic transitions of, for example, the Lyman and Balmer series, discussed earlier, correspond to transitions of the electron between different *orbital* energy levels. Recall that the energy level in a hydrogen atom only depends on the value of n, the *principle quantum number*, and not on l or m (which characterise the total angular momentum L, and L_z the angular momentum along the z-axis, respectively.)

However, both proton and electron have another purely quantum mechanical (QM) property called *spin*. The spin[4] of an elementary particle has some properties in common with a magnetic moment, and the energy of an electron in a hydrogen atom will be higher (less bound) if the electron and proton spins are parallel, and lower (more strongly bound) if they are anti-parallel[5] The curious rules of QM mean that these spins in fact can only be either parallel or anti-parallel.

Therefore, a hydrogen atom in which electron and proton spins are parallel, is very slightly less less bound than if the spins were anti-parallel. When

[3]Note that the electron and proton densities are the same, since the cloud was initially fully atomic.

[4]Spin does not feature in Schrödinger's model of the hydrogen atom, but it does appear in Dirac's version.

[5]Just as with bar magnets, who prefer to be anti-parallel rather than parallel.

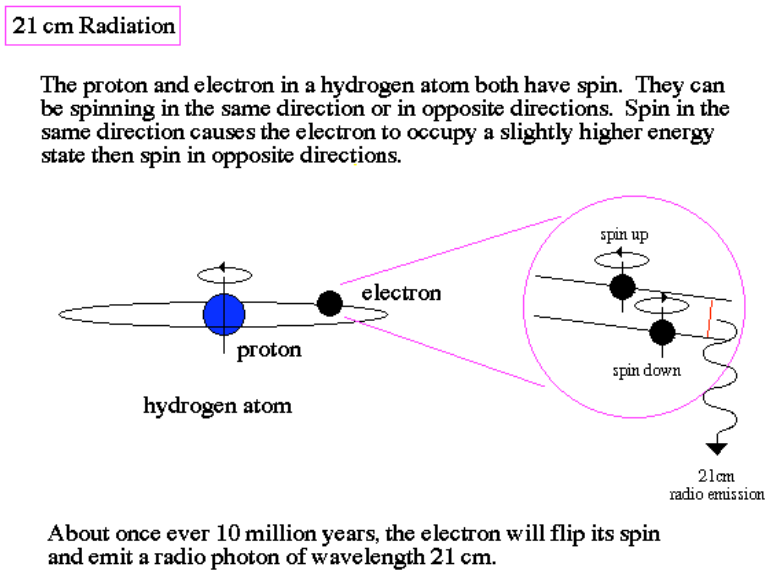

Figure 4.6: Emission of 21-cm radiation in Hydrogen due to a hyperfine transition, taken from Schombert.

a hydrogen atom flips from parallel to anti-parallel states, the energy difference is carried away by a photon with wavelength $\lambda \approx 21$ cm - this is the 'hydrogen 21-cm line', or *hyperfine* line, illustrated in Fig. 4.6. The warm H I gas in the MW, and other galaxies, can be observed in this transition. This line is a great probe of gas in the ISM. Notice that, given its long wavelength, it's propagation is not affected by dust[6].

4.3.5 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.

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

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.

4.3.6 The Jeans mass (CO p. 412)

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

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

$$\begin{aligned} K &= Mu \\ u &= \frac{3kT}{2\,\mu \mathrm{m_H}}\,, \end{aligned} \tag{4.1}$$

where k is Boltzmann's constant, and $\mu \mathrm{m_H}$ the mean molecular weight per particle. The binding energy U of the sphere is

$$\begin{aligned} U &= \frac{3}{5}\frac{GM^2}{R} \\ M &= \frac{4\pi}{3}\rho\, R^3\,. \end{aligned} \tag{4.2}$$

In virial equilibrium, $2K = U$, and the mass of the sphere is the Jeans mass[7]:

$$M_J = \left(\frac{5\mathrm{k_B}T}{\mu \mathrm{m_H} G}\right)^{3/2} \left(\frac{3}{4\pi\rho}\right)^{1/2}\,. \tag{4.3}$$

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

In clouds more massive than the Jeans mass, gravity overpowers pressure, and such clouds will collapse. In clouds less massive than M_J, pressure overpower gravity, and such clouds are not susceptible to collapse.

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*, $\rho \propto T^{3/2}$, then $M_J \propto \rho^{1/2}$, and the Jeans mass will increase as the cloud collapses. Therefore, if $M = M_J$ initially, then as the cloud collapses, the mass will becomes smaller than the Jeans mass, and the cloud will expand again (basically the restoring pressure increases faster than gravity).

However, consider now a cloud that behaves isothermally, because the gas cools as it collapses so that its temperature remains the same rather than increasing as it gets compressed. Then, as the cloud collapses and ρ increases, M_J *decreases*: smaller clouds, that were initially stable because they had $M < M_J$, now can becomes unstable (because M_J decreased): the cloud may *fragment*.

4.4 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 discs

CO §24.3

The stars in the Milky Way disc are on (almost) circular orbits, with gravity balancing centripetal acceleration. Given that most of the light of the disc comes from its central parts, we would expect the circular velocity in the outer parts of the disc 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 (CO p. 917)

5.1.1 Keplerian rotation

The velocity of a test mass m in circular motion around a point mass M ($M \gg m$) at distance R is

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

Applied to planets in the solar system, we find that the period increases as

$$P = \frac{2\pi R}{V} = \frac{2\pi R^{3/2}}{(GM)^{1/2}} \,.$$

The mass M need not be a point mass: Newton's theorem guarantees this equation is also correct for an extended spherical mass distribution, provided we use the mass $M(< R)$ enclosed by the orbit; any mass outside R does not contribute to the gravitational force:

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

To describe the motion of stars in the MW, we would like to apply this equation to a disc, with $M(< R)$ the disc mass enclosed by the orbit. However, Newton's theorem does not apply there, since obviously a disc is not spherical. Fortunately, the error is not very large.

So applying this to the MW's disc, we observe the following. Most of the *light* in the MWs disc+bulge is interior to the Sun's orbit. If this means that also most of the *mass* is enclosed, than in the previous equation $M(< R)$ remains constant for $R \geq R_\odot$, where $R_\odot$ is the distance Sun-MW centre (because (almost) all of the mass is interior to $R_\odot$). In that case, we expect that $V^2 \propto R^{-1}$ (for $R \geq R_\odot$), just as we found for planets: Newton's law, applied to the motion of stars in the outskirts of the disc, predicts that rotation speed falls with distance as $V \propto 1/R^{-1/2}$ - just as is the case of planets in the Solar system. We can test this assertion by studying the motion of stars in the solar neighbourhood, by measuring 'Oort's constants'.

The curve $V(R)$ (rotation speed as a function of distance to the centre) is called the **rotation curve** of the galaxy.

5.1.2 Oort's constants (CO p. 908-913)

Assume all disc stars are on circular orbits, with circular velocity $V(R)$ for a star at distance R from the MW centre. An observer moves on a circular orbit with radius R_0 and circular speed $V_0 \equiv V(R = R_0)$. They measure the line-of-sight speed, V_r, and the tangential speed, V_t, for a star at distance d with galactic coordinates $(b = 0, l)$ (see Fig.5.1). The orbit of the star has radius R, and the star moves with circular speed $V(r)$.

The rotation curve $V(R)$ can be inferred by measuring $V_r(d, l)$ and $V_t(r, l)$ as follow. First, use trigonometry to show that

$$\begin{aligned} V_r &= V\cos(\alpha) - V_0\sin(l) \\ V_t &= V\sin(\alpha) - V_0\cos(l)\,. \end{aligned} \qquad (5.2)$$

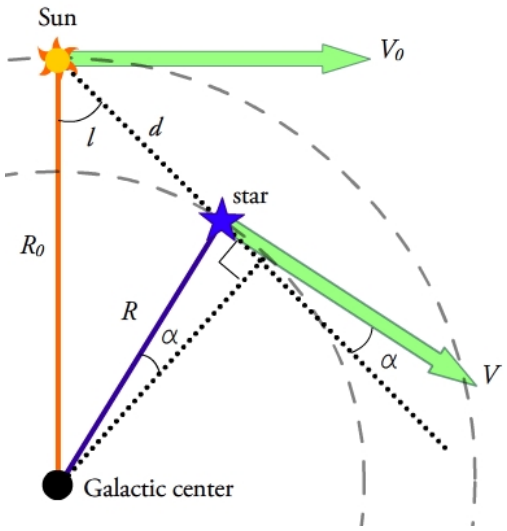

Figure 5.1: (Taken from wikipedia) The observer at the Sun is moving on a circular orbit with velocity V_0 and radius R_0. The observed star is at distance d and has galactic longitude l (with $b = 0$ since it is in the disc). The star is on a circular orbit with radius R and speed V.

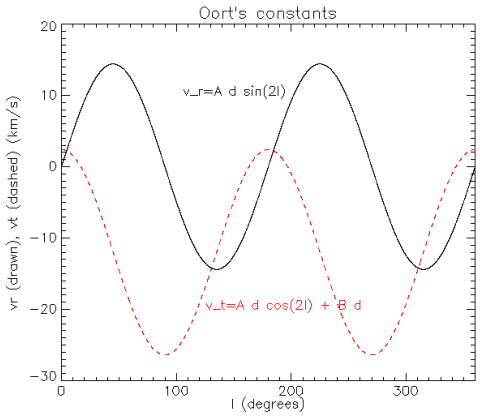

Figure 5.2: Line-of-sight velocity, V_r, and tangential velocity, V_t, of a star at distance $d = 1$ kpc, as a function of its galactic longitude, l.

Again using trigonometry 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. The last step assumes that $d \ll R_0$ so that the angle $\beta \approx 0$: we restrict the analysis to *nearby* stars[1]. A little algebra yields

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

where $\Omega_0 \equiv V_0/R_0$ is the *angular velocity* of the Sun, and $\Omega \equiv V/R$ the angular velocity of the star. For *nearby stars*, we can expand $\Omega(R)$ in Taylor series around $R = R_0$, keeping only the first terms,

$$\Omega(R) \approx \Omega(R_0) + \frac{d\Omega}{dR}|_{R=R_0}(R - R_0)\,.$$

Notice that

$$\frac{d\Omega}{dR} = \frac{d}{dR}\left(\frac{V}{R}\right) = \frac{1}{R}\frac{dV}{dR} - \frac{V}{R^2}\,.$$

Now define **Oort's constants** A **and** B by

$$\begin{aligned} A &\equiv -\frac{1}{2}\left[\frac{dV}{dR}|_{R=R_0} - \frac{V_0}{R_0}\right] \approx 14.4 \pm 1.2 \mathrm{km\,s^{-1}\,kpc^{-1}} \\ B &\equiv -\frac{1}{2}\left[\frac{dV}{dR}|_{R=R_0} + \frac{V_0}{R_0}\right] \approx -12.0 \pm 2.8 \mathrm{km\,s^{-1}\,kpc^{-1}}\,. \end{aligned} \tag{5.5}$$

Some juggling yields

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

[1]Notice that this is not true in the figure!

and the results are plotted in Fig. 5.2. 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[2] 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 $B = -12.0 \pm 2.8 \mathrm{km\,s^{-1}\,kpc^{-1}}$.

What do we expect for these constants, if light traces mass in the MW disc? For such a 'Keplerian disc', $V \equiv V_0(R_0/R)^{1/2}$, hence $dV/dR = -(1/2)V/R$. Therefore we expect

$$\begin{aligned} A_{\rm Kepler} &= \frac{3}{4}\frac{V_0}{R_0} \approx 19.4 \mathrm{km\ s^{-1}\,kpc^{-1}} \\ B_{\rm Kepler} &= -\frac{1}{4}\frac{V_0}{R_0} \approx -6.5 \mathrm{km\ s^{-1}\,kpc^{-1}}\,, \end{aligned} \tag{5.7}$$

where the numerical values use $V_0 \approx 220 \mathrm{km\ s^{-1}}$ and $R_0 \approx 8$kpc, as measured for the Sun.

Our expected values for A and B are clearly inconsistent with the values measured by Oort. In particular, let's compare the values of expected and measured gradient,

$$\frac{dV}{dR} = -(A + B)$$

which are

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

As expected, the Keplerian circular velocity drops $\propto R^{-1/2}$ and therefore $dV/dR < 0$. However, the *measured value is consistent with zero*: **the Milky Way's rotation curve is flat, *i.e.* $V(R) \approx$ 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'[3]: the speed of the Sun is ~ 16 km s^{-1} with

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

[3]Note this is not an inertial system

Figure 5.3: Prof Vera Rubin was pivotal in establishing that the rotation curves of spiral galaxies are flat in their outskirts, thereby unambiguously demonstrating that galaxies are dominated by dark matter.

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 disc, 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 gas from its Doppler shift, and hence the rotation curve of the gas disc[4]. To see how, use Eqs. (5.2) to show that V_r has a *maximum*

$$V_{r,\max} = V - V_0 \sin(l)\,, \tag{5.9}$$

(which occurs for $\alpha = 0$ and for which $d = R_0 \cos(l)$ and $R = R_0 \sin(l)$). Therefore

$$\frac{dV_{r,\max}(l)}{dl} = \frac{dV(R)}{dR}\frac{dR}{dl} - V_0\,\cos(l)\,, \tag{5.10}$$

which can be simplified, using $R = R_0 \sin(l)$, to

$$\frac{dV(R)}{dR} = \frac{dV_{r,\max}(l)}{dl}\,/(R_0\,\cos(l)) + V_0/R_0\,. \tag{5.11}$$

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

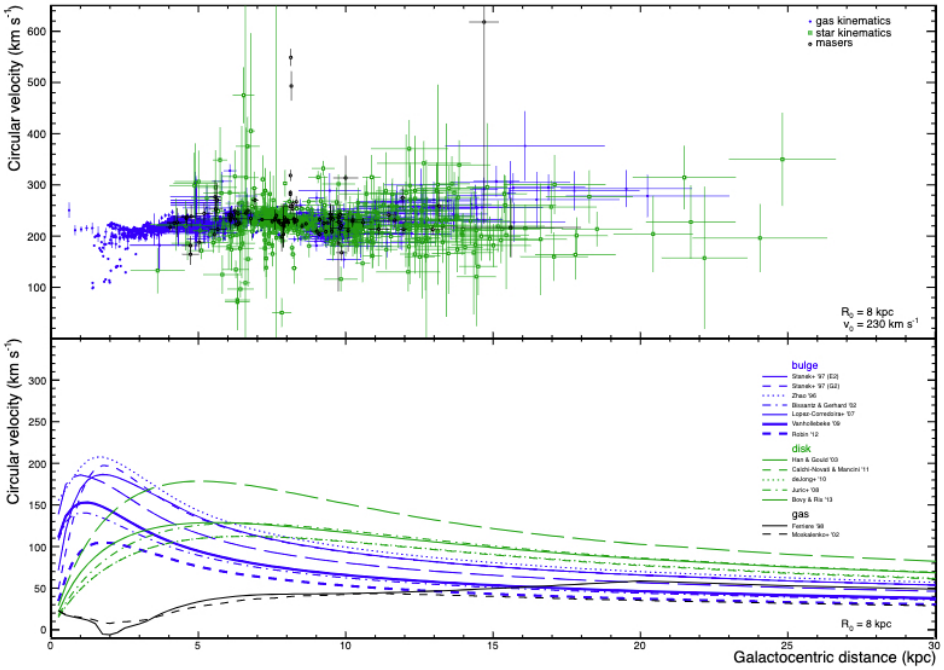

Figure 5.4: Compilation of data (top) and models (bottom) for the Milky Way's rotation curve. Fig 1 from Iocco et al., arXiv:1502.03821

So we don't need to measure the distance to the H I clouds, simply measure their maximum speed as a function of l. 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 now available for tens of spirals, and they all show flat rotation curves in their outskirts; the MW is definitely not unusual in this respect. Prof Vera Rubin, Fig. 5.3, pioneered such measurements. Figure 5.4 shows a recent compilation of measurements of the MW's rotation curve.

5.2 Rotation curves and dark matter (CO p. 914)

A flat rotation curve, $V \sim$ const, rather than the Keplerian expectation, $V \propto R^{-1/2}$, implies that the MW's 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[5] with respect to R of

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

for V is constant:

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

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.

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[6] in nature.

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

[6] 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.

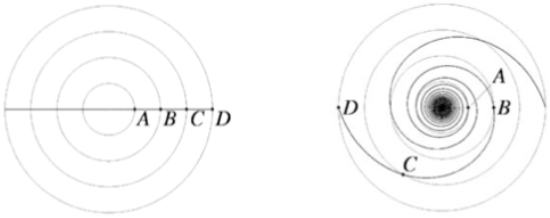

Figure 5.5: A spiral pattern made out of stars will rapidly wind-up in a disc undergoing differential rotation.

5.3 Spiral arms (CO §25.3)

There is quite a variety in spiral patterns. For example, contrast the relatively poorly defined pattern in a *flocculent* spiral galaxy to the well-defined structure of a *grand design* spiral galaxy. Maybe the physical processes that shape these are different?

However, we first need to discuss a problem. Consider two stars, A and B, which are in the same spiral arm but at distances R and $2R$ from the centre, respectively. Assume further that the galaxy has a flat rotation curve. Since the stars move with the same speed (flat rotation curve), the period of the orbit of star B is twice that of star A, $P_B = 2P_A$: after half a period of B, A will have completed a full period. As a consequence of this 'differential rotation', the spiral arm winds up. This means that we expect spiral arms to be very tightly wound - but that is not what we see. This is called the *winding problem*, see the illustration of Fig. 5.5.

The resolution of the paradox is the realisation that spiral arms are **not** material structures (meaning stars remain in spiral arms all the time). Rather, they are a pattern with stars (and gas) moving in and out of spiral arms. The pattern spins with some angular speed, which may be different from the angular speed of the stars. The *density wave* theory of spiral arms states that, when a disc of stars and gas is disturbed, a spiral pattern is induced. To make an analogy: if you disturb a guitar string by plucking it, a series of standing waves is induced. Similarly, disturb a galactic disc, and a spiral pattern is excited. A flocculent pattern may result from a cooling instability in the gas, a grand design pattern from a tidal perturbation caused by another galaxy passing close by.

Figure 5.6: The spiral pattern in NGC 1566, rotating clock wise.. The dark clouds are the location where gas enters the arm from the inside, gets compressed and starts making the stars. By the time many stars have formed, the cluster has overtaken the arm, and we see the shiny new clusters on the 'outside'of the arm.

If the spiral pattern spins with constant angular velocity Ω, say, the the tangential speed of the arm at distance R is $V_t = \Omega R$, which clearly increases with R. This implies there is a critical radius R_c, given by $\Omega R_c = V$, such that for $R < R_c$, stars overtake the spiral pattern (because $V > V_t$), and for $R > R_c$, the spiral pattern overtakes stars (because $V < V_t$.) The consequence of this can be seen in Fig. 5.6, where the spiral pattern spins clockwise, and we are looking at the region where $R < R_c$. Gas enters the spiral arms from the inside, gets compressed and makes stars in the dark clouds. It takes a while to make these stars, by which time the gas has overtaken the spiral arm. This is why we see the shiny new clusters on the outside of the arm. (At larger R, it would be the other way around. In real galaxies, V tends to decrease close to the centre: this means that there is another, smaller co-rotation radius - where $V = V_t$- closer in.)

5.4 Summary

After having studied this lecture, you should be able to

- Derive the rotation curve for a Keplerian disc
- Derive the equations for the radial and tangential velocity of stars on circular orbits in a disc in differential rotation, and derive expressions for Oort's constants.
- Compute Oort's constants A and B for a Keplerian disc. 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.
- 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 Matter Halo

CO p. 896-897

Measurements of the rotation curve using HI 21-cm emission, analysis of the motions of stars in the solar neighbourhood with Oort's constants, and the Oort limit[1], 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[2] 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 requires that the speed of the star, $v_\star$, is lower than the local escape speed[3].

[1]Not discussed in detail.

[2]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 disc, that have acquired their high velocity following a super nova explosion. For a recent discussion based on GAIA, see Deason et al, '20

[3]The escape speed in a given potential, is the minimum speed a particles needs to have to be able to escape to infinity.

6.1.1 Point mass model

For a *point mass model* (all the mass in the centre), it is easy to find the relation between escape speed, v_e, and circular speed, V_c. For such a model, the circular speed at radius $R_\odot$ is $V_c^2 = GM/R_\odot$, where M is the mass of the MW[4]. The gravitational potential is $\Phi = -GM/R_\odot = -V_c^2$. A star moving with the *escape speed* has zero specific energy[5],

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

Therefore $v_e = 2^{1/2}\, V_c \approx 311\text{km s}^{-1}$ (Using $V_c = 220\text{km s}^{-1}$.) So for a point mass model of the MW, most high velocity stars are not bound to the MW. This analysis also shows that we cannot resolve the discrepancy by simply increasing M. Indeed, although increasing M would increase v_e - it would also increase V_c - yet V_c is *measured*. The only way to increase v_e but not V_c is by changing the *mass distribution* - as we show below.

6.1.2 Dark matter halo model

Given the failure of the point mass MW model, let's assume there to be a dark matter halo, which is spherically symmetric (to make the calculations easy). Let's further assume that the MW's rotation curve is flat, $V_c \approx$ constant, out to some radius R_h - the edge of the halo. In this case, given that $V_c^2 = GM/R$ is constant out to R_h,

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

The gradient of the gravitational potential is the force per unit mass, which is V_c^2/R for $R \leq R_h$, therefore

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

[4]To compute the escape speed at the location of the Sun, we will take $R_\odot$ the distance of the Sun to the centre of the MW, $R_\odot \approx 8$ kpc.

[5]*Specific energy* is energy per unit mass.

Integrating this equation between $R \leq R_h$ and R_h yields $\Phi(R) = \text{constant} - V_c^2 \ln(R_h/R)$. We can determine the value of the constant at $R = R_h$, since then $\Phi(R = R_h) = -G\,M/R_h = -V_c^2$. Hence

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

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

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

For the Sun, $R \approx 8.5$kpc, $V_c = 220\text{km s}^{-1}$, $v_e \geq 500\text{km s}^{-1}$ requires $R_h \geq 40$kpc corresponding to a dark matter halo mass of at least

$$M(R = R_h) \geq 4.4 \times 10^{11} M_\odot\,. \tag{6.6}$$

Even this lower limit to the mass is significantly higher than the MW's stellar mass of $M_\star \approx 7 \times 10^{10} M_\odot$ from Chapter 3. A recent application of this method put $M_h \approx 10^{12} M_\odot$, see Deason et al, '20. An independent measure of the MW's halo mass is based on the motion of the Andromeda galaxy in the *Local Group*.

6.2 The Local Group (CO p. 1059-1060)

The MW is located in a rather average part of the Universe, away from any dense concentrations of galaxies[6]. The 'Local Group' consist of the MW, Andromeda (M31), and a few hundred small, irregular galaxies, all gravitationally bound to each other.

6.2.1 Galaxy population

The MW is orbited by ~ 10 'classical dwarf' satellites, which include, for example, the *Large* and *Small Magellanic Clouds*. These satellites are gravitationally bound to the MW and orbit inside its dark matter halo. The advent of digital sky surveys has resulted in an explosion in the discovery of much fainter dwarf galaxies, also gravitationally bound to the MW, see for example this recent CalTech review. The tally of these ultra-faint galaxies now stands at ~ 100, with likely many more to be discovered.

[6]Such dense concentrations are called *clusters*, see later chapters.

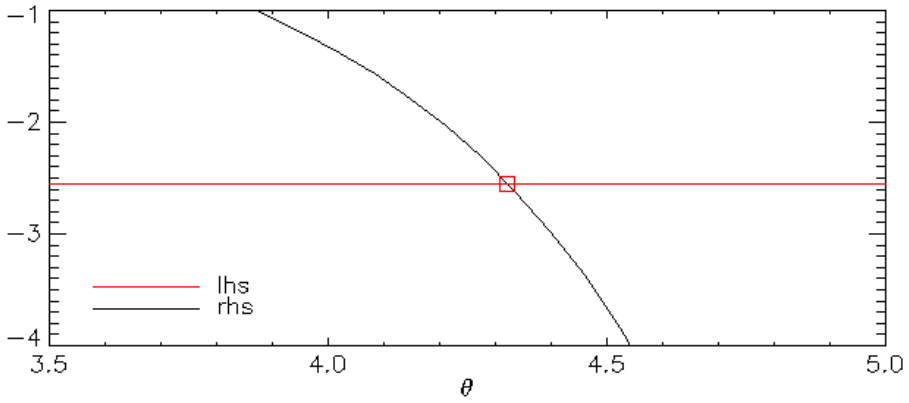

Figure 6.1: The left-hand-side of Eq. 6.11 plotted against the right-hand-side, as a function of the parameter θ. The point with coordinates (4.32, -2.55) is shown by a square.

The tidal force of the MW can rip satellite galaxies apart if they venture too close to the disc and/or bulge. An example is the 'Sagittarius dwarf galaxy' of which we can trace the tidal debris all around the MW.

The Andromeda galaxy, M31, is very similar in mass and luminosity to the MW, and it has its own set of satellites. Interestingly, M31 and the MW are also gravitationally bound to each other. In fact, M31 is on a collision course with the MW, with the impact expected to be about 5 Gyr from now. The tidal force between both galaxies will be so large that we expect both discs to be destroyed in the collision[7].

The bound system of the MW and its satellites, together with M31 and its satellites, and some further smaller galaxies such as the *triangulum galaxy*, constitute the *Local Group*. The motion M31 as seen from the MW can be used to estimate the mass of the MW, 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 Local Group and in the MW as follows. From the Doppler shifts of

[7]Distances between stars are so large that it is very unlikely that two *stars* will collide when M31 and the MW merge.

spectral lines, we can determine the line-of-sight velocity of M31 with respect to the MW[8],

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

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 the 1950's 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 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[9], then Newton's law gives for the equation of motion

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

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

The solution can be written in the well known parametric form as

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

The distance r increases from 0 (for $t = \theta = 0$) to some maximum value $R_{\rm max}$ (for $\theta = \pi$), and then decreases again. The relative velocity follows from taking the derivative and use the chain rule,

[8]What one measures is the radial velocity wrt to the *Sun*. Since the Sun is on a (nearly) circular orbit around the MW, we need to correct the measured heliocentric velocity of M31 to obtain the radial velocity of Andromeda wrt the MW.

[9]We'll make this assumption for simplicity; GAIA recently measured the *tangential* velocity of M31, see van der Marel et al, 2019.

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

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

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

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[10], but even using ages of the oldest MW stars as Kahn & Woltjer did, $t \sim 15$Gyr, still gives a relatively accurate and interesting value.

So, taking $v = -118$ kms^{-1}, $r = 710$ kpc, and $t = 15$ Gyr, yields $\theta = 4.32$ radians, as shown graphically in Fig. 6.1, when *assuming M31 is on its first approach to the MW*[11].

Substituting these value in the previous equations yields, amongst others, $M_{\rm total} \approx 3.66 \times 10^{12} M_\odot$. Making the reasonable assumption that the MW's halo mass is half of the M31's (given that M31 is twice as bright), yields a total mass of the MW (stars + halo), of

$$M \approx 1.2 \times 10^{12} M_\odot , \qquad (6.12)$$

comfortably higher than the lower limit to M_h of Eq.(6.6).

Notice that this mass is *much* higher than the MW's stellar mass, of $M_\star \approx 7 \times 10^{10} M_\odot$: provided with did our calculations right, the mass in dark matter is ~ 20 times that in stars.

Taking $M_h \approx 1.2 \times 10^{12} M_\odot$, we can estimate the extent of this halo, R_h,

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

If, as is more likely, the rotation speed eventually drops below 220km s^{-1}, then R_h 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 disc is ~ 15kpc, therefore the halo's radius is probably about 7 times that.

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

[11] 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.

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.

CO S 25.1

Elliptical galaxies are spheroidal stellar systems with smooth luminosity profiles. They contain old, typically metal rich stars. Unlike spiral galaxies, ellipticals exist with a very large range of stellar masses. The nature of their ISM is also different, typically consisting of tenuous hot gas detected in X-rays. The dynamics of this gas and of their observed stars both suggest that the mass of eliptical galaxies is also dominated by dark matter.

7.1 Luminosity profile (CO p. 892 & 950)

Elliptical galaxies have smooth surface brightness (SB) profiles, which in projection are ellipsoidal in shape, see examples of such profiles in Fig.7.1[1]. Average intensities as function of radius for the same galaxies are shown in Fig. 7.2. The drawn line, which fits the data well, is the **de Vaucouleurs** or '$r^{1/4}$' profile introduced previously in Eq. (3.2) for the MW's bulge[2].

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

Here, $I(r)$ is the intensity[3] at projected distance r from the centre. When

[1]Notice that in these observations, isophotes near the centre are well defined, whereas they become fainter and hence noisier in the galaxies' outskirts.

[2]The curious factor 7.67 is simply a normalisation to make r_e the half-light radius - meaning half the light is emitted interior to r_e.

[3]The amount of light emitted per unit area.

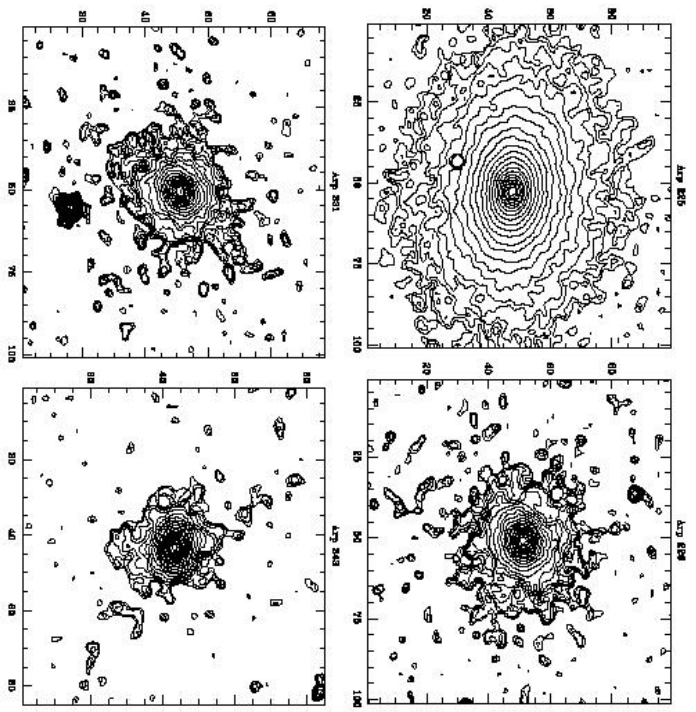

Figure 7.1: (Taken from astro-ph/0206097) Lines of constant surface brightness (isophotes) in the K-band for four elliptical galaxies. In successive isophotes, the surface brightness increases by 0.25 magnitudes.

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.

The galaxy at the very centre of a cluster of galaxies (see next chapter) often has a very extended halo of light - much more extended than the de Vaucouleurs profile. This 'stellar halo' is build-up by the large number of mergers of the central galaxy with other elliptical galaxies. An example of such a galaxy is NGC 1399, the central in the nearby 'Fornax' cluster of galaxies, shown in Fig. 7.4. The extent of NGC 1399s halo is enormous; it can be traced[4] out to ≈ 1 Mpc.

Ellipticals also have many globular clusters (CO p. 962). 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 1000s of high surface brightness objects, indistinguishable on this plate from foreground stars, which are in fact globular clusters in the halo of NGC 1399.

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 of ellipticals

The redder colours of elliptical galaxies is mostly due to the absence of massive and hence hot main sequence stars - itself a consequence of the fact that elliptical galaxies typically have undergone no, or very little, recent star formation. There is also a secondary effect due to the metallicity of the

[4]Which means that, given the distance to this galaxy, its extent on the night sky is comparable to that of the moon.

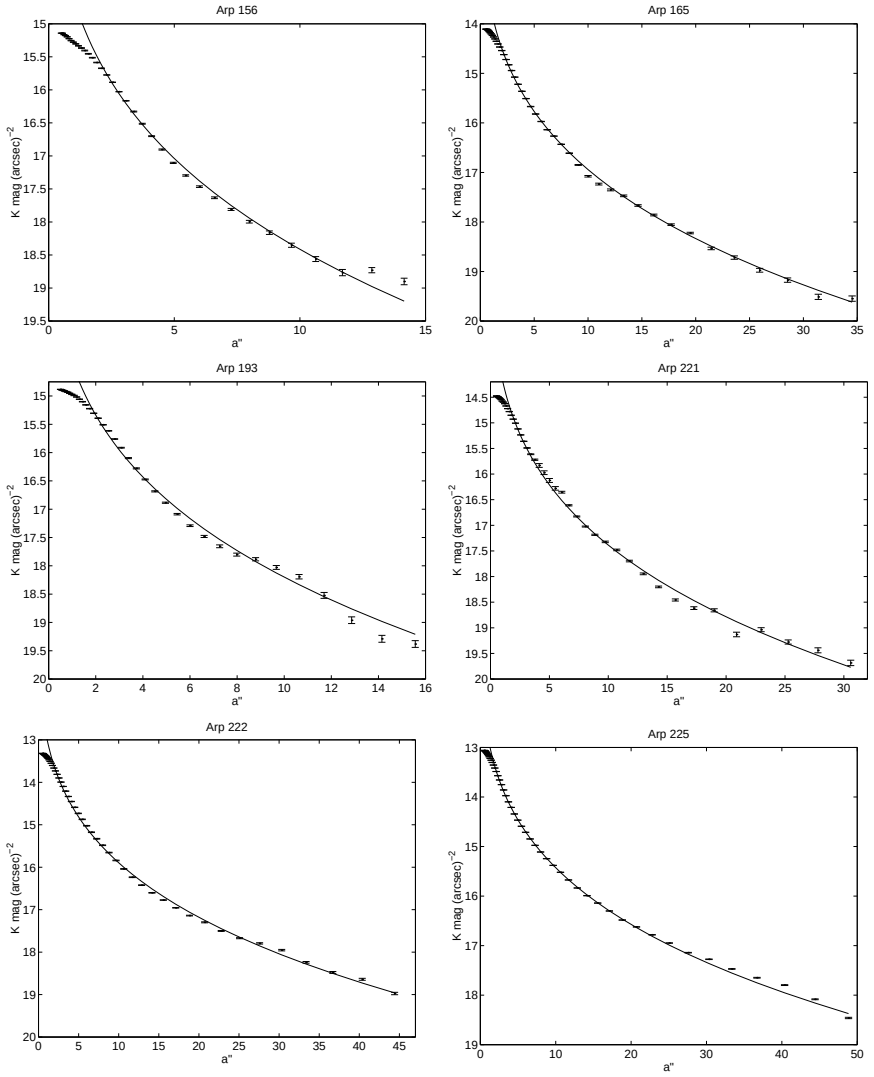

Figure 7.2: (Taken from astro-ph/0206097) Surface brightness as function of distance to the centre for the galaxies from Fig. 7.1. The drawn line is the best $r^{1/4}$ fit.

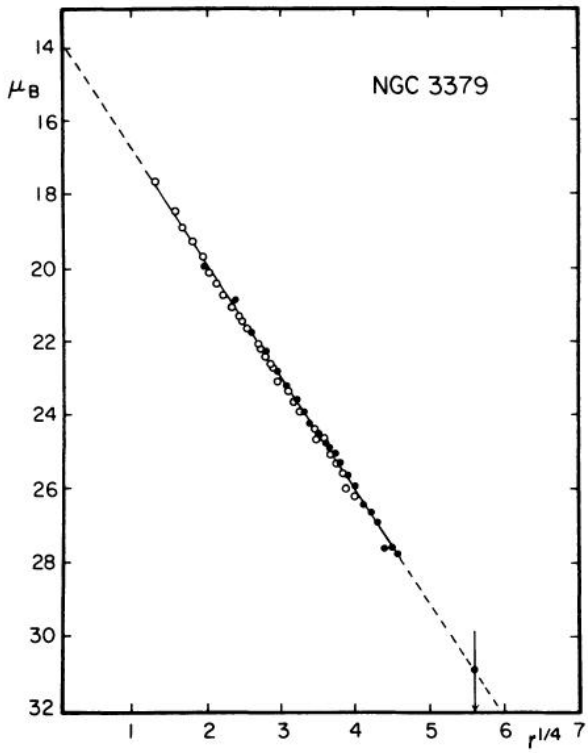

FIG. 2.—Mean E-W luminosity profile of NGC 3379 derived from McDonald photoelectric data. ●, Pe 4 data with 90 cm reflector; ○, Pe 1 data (M + P) with 2 m reflector. Note close agreement with $r^{1/4}$ law.

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

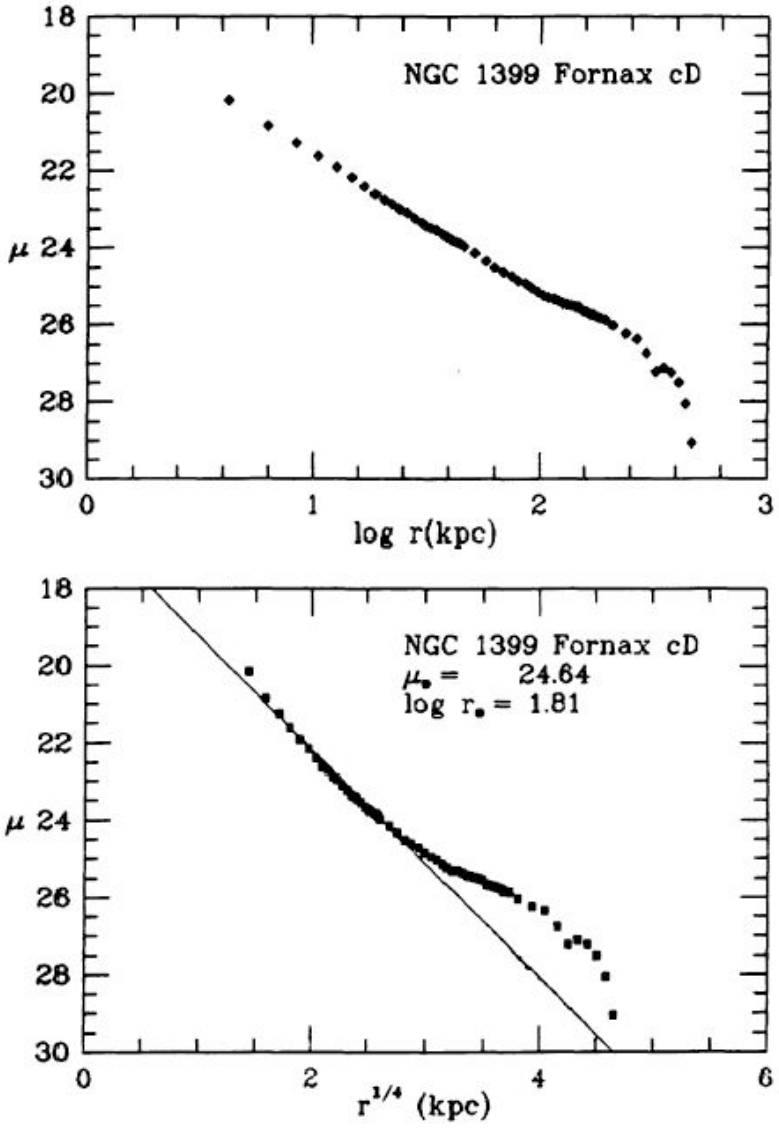

Figure 7.4: SB profile for NGC 1399 from Schombert (filled symbols; ApJS, 1989). The outer parts of the profile do not follow the $r^{1/4}$ fit (bottom panel), but are nearly a straight line in a SB-log(r) (i.e. a log $-$ log) plot (top panel), meaning the profile is close to a power-law.

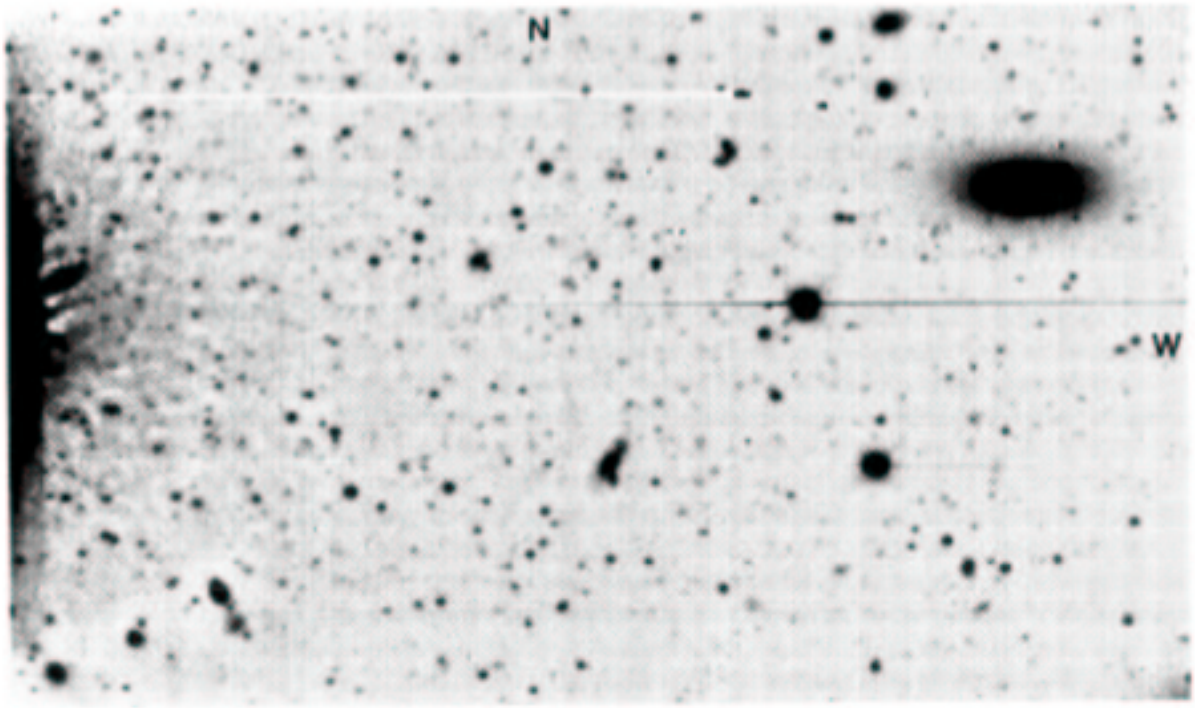

Figure 7.5: Image of NGC 1399 in the Fornax cluster of galaxies, with smooth, best fit $r^{1/4}$-profile subtracted (Bridges et al., AJ 101, 469, 1991). The centre of NGC 1399 is on the left, the object toward the top right is another galaxy in the same cluster. The other extended objects in the image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

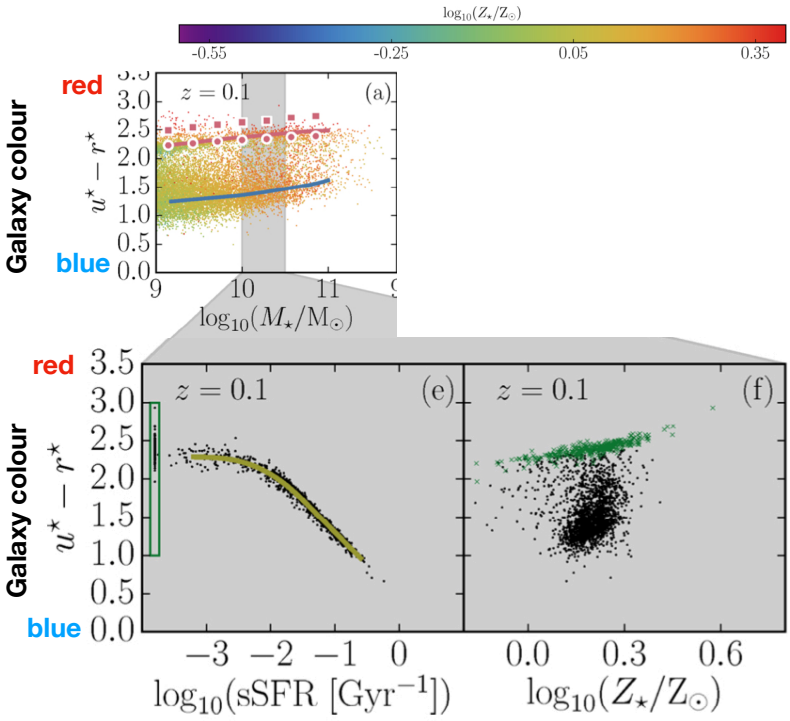

Figure 7.6: The effects of star formation and metallicity on the colours of galaxies, from Trayford+'16. The top panel shows the intrinsic $u^\star - g^\star$ colour of a galaxy plotted against its stellar mass, $M_\star$ (every dot is a galaxy), with high values of $u^\star - g^\star$ denoting a red galaxy, and low values a blue galaxy. The colour of each dot is a measure of the metallicity of the galaxy. The plot shows two sequences: a red sequence with $u^\star - g^\star \approx 2.5$ of elliptical galaxies, and a blue sequence with $u^\star - g^\star \approx 1.5$ and large scatter of star forming galaxies. The bottom left panel shows that, the higher the specific star formation rate, sSFR$\equiv \dot{M}_\star/M_\star$, the bluer the galaxy. The bottom right panel shows that for galaxies on the red sequence, redder galaxies have higher stellar metallicities, $Z_\star$.

stars: a higher metallicity makes a star redder[5]. Both effects are illustrated in Fig. 7.6.

Dust and gas The dust and cold gas that is present in the ISM of spiral galaxies is mostly absent in ellipticals. However, sometimes an elliptical galaxy may accrete a smaller galaxy and tidally tear it apart. This may result in a dust lane running across the elliptical, with the dust originating from the ISM of the ingested galaxy (see for example Fig. 7.7). Such merger events may be quite common: a number of ellipticals have faint rings of stars around them (see for example Fig. 7.8), probably also a result of such 'galactic cannibalism'.

7.3 X-rays

X-rays are absorbed by the Earth's atmosphere hence require observations from rockets or satellites. X-rays would simply be absorbed by a normal mirror and to focus them an X-ray telescope consists of many nearly parallel plates that gently nudge the incoming X-ray onto a detector [6].

Two unrelated sources contribute to the X-rays detected from ellipticals: (1) hot, tenuous gas in their interstellar medium and (2) mass-transfer in close binary stars. The physical process that produces them is the same in both cases: **thermal bremsstrahlung**. Thermal bremsstrahlung[7] is the process whereby high energy radiation (X-rays) is produced in a plasma of ions and electrons, due to the electromagnetic deflection of 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[8] . The plasma is highly ionised because it is so hot, and the electrons are not bound to any particular ion. For this reason, thermal bremsstrahlung is sometimes

[5]Partly due to absorption of blue light in the stellar atmosphere, partly because of subtle changes in the stellar structure.

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

[7]German for 'braking radiation'.

[8]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 7.7: Image of the ‘CenA’ galaxy, with its striking dust lane.

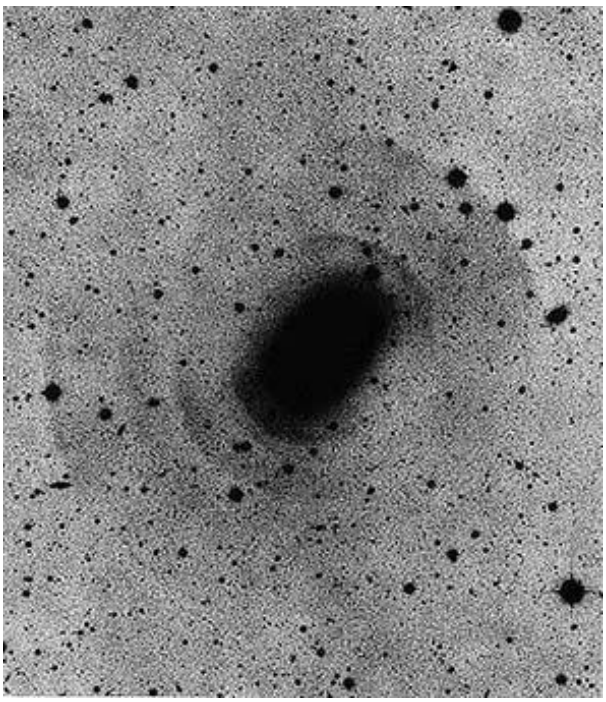

Figure 7.8: Image of NGC 3923 taken by David Malin. Several faint shells of stars appear as ripples in the outer parts of the galaxy, probably resulting from the merger of the central galaxy with a much smaller system. (see http://www.ast.cam.ac.uk/AAO/images/general/ngc3923.html)

called 'free-free' radiation.

Subtracting the X-ray point sources from the X-ray image of an elliptical galaxy - produced by its binary stars - reveals extended X-ray emission. The X-ray *emissivity* ϵ - the amount of energy radiated per unit volume - depends on density and temperature as

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

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 of radiation up to a cut-off frequency $\nu_{\rm max}$ which depends on the temperature of the gas, approximately as $h\nu_{\rm max} \sim kT$. This can be use to measure the temperature of the hot plasma. In addition, observed X-ray spectra of elliptical galaxies also show *spectral lines*, where electronic transitions in some highly ionised elements present in the hot plasma contribute to the spectrum. These lines also allow a measurement of T, as well as a determination of the composition of the plasma (i.e its metal content).

7.4 Evidence for dark matter from X-rays (CO p. 1063)

The hot gas only remains bound to the elliptical galaxy because of the depth of its dark matter halo potential. If we make the reasonable assumption that the gas is in hydrostatic equilibrium with the dark matter potential, then we can measure the required halo mass - in much the same way as you did in the course on stars, as follows.

The hydrostatic equilibrium equation balances the outward force on a gas shell due to the pressure gradient across it to the inward pull due to gravity:

$$\frac{GM}{r^2} = -\frac{1}{\rho}\frac{dp}{dr} \,, \tag{7.3}$$

where M is the mass enclosed in sphere with radius r, p is pressure and ρ is density. We will assume that the gas follows an ideal equation of state, so

that

$$p = \frac{kT}{\mu m_p}\rho\,, \tag{7.4}$$

where k is Boltzman's constant, m_p the proton mass, and we take the mean molecular weight $\mu = 1/2$ - appropriate for a fully ionised hydrogen gas[9].

Observationally, the temperature T is nearly constant across the gas - so, to simplify the calculations, we will assume T to be independent of r. The rhs of Eq. (7.3) then becomes

$$\frac{1}{\rho}\frac{dp}{dr} = \frac{kT}{\mu m_p}\frac{d\ln(\rho)}{dr}\,. \tag{7.5}$$

Now multiply the lhs of Eq. (7.3) with r^2, and take the derivative,

$$\frac{d}{dr}GM = 4\pi G r^2 \rho_{\rm dm}\,, \tag{7.6}$$

where we assumed that the mass M is dominated by dark matter, with density $\rho_{\rm dm}(r)$. Combining the last two equations yields

$$4\pi G r^2 \rho_{\rm dm}(r) = -ccre\frac{kT}{\mu m_p}\frac{d}{dr}r^2\frac{d\ln(\rho)}{dr}\,. \tag{7.7}$$

Now, numerical simulation show that the dm profile of halos is very well fitted by the following functional form[10],

$$\rho_{\rm dm}(r) = \frac{\rho_c}{r/r_s\,(1+r/r_s)^2}\,, \tag{7.8}$$

characterised by two parameters, a density, ρ_c, and a scale-radius, r_s. To find the corresponding gas density profile, consider the following *Ansatz*

$$\rho(x) = \rho_0\,\exp\left(-B\left[1-\frac{\ln(1+x)}{x}\right]\right)\,, \tag{7.9}$$

where ρ_0 and B are constants to be determined, and $x \equiv r/r_s$. A little algebra shows that

$$\frac{d}{dx}x^2\frac{d\ln(\rho(x))}{dx} = -\frac{Bx}{(1+x)^2}, \tag{7.10}$$

[9] As an exercise, compute how this would change if Helium were present as well.

[10] The famouw Navarro-Frenk & White profile, no so long ago projected on Durham cathedral during a *Lumiere* festival.

so that the rhs of Eq. (7.7) becomes

$$\frac{kT}{\mu m_p}\frac{d}{dr}r^2\frac{d\ln(\rho)}{dr} = B\frac{kT}{\mu m_p}\frac{1}{r_s^2}\frac{r/r_s}{(1+r/r_s)^2}\,. \tag{7.11}$$

Therefore

$$\rho_{\rm dm} = \frac{B}{4\pi G}\frac{kT}{\mu m_p}\frac{1}{r/r_s(1+r/r_s)^2}\,. \tag{7.12}$$

Comparison with the NWF profile of Eq.(7.9) shows that the gas distribution is in hydrostatic equilibrium with the dark matter, provided

$$\rho_c = \frac{B}{4\pi G}\frac{kT}{\mu m_p}\,. \tag{7.13}$$

Therefore measuring T and fitting B to the observed profile yields ρ_c as well as r_s - the parameters of the NFW profile. To do better, we would also have to include the contribution of both stars and gas to the potential.

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 profile to the underlying gravitational potential.

Chapter 8

Groups and clusters of galaxies

CO §27.3
Galaxies are *not* sprinkled randomly throughout the Universe. Instead, galaxies like the Milky Way tend to huddle together in small groups similar to the Local Group, with more massive elliptical galaxies clumping together in bigger groups and clusters containing thousands of galaxies. Regions of the Universe with a low density of galaxies are called *voids* - they typically contain smaller galaxies. The origin of all this structure is the amplification by gravity of the tiny fluctuations seen in the cosmic micro-wave background (CMB). But why do the properties of the galaxies depend on their surroundings?

8.1 Introduction

The Milky Way, Andromeda, and several hundred smaller irregular galaxies within $\sim$ 2 Mpc or so from the MW, are part of a gravitationally bound system, called the Local Group, discussed previously. Most spiral galaxies like the MW are found in such small **galaxy groups**.

Galaxy clusters, on the other hand, are gravitationally bound systems of 10s-100s of mostly elliptical and S0 galaxies, together with 1000s of smaller dwarf galaxies. Clusters were discovered by eye - by staring at photographic plates of the night sky, and recognising that patches of the sky contained vastly more galaxies per unit area than average, for example by Abell whose numbering scheme is still in use. Examples of nearby clusters include Fornax

and Virgo, both at a distance of $\sim 20\,\mathrm{Mpc}$ from the MW. Galaxy clusters are the most massive gravitationally bound structures in the Universe. They were the first systems in which there was evidence for the presence of dark matter.

8.1.1 Evidence for dark matter in clusters from galaxy motions (CO p. 960)

The Swiss astronomer Fritz Zwicky measured Doppler velocities of galaxies in clusters of galaxies. He assumed - correctly - that the *velocity dispersion* of these galaxies could be used to characterise the depth of the gravitational potential in which the galaxies move, and from that the mass of the cluster could be inferred. The values of these *dynamical masses* obtained by Zwicky were *much* higher than the corresponding masses in stars: for any reasonable uncertainty in stellar mass, he concluded that most of the mass in the cluster is not visible[1]. His argument goes as follows.

Assume for simplicity that all cluster galaxies have the same mass m, and that the cluster contains N galaxies. The kinetic energy of the system, due to the random motions of the galaxies, can be written as

$$K = \frac{1}{2}\sum m\,v^2 = \frac{1}{2}\,M\,\sigma^2\,, \tag{8.1}$$

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

$$\sigma^2 = \frac{\sum m\,v^2}{M}\,. \tag{8.2}$$

Note that these velocities are measured with respect to the *mean* velocity of the cluster. The potential energy in the system is of order[2]

$$U = -\frac{3}{5}\frac{G\,M^2}{R}\,, \tag{8.3}$$

[1]Zwicky was not aware that the gas mass of a cluster is much larger than the mass in stars. Yet even including this extra mass, it remains true that most mass is invisible.

[2]The factor 3/5 assumes the density of galaxies is constant inside the cluster, which is clearly an approximation.

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

$$M = \frac{5}{3}\frac{\sigma^2 R}{G}\,. \tag{8.4}$$

We have now two estimates for the mass of the cluster: (1) the mass obtained from Eq. (8.4) (a *dynamical* mass - inferred from the dynamics of galaxy motions), and (2) the *stellar* mass, $M_\star$. The stellar mass is not directly observable: what we can measure is the total luminosity of all galaxies in the cluster, $L_\star$ (provided we can measure the flux of the cluster, as well as the distance to the cluster.). **If** all cluster stars were the same as the Sun, then the cluster mass would be $M_\odot \times (L_\star/L_\odot)$. However, this not likley to be correct. Zwicky estimated that, if the stellar population of cluster galaxies were similar to that of the Milky Way, then a better estimate of the stellar mass of the cluster would be $M_\star \approx 3M_\odot \times (L_\star/L_\odot)$. This is because low mass stars contribute very little to $L_\star$, but they do contribute to $M_\star$. Therefore, if the cluster galaxies contain low mass stars as well, then this would increase the value of $M_\star$. It turns out this correction is not very important, since in any case,

$$M \gg M_\star\,, \tag{8.5}$$

the dynamical mass is **much** larger than the stellar mass - Zwicky estimated that $M \approx 400M_\star$. The alternative is that the observed cluster is simply unbound, with galaxies now escaping the system.

8.2 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, ρ, and pressure, p, as (see Eq. 7.3)

[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.

$$\frac{GM(<r)}{r^2} = -\frac{1}{\rho(r)}\frac{dp}{dr}\,. \tag{8.6}$$

As for ellipticals, the X-ray data can be used to measure the right hand side, allowing a determination of the encloded mass, $M(<r)$. Comparing this to the mass in stars $M_\star$, inferred from the luminosity of the galaxies, and the mass in gas, $M_{\rm gas}$, inferred from the X-ray observations, yields $M_\star + M_{\rm gas} \approx M_{\rm gas} < M$: gas dominates over stars (by a large factor), but the gas mass is still significantly below the dynamical mass: 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 attract gas (and dark matter) 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*. This works as follows: assume that a parcel of gas starts at infinity with velocity $v = 0$. The parcel feels the gravitational pull from the cluster, gets accelerated and eventually hits the cluster itself, at a distance R from the centre - R is the radius of the cluster with mass M.

Since energy is conserved along the orbit of the parcel of gas, we can compare its energy at infinity to its energy at R:

$$0 = E = \frac{1}{2}v^2 - \frac{G\,M}{R}\,, \tag{8.7}$$

since the energy of the parcel at infinity is zero (its speed is zero, and its gravitational energy GM/r is also zero for $r \to \infty$. So that sets the speed with which the gas slams into the cluster, in terms of M and R. The gas will now converts its kinetic energy into thermal energy, which means its gets heated to temperature T, given by

$$\frac{1}{2}mv^2 = \frac{3}{2}m\frac{kT}{\mu m_p}\,, \tag{8.8}$$

where μm_p is the mean molecular weight per particle. Combining the last two equations yields

$$kT = \frac{2}{3}\frac{\mu m_p GM}{R} = 15\times 10^3 \frac{(\mu/0.5)(M/10^{15}M_\odot)}{R/1{\rm Mpc}}{\rm eV}\,, \tag{8.9}$$

corresponding to $T \approx 1.7 \times 10^8$K. This temperature is called the *virial temperature*, and its high value explains why we detect X-rays with energies of the order of 10^{3-4} eV.

8.3 Metallicity of the X-ray emitting gas.

The X-ray spectrum of a galaxy cluster consists of a power-law component with a cut-off at high energy resulting from thermal bremsstrahlung with additional emission lines from highly ionised metals such as Si, N and Fe, see Figure 8.1. The location of the cut-off in the continuum shape, as well as the ratio of emission lines for a given element, can be used to infer the temperature of the gas, T. Once T is known, the gas density - and gas mass - follows from measuring the X-ray luminosity, since the emissivity is $\propto \rho^2 T^{1/2}$ for thermal bremsstrahlung (with a known proportionality constant).

The first surprise of such an analysis is that most of the baryonic mass in the cluster is in the X-ray emitting gas: $M_\star \ll M_{\rm gas}$. Although we originally 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.

The second surprise is the high abundance of metals in this gas. For example for Fe, we find that the ratio $M_{\rm Fe}/M_{\rm gas} \approx 1/3 (M_{\rm Fe}/M)_\odot$, that is: the ratio of iron mass to total gas mass in the cluster is of order 1/3 of the iron mass fraction in the Sun. Such a ratio is higher than for most *stars* in globular clusters. Why is this so surprising? Well - how did these metals - synthesized in *stars inside galaxies* - manage to be flung out of the galaxy and into the gas in between the galaxies? Clearly, we must conclude that galaxies are not simply closed boxes: elements synthesized inside a galaxy manage to escape the galaxy. Our present bet is that the combined action of many super nova explosions manages to eject a considerable fraction of the metal enriched gas outside of the galaxy - see the workshop for some exercises on this.

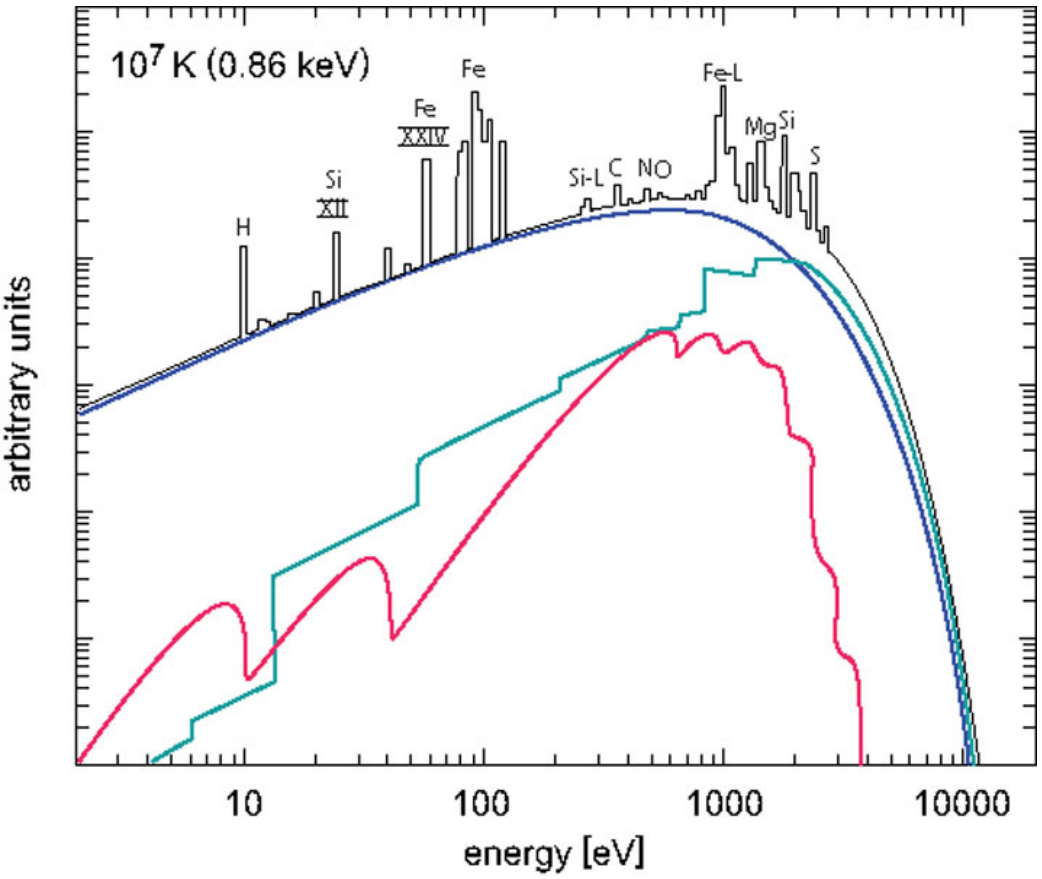

Figure 8.1: Model X-ray spectrum of a plasma with temperature $T = 10^7$ K and a solar abundance pattern. The underlying continuum (the smooth line) is due to thermal bremsstrahlung - notice the sharp cut-off in the emissivity above $\sim 2 \times 10^3$ eV. The emission lines are due to electronic transitions in highly ionised gas.

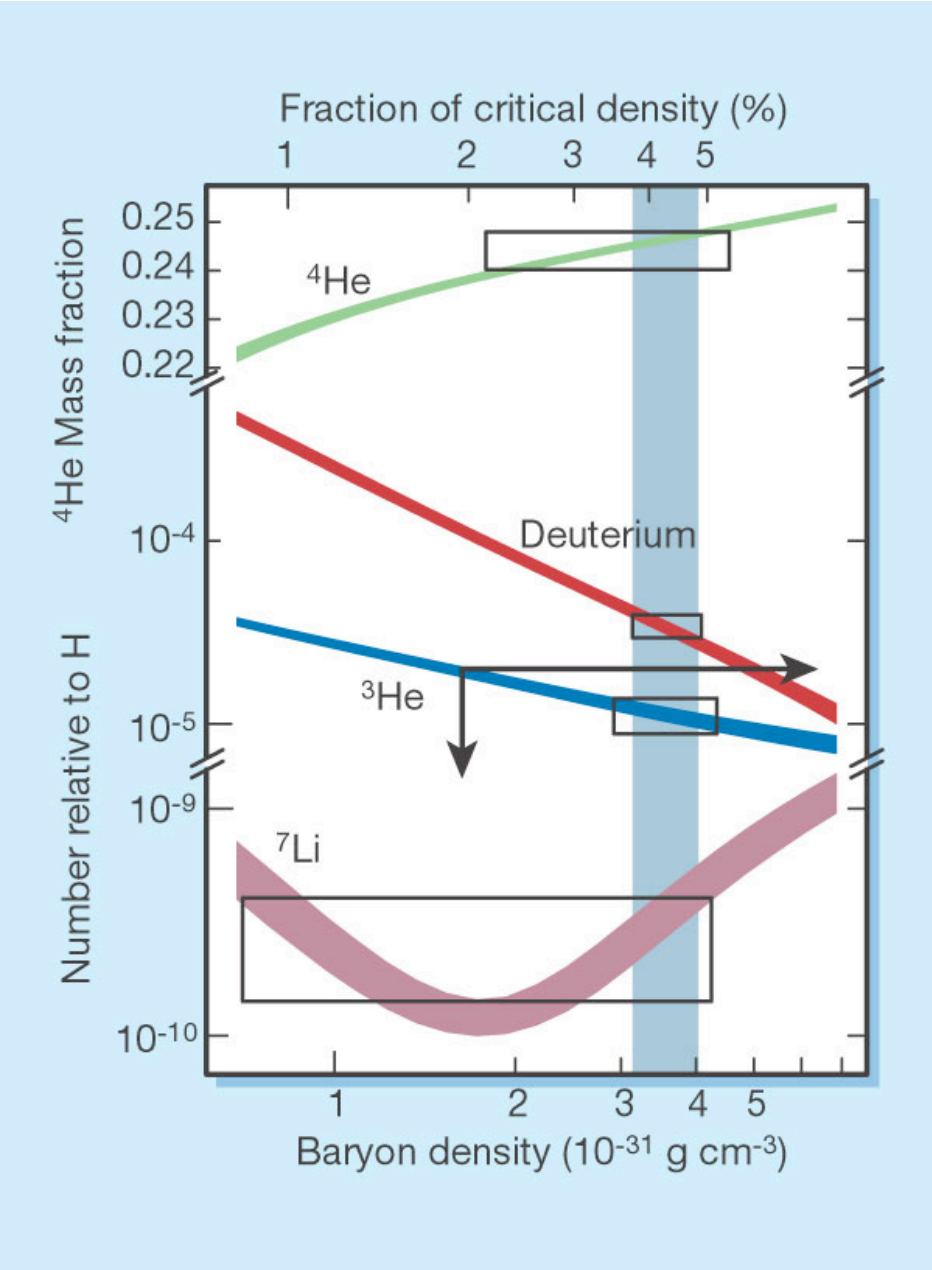

Figure 8.2: The production of various elements during Big Bang nucleosynthesis as a function of the baryon density. (*Nature* **415**, p. 27, 2002)

8.4 The dark matter density of the Universe

Clusters have been used to estimate the mean dark matter density of the Universe as follows. Assume that the Universe starts-out smooth, with a (nearly) constant ratio ω of dark matter to baryons[4]

$$\omega = \frac{\rho_{\rm dm}}{\rho_{\rm b}} \,. \tag{8.10}$$

A cluster forms by accreting both dark matter and baryons. As the gravitational potential well of the forming cluster deepens, it is (probably) a good approximation to assume that eventually, neither dark matter nor baryons can ever escape from the cluster's potential well[5]. As a result, the ratio of dark matter mass to baryon mass of the cluster, is also equal to ω:

$$\omega \approx \frac{M_{\rm dm}}{M_{\rm b}} \,. \tag{8.11}$$

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. 8.4) 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 ordinary hydrogen*, $\rho_{\rm D}/\rho_{\rm H}$. The reason is that deuterium is produced during Big Bang nucleosynthesis, with the ratio $\rho_{\rm D}/\rho_{\rm H}$ depending on the total baryon density, as illustrated in Fig.8.2.

We can measure the deuterium fraction in intergalactic gas clouds. This, together with Big Bang nucleosynthesis calculations yields ρ_b, and given ω from cluster observations, finally yields $\rho_{\rm dm}$:

[4]According to Cern, baryons are composed of three quarks. Astronomers use the term *baryon* to mean ordinary matter (such as stars and gas) composed of protons and neutrons - as opposed to dark matter whose composition is presently unknown.

[5]We know for a fact this is not true for the gravitational potential of the Milky Way - recall our discussion on high and hyper velocity stars, for example.

[6] Deuterium is an isotope of hydrogen: a deuterium nucleus consists of a proton and a neutron, a opposed to the nucleus of ordinary hydrogen which is just a single proton. Deuterium is not produced but is destroyed in stars.

$$\rho_{\rm dm} \approx 4 \times 10^{-31} {\rm g\ cm^{-3}}\,. \tag{8.12}$$

This is astonishingly low compared to the density of you, the reader, which is ~ 30 *orders of magnitude higher!* More recent estimates of $\rho_{\rm dm}$, based on the cosmic microwave background, are consistent with the value found from clusters.

8.5 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 9

Galaxy statistics

9.1 Introduction

To better understand how galaxies form and evolve, we require a large and unbiased *sample* of them. Acquiring such a sample is difficult for many reasons. Firstly, we detect galaxies based on their *flux* and as a consequence we can detect luminous galaxies to large distances but intrinsically faint galaxies only nearby. So if we simply studied all galaxies above a given flux limit, then intrinsically bright galaxies would be over represented compared to intrinsically faint galaxies. Secondly, measuring reliable distances is difficult with Cepheid distance measurements possible to out to $\sim$ 20 Mpc but not much further.

Galaxy redshift surveys use the redshift of a galaxy to infer the distance by assuming that the Universe expands at a known rate,

$$\lambda = \lambda_0(1 + \frac{v}{c}) = \lambda_0(1 + H_0\frac{r}{c}) \,, \tag{9.1}$$

where H_0 is the Hubble constant and λ the observed wavelength of a line in the galaxy's spectrum with laboratory wavelength λ_0; r is the (Hubble) distance to the galaxy[1]. Surveys select objects on the night sky identified in photographics plates - or more recently CCD images - based on size, luminosity and colour to distinguish galaxies from stars or other foregrounds.

[1]The relation between physical size l and and angular extent θ defines the distance r_A: $\theta = l/r_A$. The relation between flux and luminosity defines the distance r_L: $F = L/(4\pi r_L^2)$. Locally, these distances are the same, $r_A = r_L = r$, but on cosmological scales they are not because space is curved.

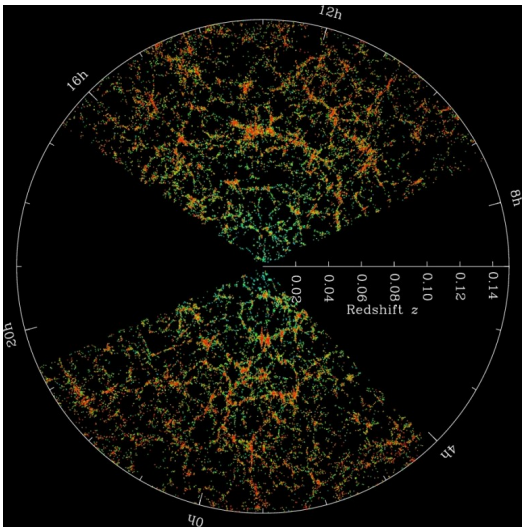

Figure 9.1: Large-scale structure in the distribution of galaxies around us (we are at the centre of the map), from the *dark energy survey*. Galaxies are surveyed in two pie-shaped regions on the sky, and are plotted as a function of right-ascension and distance. Each coloured dot is a galaxy, with red dots representing galaxies with little current star formation (ellipticals), and green and blue dots representing star forming galaxies (spiral and irregular galaxies). The striking 'fingers of god' correspond to galaxy clusters. Notice also the characteristic 'filamentary' distribution of galaxies, and the large dark regions with few galaxies - called 'voids'. For reference, the distance to $z = 0.14$ is ≈ 600 Mpc.

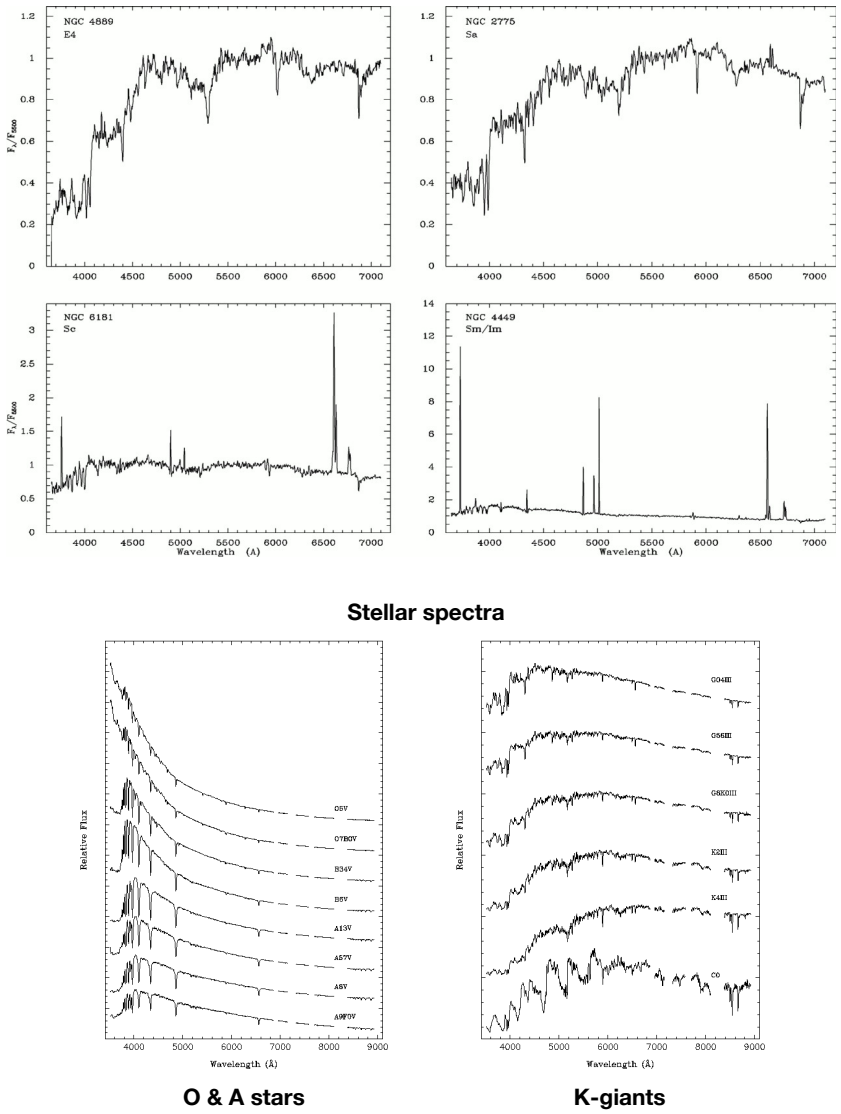

Figure 9.2: Top panels: Sample galaxy spectra from Kennicutt 1992. Galaxy type is indicated in each panel (top left: elliptical, top right and bottom left: spiral galaxies, and bottom right: irregular galaxy.). Notice in particular the deep absorption troughs (corresponding to light from K-giant stars) and the absence of blue continuum in the spectrum of the elliptical, and the striking emission lines (in particular Hα at $\lambda \approx 6800$Å) and the light from A-type stars in the spectra of spirals and irregulars. Bottom panels: sample spectra of O & A type stars (left) and of K-gaints (right).

By measuring a spectrum, they determine the galaxy's Doppler shift, v/c, and from that infer the distance to the galaxy. This allows us to make a 3 dimensional map of galaxies around us, see Fig. 9.1. Given the distance to the galaxy, we can infer its luminosity by measuring its flux. The galaxy's spectrum allows us to infer many other properties for the galaxy as well, for example characterise its stellar population and from that its stellar mass, $M_\star$, measure its star formation rate (basically by counting the combined luminosity of all its H II regions), and measure the mean metallicity of its stars. Compare the galaxy spectra with spectra from individual stars in Fig. 9.2. Future surveys such a *Euclid* aim to observe ~ 2 billion galaxies [2]. We discuss some striking correlations of galaxy properties next.

9.2 The galaxy luminosity and stellar mass function

Counting galaxies as a function of luminosity allows us to compute the *galaxy luminosity function* - the number density of galaxies as a function of luminosity. Combined with modelling the stellar population of these galaxies, we can compute the *galaxy stellar mass function* (GSMF) - the number density of galaxies as a function of stellar mass, see Fig. 9.3. Notice how the number density of low mass galaxies increases with decreasing mass as a power-law. At masses above $M_\star \sim 10^{10.5}\mathrm{M}_\odot$, the number density of galaxies drops very rapidly. To obtain the mass in galaxies of a given mass, we would need to multiply Φ by $M_\star$ - showing that galaxies with mass $\sim 10^{10.5}\mathrm{M}_\odot$ - i.e. with masses similar to the MW - dominate the mass budget[3].

The shape of the GSMF is well captured by the following fit due to[4] Schechter in 1976,

$$\Phi = \frac{dn}{d\log M_\star} = \Phi_0 \left(\frac{M_\star}{M_{\star,c}}\right)^\alpha \exp(-M_\star/M_{\star,c})\,, \tag{9.2}$$

[2]The number of galaxies in the observable Universe that are brighter than the Milky Way is estimated at ~ 200 billion.

[3]Or, with other words, most stars in the Universe today are in galaxies with mass similar to that of the MW.

[4]The Schechter fit was aimed at fitting the luminosity function, but it works equally well for the mass function.

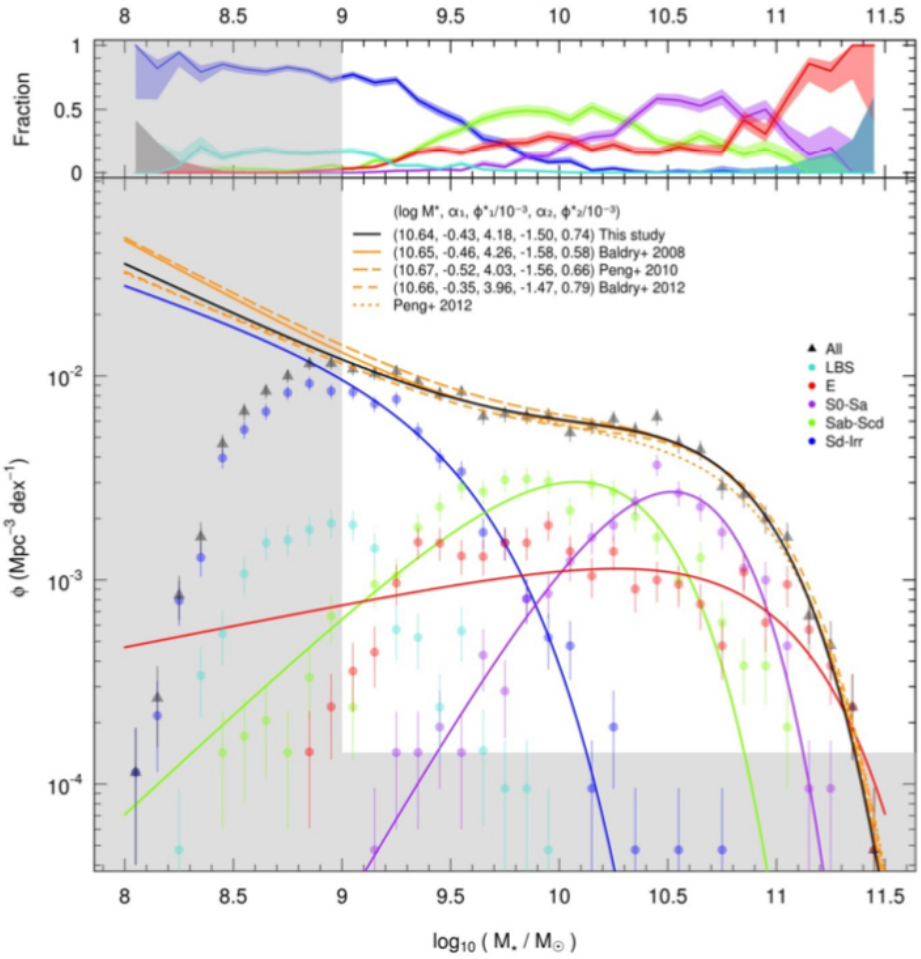

Figure 9.3: Black line: Φ - the number density of galaxies per decade in stellar mass. The different colours decompose this galaxy stellar mass function in terms of different galaxy types. Notice the power-law increase in the number of low-mass galaxies, and the rapid (exponential) drop at high masses. Figure taken from Kelvin 2015.

where n is the number density of galaxies. The function has three fitting parameters:

1. Φ_0 - a measure of the mean number density of galaxies
2. α - a measure of the low-mass slope
3. $M_{\star,c}$ - a characteristic mass above which Φ drops exponentially

9.3 The density-morphology relation

The *density-morphology relation* (Dressler 1980)is the observed correlation between the morphology of galaxies (elliptical versus spiral, or non-star forming versus star forming) and the local density of galaxies (the number of galaxies per unit volume): elliptical galaxies tend to live in dense regions, spirals avoid dense regions - see Fig. 9.4.

There are probably several processes that drive this correlation, but the following three processes probably play a major role:

1. Galaxies in a high density regions undergo frequent tidal interactions with other galaxies - these interactions may destroy fragile discs.
2. The hot gas present in high density regions (clusters) may strip star forming gas from a galaxy, shutting down its star formation.
3. A galaxy in the halo of a more massive galaxy (such as a galaxy in a cluster) - may no longer accrete gas from the intergalactic medium - it is being starved of fuel for star formation.

9.4 Galaxy scaling relations

A *galaxy scaling relation* is a relation between different properties of the galaxy, for example between galaxy mass ($M_\star$) and galaxy size ($R_\star$), or galaxy mass and stellar metallicity ($Z_\star$). The origin of these relations is not always well known. Here we discuss some really striking relations.

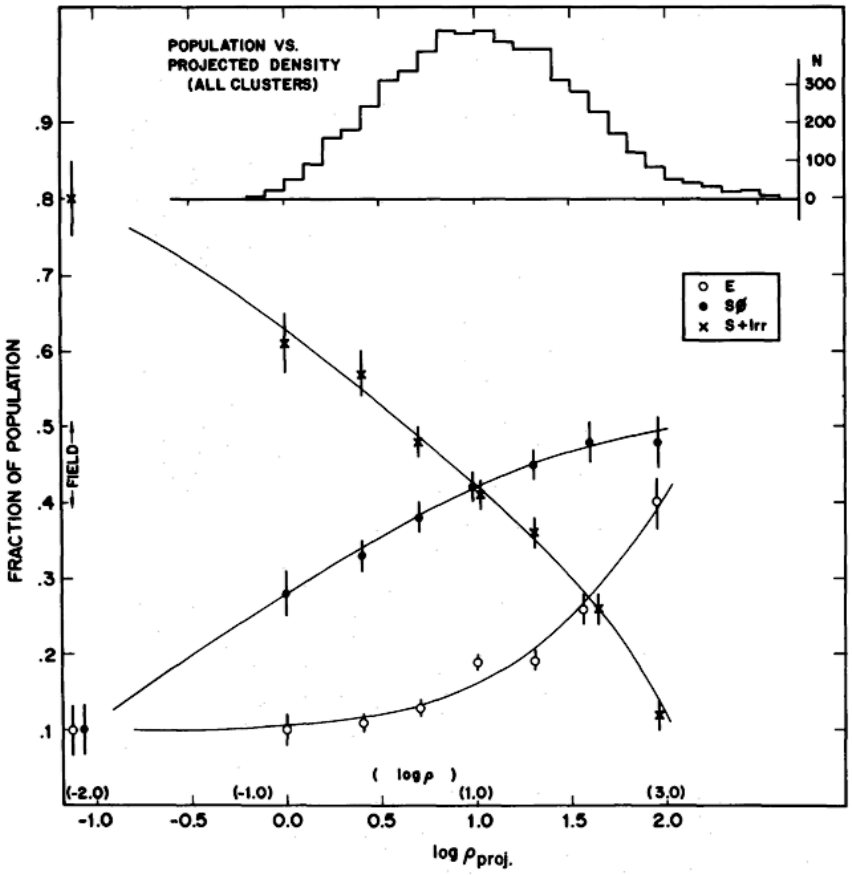

Figure 9.4: The fraction of elliptical (E), S0, and spiral + irregular (S+I) galaxies as function of the logarithm of the projected density($\rho_{\rm proj}$, in galaxies $\rm Mpc^{-2}$). At low density (small $\rho_{\rm proj}$), most galaxies are of type S or Irr, whereas at high densities, most galaxies are S0 or E: there appears to be a relation between galaxy density and galaxy morphology. Figure taken from Dressler 1980.

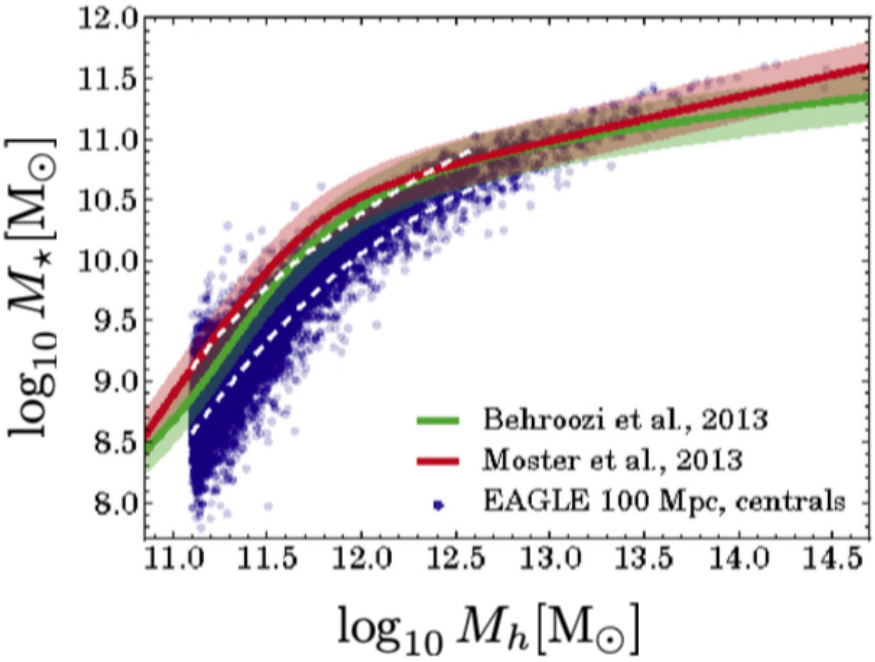

Figure 9.5: Relation between dark matter halo mass, M_h, and galaxy stellar mass, $M_\star$, from Matthee et al, 2017. Red and green drawn lines are relations inferred from the observed galaxy stellar mass function, purple points are galaxies from the EAGLE cosmological hydrodynamical simulation.

9.4.1 The stellar mass - halo mass relation

We discussed at length the evidence that galaxies form inside massive dark matter halos, with galaxy stellar mass much smaller than halo dark matter mass, $M_\star \ll M_h$. Directly measuring M_h is not easy, and the methods we described so far are not able to measure M_h for a large number of galaxies. *Gravitational lensing*, discussed in Chapter 11, can in principle be used to measure M_h. But an indirect method is called *abundance matching*. This method uses the fact that we can compute very accurately the number density of dark matter halos as a function of their mass. If we make the reasonable assumption that *more massive galaxies inhabit more massive halos*, than we can relate $M_\star$ to M_h using *abundance matching* - meaning that we identify galaxies of a given $M_\star$ to halos with a given M_h, provided they have the same number density.

So, combining the observed number density of galaxies with given $M_\star$ from §9.2 with the theoretical number density of *halos* with given M_h, yields the $M_\star - M_h$ relation of Fig. 9.5. Reproducing the observed relation is a stringent test of galaxy formation theories. In addition, some observed properties of galaxies can be understood in terms of this $M_\star - M_h$ relation.

9.4.2 The Tully-Fisher relation (CO p. 952-956)

Spiral galaxies have flat rotation curves: the circular velocity is independent of radius away from the centre. The *Tully-Fisher* relation is an observed relation between that constant circular speed, V_c, and the luminosity of the galaxy, L, of the form

$$L = L_0 \left(\frac{V_c}{V_0}\right)^4 . \tag{9.3}$$

The value of the exponent depends (weakly) on the choice of filter in which the luminosity is measured; the ratio L_0/V_0^4 is a normalisation constant[5].

We can determine the normalisation constant L_0/V_0^4 as follows. Suppose we determine the distances to (nearby) spiral galaxies using the period-luminosity relation of Cepheid variables. Knowing the distance, we can compute the luminosity of the galaxy once we've measured its flux. If we also measure the rotation speed, we have both L and V_c, and hence can compute

[5]Notice that we only need to know the ratio L_0/V_0^4, we never need to know them separately.

L_0/V_0^4:

measure distance d, flux F, rotation speed $V_c \rightarrow L = (4\pi d^2)F \rightarrow \dfrac{L_0}{V_0^4} = \dfrac{L}{V_c^4}$. (9.4)

This turns the TF relation into a **standard candle**! Indeed, suppose we measure the flux and rotation speed of a distant spiral galaxy. Assuming it is on the TF relation, we can estimate L from measuring V_c, and hence determine d:

$$\text{measure } V_c \text{ and flux, } F \rightarrow L = L_0 \left(\frac{V_c}{V_0}\right)^4 \rightarrow d = \left(\frac{L}{4\pi F}\right)^{1/2} . \qquad (9.5)$$

Is is not surprising that more massive galaxies are both brighter and have a higher value of circular speed. What is surprising is that observed galaxies follow the relation with so little scatter. Ultimately, the TF relation is closely related to the $M_\star - M_h$ relation, but there is more to it than that.

The usual textbook explanation for the origin of the TF relation goes as follows. The circular velocity depends on enclosed mass, $M(< R)$, as,

$$V_c^2 = \frac{\mathrm{G}\, M(< R)}{R} . \qquad (9.6)$$

Now define the *mass-to-light ratio*, Υ, using the galaxy's luminosity, L,

$$\Upsilon \equiv \frac{M(< R)}{L} , \rightarrow V_c^2 = \mathrm{G}\Upsilon \frac{L}{R} . \qquad (9.7)$$

The mass-to-light ratio Υ will depend on the number and types of stars in the galaxy (which determines L), and the amount of dark matter (which mainly determines M). Eliminate R, the radius of the galaxy, by using the intensity I - the luminosity per unit area,

$$I = \frac{L}{\pi R^2} \rightarrow R = \left(\frac{L}{\pi I}\right)^{1/2} . \qquad (9.8)$$

Combining the above equations yields

$$L = \frac{1}{\mathrm{G}^2 \Upsilon^2 \pi I} V_c^4 . \qquad (9.9)$$

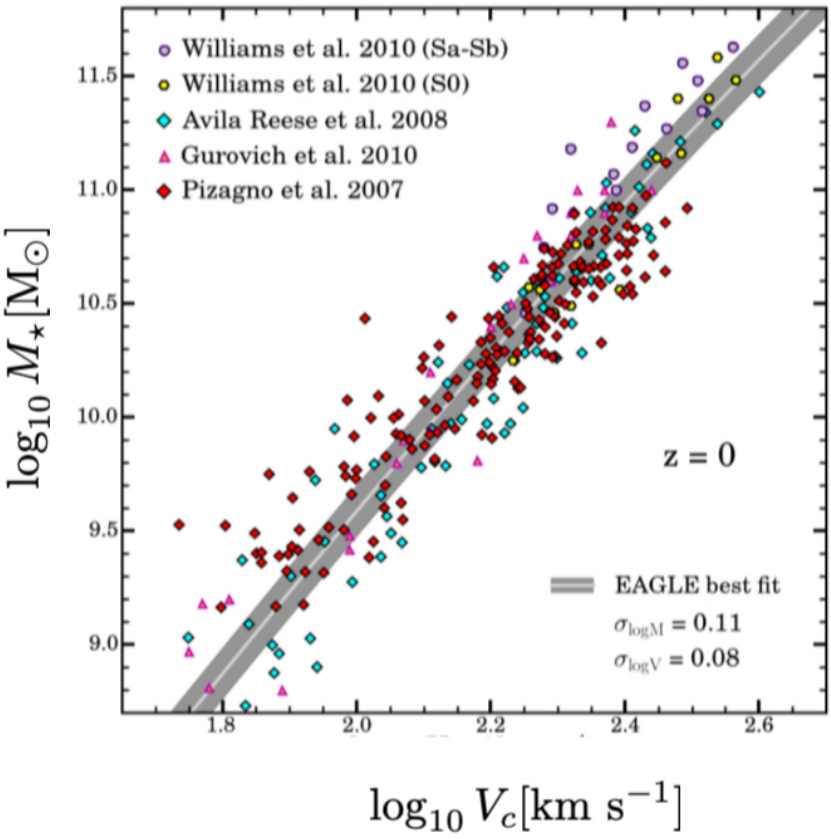

Figure 9.6: Observed 'Tully-Fisher' relation, which relates a galaxy's stellar mass, $M_\star$ (or galaxy luminosity), to circular speed, V_c. Coloured symbols are observed galaxies, grey band is a prediction from a simulation of galaxy formation. Figure from Ferrero et al, 2014.

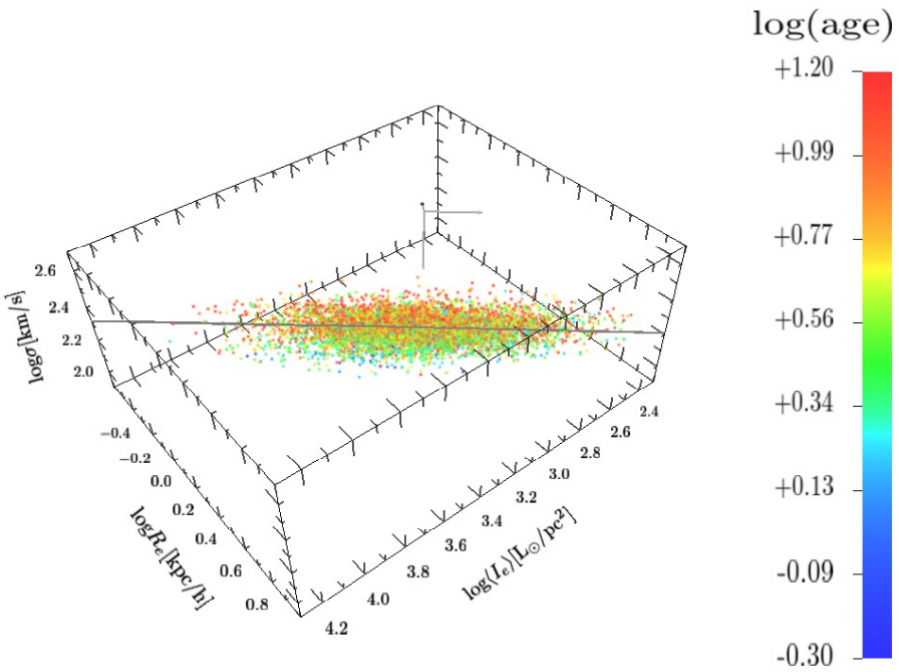

Figure 9.7: Observed correlation between the size of galaxies, R_e, their stellar velocity dispersion, σ, and their central intensity, I_e. Every dot is an elliptical galaxy, the colour of the dot is a measure of the galaxy's stellar age. Striking is that galaxies do not scatter randomly in this figure, but lie on a plane - the fundamental plane. Figure from Magoulas et al, 2012.

This relation has the form of the TF relation, provided $\Upsilon^2 I$ is (approximately) constant. But why would that be true? Since L is related to stars, but M (mostly) to dark matter, the product $\Upsilon^2 I$ can only be approximately constant if the stellar mass (which sets L) and the size of the galaxy (which sets R) are correlated with the mass of the halo (which sets M). Or in other words, provided the properties of the galaxy are closely related to those of its halo.

9.4.3 The Faber-Jackson relation and the fundamental plane in ellipticals (CO p. 987)

To find an equivalent scaling relation for ellipticals, we start from

$$\sigma^2 = \frac{5}{3}\frac{GM(< R)}{R}\,, \tag{9.10}$$

motivated by Eq. (8.4). Following the TF reasoning predicts

$$L \propto \sigma^4\,, \tag{9.11}$$

called the **Faber-Jackson** relation. Similar to the TF relation, it acts as a distance indicator: suppose we first determine the proportionality constant by measuring σ and L for nearby galaxies (requiring a distance measurement based on Cepheids, say, and measuring the galaxy's flux). We can now measure distances to distant ellipticals, by measuring σ from a galaxy spectrum, computing L by assuming the galaxies lies on the FJ relation, and then obtain distance from the computed value of L and the measured flux.

Elliptical galaxies do follow the FJ-relation approximately, but the *scatter* around the relation is significantly larger than the scatter around the TF relation. This motivated Djorgovski & Davis (1987) to introduce a dependence of Υ on L to improve the correlation (i.e. reduce the scatter). They assumed that

$$\Upsilon \propto L^\alpha\,, \tag{9.12}$$

for some value of the exponent α that is to be determined. Combining Eq. (9.10) with the relations from the previous section, yields the following relation between R, L and σ,

$$R \propto \sigma^{2/(1+2\alpha)}\, I^{-(1+\alpha)/(1+2\alpha)}\,. \tag{9.13}$$

Figure 9.7 plots elliptical galaxies in the 3D space of $R-\sigma-I$ - where they tend to lie close to a plane, of the form

$$R \propto \sigma^{1.25} I^{-0.89}\,. \tag{9.14}$$

This plane is called the **fundamental plane**. Equation (9.13) reduces to Eq. (9.14) for $\alpha \approx 0.24$. The origin of this relation is perhaps even less well understood than the origin of the TF relation. But the fact that ellipticals follow this fundamental relation can (again) be used to measure distances.

9.5 Tully-Fisher and Fundamental plane relations as standard candles

The importance of the TF and fundamental plane relations are twofold

- they show that the growth of a galaxy is closely related to that of its dark matter halo (although detailed understanding of the underlying physics of why the scatter is so small is currently lacking)
- both can be used to measure distances to galaxies

The reason they can be used to measure distances is because they relate parameters of the galaxy that are easy to measure and are *distance independent* , to properties of the galaxy that *do* depend on distance.

- For the TF relation: V_c is independent of distance, d, but when L is inferred from V_c using the TF relation, d follows by combining the measured flux and the computed luminosity.
- For the FP relation: σ and I can be measured independently of distance[6]. Combining these with the FP relation yields R. The distance d follows from measuring the angular extent of the galaxy.

[6]Recall that surface brightness is distance independent

9.6 Summary

After having studied this lecture, you should be able to

- explain what the galaxy luminosity and stellar mass functions are, and sketch them
- describe how the stellar-mass halo mass relation is determined
- describe the density-morphology relation for galaxies
- describe the Tully-Fisher relation, and explain how it is used as a distance indicator
- describe the Faber-Jackson and fundamental plane relations, and explain how they are used as a distance indicator

Chapter 10

Active Galactic Nuclei (AGN)

CO §28

Every sufficiently massive galaxy is thought to harbour a super massive black hole (SMBH, BH for black hole) in its centre. Evidence for the presence of an SMBH with mass $M_{\rm BH} \approx 4 \times 10^6 {\rm M}_\odot$ at the centre of the Milky Way galaxy, from the motion of stars in the vicinity of the BH, is especially convincing, as is the recent detection of an SMBH with mass $M_{\rm BH} \approx 6.4 \times 10^9 {\rm M}_\odot$ in the centre of galaxy M87, though the detection of its 'shadow'. The origin of these SMBHs is presently unclear. However, they can grow in mass by mergers with other SMBHs (during a galaxy-galaxy merger) or through accretion of gas through an *accretion disc*. Energetic phenomena in and around the accretion disc turn SMBHs into the most luminous objects in the Universe, displaying a rather baffling variety of phenomena from extremely luminous radio sources to optically luminous quasars. The generic name for an accreting SMBH that emits copious radiation is 'Active Galactic Nucleus'- or AGN for short. It is becoming clear that such AGN can dramatically affect their host galaxy, suppressing or indeed preventing star formation.

10.1 Discovery

Radio waves from outside the solar system were first detected by Jansky in 1933, who also correctly identified the physical process that generates

them - synchrotron radiation[1] - yet the source of the emission was unknown. Following-up from this, Hey et al., 1946 reported strong radio-emission emanating from the direction of Cygnus - later identified as coming from the AGN now called 'Cygnus A'. Synchrotron spectra are power-laws therefore there is little information in the spectrum about the nature of the source. A more systematic investigation of what caused this radiation was made possible by the 'third Cambridge all-sky radio-survey' (3C) published by Edge et al., 1959 and refined by Bennett 1962 . Identifying optical counterparts (*i.e.* optical sources at the exact same position in the sky as the radio source) allowed detailed study of the sources of radio emission, heralding the era of AGN studies.

10.2 Observational manifestations of AGN

Although originally discovered by their radio-emission, AGN display a baffling variety of observational manifestations. For some AGN, light is detected from the radio over IR and optical-UV to X-rays to gamma rays. It is thought that this light is produced by gas accreting onto the SMBH through an accretion disc, but many details remain to be understood.

Figure 10.1 shows two famous examples. The top left panel is a radio image taken by the VLA of Cygnus-A. Notice the two very extended radio-lobes, and the thin 'jet' of emission that connects them to the bright source in the centre. The bright source is the central BH, located at the centre of a galaxy (which you can't see in this radio image). Notice also the shear scale of the lobes, with the image extending over 150 kpc, 5 times the diameter of the MW's disc.

The top right panel shows the quasar 3C273, indicated by an arrow. At a distance of 740 Mpc, this quasar appears almost equally bright as the MW foreground star next to it, even though this solar luminosity star is much closer, at a distance of ~ 0.5 kpc - implying the luminosity of 3C272 is $\sim 2 \times 10^{12} L_{\odot}$ - or about 100 times the luminosity of the MW. The most luminous quasar currently known has $L \sim 4 \times 10^{14} L_{\odot}$ - or 10^4 times as bright as the MW!

The bottom left panel is a deep X-ray image taken by the Chandra telescope. Visible in this image are two clusters of galaxies detected in X-rays

[1]The radiation emitted by electrons on helical trajectories when moving through a magnetic field.

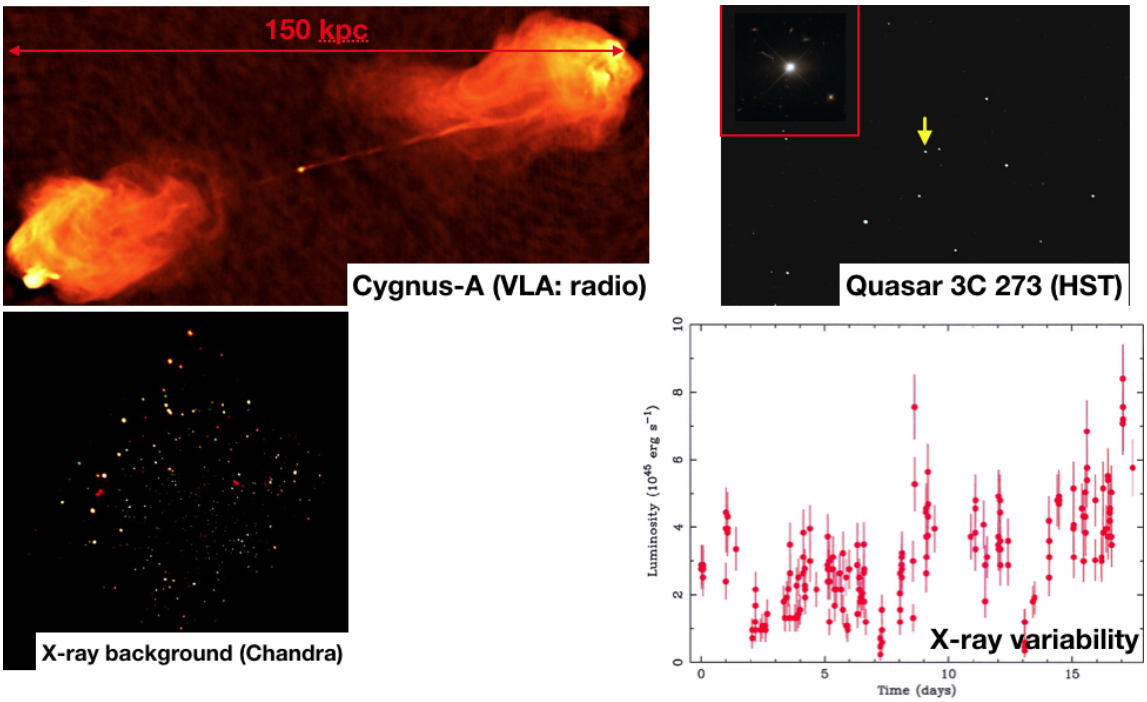

Figure 10.1: Manifestations of AGN and cartoon of the central engine. *Top left panel:* Cygnus A, the brightest radio source in the sky located outside our galaxy, as imaged with the *Very Large Array*. *Top right panel* (main panel): the optical AGN, quasar 3 C273, does not look very impressive in this optical image, until you realise that the distance to 3C273 is about750 Mpc, yet it appears approximately equally bright as the MW foreground star next to it. Assuming this star has the luminosity of the Sun and is at a distance of 1/2 kpc the shows that 3C273 has a luminosity of about $4 \times 10^{12} L_{\odot}$ or ~ 100 times the total luminosity of the MW. The inset in red shows an HST image of 3C273. *Bottom left panel*: The *Chandra deep field* X-ray image of the sky. The two red objects are clusters of galaxies, all others are AGN that are emitting hard X-rays. *Bottom right panel:* Time variation of the X-ray luminosity of AGN PHL1092, from Brandt '99.

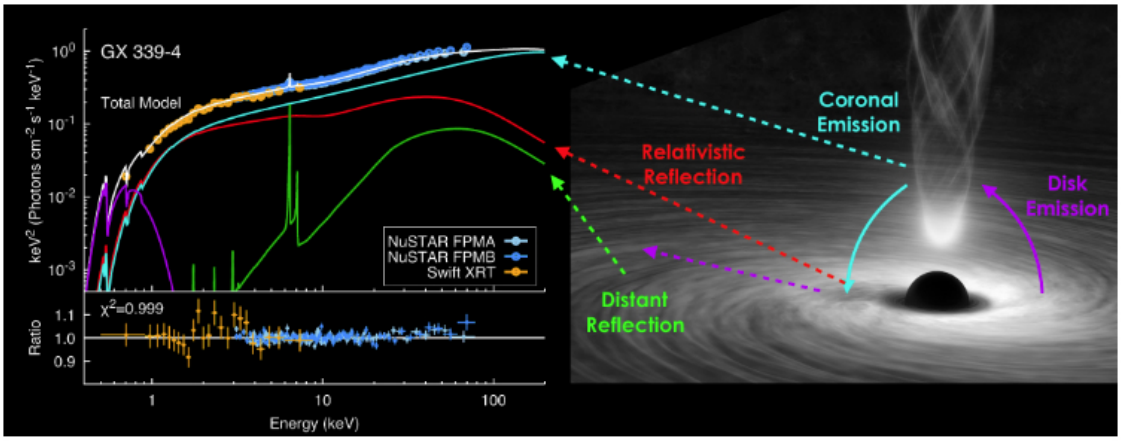

Figure 1: (*left:*) The *Swift* and *NuSTAR* spectrum of the BHXB GX 339−4 during its 2017 outburst, fitted with a Compton continuum (blue); relativistic reflection (red); distant reflection (green); and thermal disk emission (violet). The lower panel shows the fit residuals (García et al. in prep.) (*right:*) Schematic representation for the origin of the spectral components in an accreting black hole. The disk's thermal emission (violet) is Compton scattered into a power-law (blue) by electrons in a hot and compact corona. A fraction of this component illuminates the disk thereby generating the relativistic reflection component (red arrows), as well as a distant reflection component (green). These models are commonly used to constrain the black hole spin, among other important parameters. Adapted from original artwork by NASA/JPL-Caltech/R. Hurt (IPAC).

Figure 10.2: Taken from Garcia '19

(the two reddish objects), and hundreds of unresolved X-ray sources: these are AGN emitting X-ray. The bottom right panel shows that the X-ray luminosity of AGN PHL1092 is **highly variable** in time, with order of unity variations in luminosity occurring over time scales of order of hours or less.

10.3 Central engine of AGN

The rapid variability detected in the X-rays is in fact very surprising. Indeed, it implies that *the size of the AGN engine is of order of light hours*[2]. This argument is based on *causality:* an object that varies on a time-scale t must have an extent smaller than $\sim ct$ - if not, the time-variation would be smoothed out over time and hence smaller in amplitude. The order unity variation in the X-rays on time-scale of $\sim$ hours therefore implies that the engine is of order of light hours - or smaller. How can an object be so small

[2]For comparison, light takes ≈ 1.3 light hours to cover the distance Sun-Saturn.

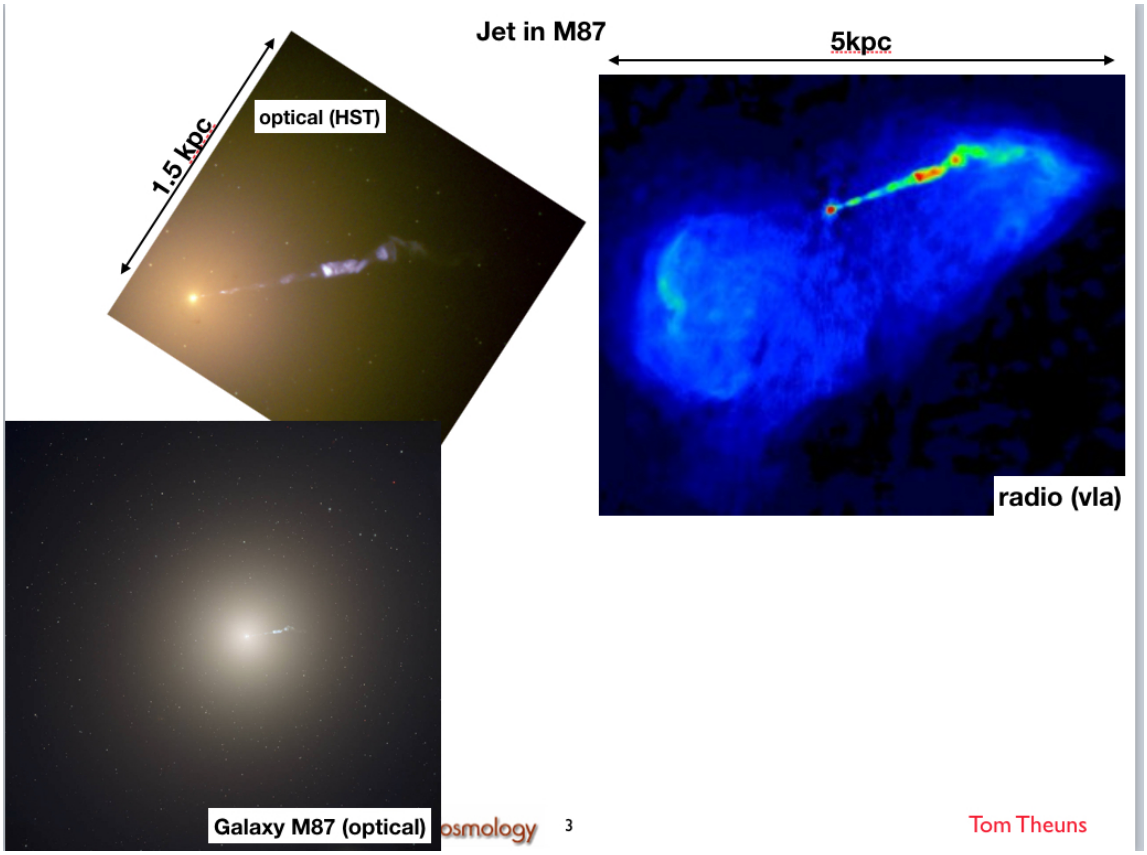

Figure 10.3: Illustrating the jet in galaxy M87. *Bottom panel:* optical image of the galaxy M87 - the central galaxy in the Virgo galaxy cluster. *Top left panel:* HST image in optical light of the jet in M87, tracing the jet all the way to the centre of the galaxy. *Top right panel:* radio image of M87. Notice how the jet is bright in the radio, and seems to 'feed' the radio 'fuzz' surrounding the galaxy.

be *so* luminous[3]?

The ultimate source AGN energy is gravity: gravitational energy is converted into radiation. This should surprise you, since you may recall from the stars' part of this course that the Kelvin-Helmholtz time-scale[4] of the Sun is only $\sim$ 30 Myr. Gas falling towards the SMBH enters an *accretion disc* - because of its angular momentum - and over many orbits slowly spirals in. How this results in the observed spectrum - which extends from radio to gamma rays - is complex, not universally agreed, and not suitable for a first intro to AGN. However if you're interested, do read the next paragraph.

The (nearly) Keplerian disc heats the gas thermally, basically because of its differential rotation and the presence of viscous forces in the disc. This implies that the further in, the hotter the disc. The spectrum of the disc is then a sum of the spectra of each ring in the disc, each approximately a Black Body spectrum. Combined, this results in a broad spectrum of emission peaking around 1 keV, and is thought to be the origin of the optical-UV emission for AGN. Hot electrons above the disc scatter optical-UV photons[5] creating a power-law of high-energy photons - these are observed as X-rays and gamma rays, see also Fig. 10.3. This figure also illustrates the *jet* - a narrow energetic beam that is thought to be launched due to magnetic fields in the disc and which feeds extended radio-lobes.

10.3.1 Making light of gravity: Eddington limited accretion

The[6] basic mechanism that powers AGN is the conversion of gravitational energy into light. To describe this, consider an object of mass $M_{\rm BH}$ (a super massive black hole) accreting mass at a rate $\dot{M}$. If all the rest mass of the accreting gas were converted into radiation, then the luminosity of the object would be[7] $L = \dot{M}c^2$. It is thought that AGN come close to radiating at this

[3]Recall, some AGN have luminosities op to $\sim 10^4$ times the combined luminosity of all stars in the MW.

[4]The time it would take the Sun to radiation all its gravitational energy at its current luminosity.

[5]Through Thomson scattering's relativistic version called Compton scattering - discussed in the L1 lectures.

[6]'Making light of gravity' was the title of Martin Rees' birthday conference. Martin Rees, together with Donald Lynden-Bell, were the first to point to gravity as the ultimate source of AGN power.

[7]Using Einstein's famous $E = Mc^2$ equation.

maximum efficiency, therefore we write

$$L = \eta \, \dot{M} \, c^2 \, ; \quad \eta \approx 0.1 \, . \tag{10.1}$$

The radiative efficiency η is estimated from considerations of the last stable circular orbit around a black hole[8]. The value of $\eta = 0.1$ is much higher than the energy efficiency of *stars* since Hydrogen fusion only manages a meagre efficiency of 0.007, as shown in the Stars section of this course.

Curiously, the mass of the black hole does *not* increase as $\dot{M}_{\rm BH} = \dot{M}$. Indeed, the luminosity is so high[9] that we cannot neglect the 'mass loss' associated with this luminosity - the energetic photons carry energy and hence mass away. The correct relation is then

$$\dot{M}_{\rm BH} = \dot{M} - \frac{L}{c^2} = (1 - \eta) \, \dot{M} = \frac{1 - \eta}{\eta} \frac{L}{c^2} \, . \tag{10.2}$$

When in a steady state, this accretion rate is limited by the *Eddington limit*. Consider a spherical shell of gas around the BH. This shell feels gravity exerted (mostly) by the black hole pulling the gas inward. But radiation streaming through the shell pushes the shell away due to radiation pressure. If the radiation pressure is too large, the net force is outward, and the black hole can no longer accrete. To compute this maximum luminosity - called the 'Eddington luminosity'- we equate the gravitational force to the force exerted by the radiation pressure[10].

For simplicity we consider a shell consisting of Hydrogen gas with number density n (neglecting other elements) which is fully ionised (by the ionising radiation of the black hole). The shell is at distance r from the BH and has thickness dr, with $dr \ll r$. Its mass density $\rho = m_{\rm H} \, n$, the volume of the shell is $4\pi r^2 dr$, and hence the (inward) gravitational force (neglecting self gravity) is

$$F_{\rm G} = \frac{G \, M_{\rm BH} \, (4\pi r^2 \, dr) \, \rho}{r^2} \, . \tag{10.3}$$

[8] In Newton mechanics, circular orbits around a point mass are stable - but this is no longer the case in general relativity when the radius of the orbit is close to the event horizon. The value of $\eta = 0.1$ quoted in the text applies to a Schwarzschild - or non-spinning - black hole.

[9] Again use $E = M \, c^2$. A large luminosity effectively means that the total energy of the object is decreasing rapidly - meaning its mass is decreasing.

[10] Recall the identical derivation in the *Stars* part of this course

The (energy) flux impinging on the shell at distance r is $F_{\rm E} = L/(4\pi r^2)$. Since a photon of energy E has momentum E/c, this corresponds to a momentum flux of $F_{\rm p} = L/(4\pi r^2\, c)$. To calculate the radiation pressure we need to know the interaction cross section of the radiation with the matter. One interaction process is Thomson scattering of photons off electrons, with the wavelength independent Thomson cross section, $\sigma_T = 6.625 \times 10^{-29}$ m^2. The force on a single electron is thus $F_{\rm p}\,\sigma_T$, which, multiplying with the number of electrons in the shells yields the radiation force on the shell as

$$F_{\rm L} = \frac{L}{4\pi r^2\, c}\,\sigma_T\,(4\pi\, r^2\, dr)\, n\,. \tag{10.4}$$

Setting $F_{\rm G} = F_{\rm L}$ yields the expression for the Eddington luminosity,

$$L_{\rm Edd} = \frac{4\pi G\, M_{\rm BH}\, c\, m_{\rm H}}{\sigma_T} \approx 3.3 \times 10^{12}\, \frac{M_{\rm BH}}{10^8\, M_\odot}\, L_\odot\,. \tag{10.5}$$

Given the extremely high luminosities we observed for AGN, values of $10^{12} L_\odot$ or more, shows that the black holes that power bright AGN have masses of order $10^8\, M_\odot$, a conclusion first reached by Lynden-Bell (1969)

10.3.2 Growth of black holes - Salpeter time

Once a seed black hole exists, it can grow in mass through accretion and merging with other black holes, shining as an AGN as described above. How black hole seeds form is unclear: they may simply be remnants of massive stars ending their lives as black hole remnants, or may form through a completely different route, for example direct collapse of a gas cloud.

The Eddington luminosity also limits the rate at which a BH can grow as follows. Suppose the BH always accretes at its maximum rate, *i.e.* has luminosity $L = L_{\rm Edd}$. Substituting Eq. (10.5) into Eq. (10.2) yields

$$\dot{M}_{\rm BH} = \frac{1-\eta}{\eta}\frac{M_{\rm BH}}{\tau_S}\,; \quad \tau_s = \frac{c\sigma_T}{4\pi G m_h} \approx 4.5 \times 10^8 {\rm yr}\,. \tag{10.6}$$

Solving the differential equation for $M_{\rm BH}$ shows that the mass grows exponentially in time, $M_{\rm BH} = M_{\rm BH}(t=0)\exp(t/t_S)$, with a characteristic time $t_s = \eta/(1-\eta)\tau_S$ called the **Salpeter time**.

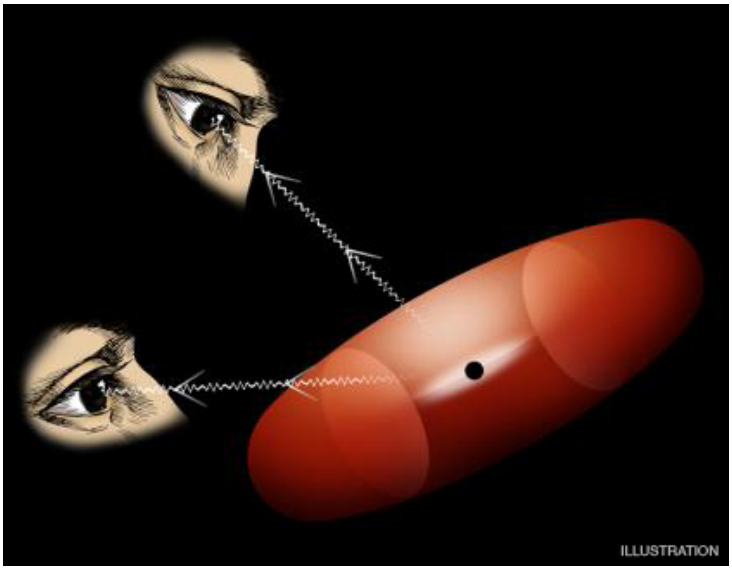

Figure 10.4: Depending on the orientation of the observer with respect to the torus of gas surrounding the black hole, the AGN may look different. Seen from above, the observer sees very close to the centre and may see an optically bright source - a QSO. Seen from the side, optical/UV light gets obscured and the observer does not see a QSO. Figure credit: Chandra observatory.

The mass of the SMBH in the most distant AGN currently known[11], ULAS J1342+0928, is estimated to be $M_{\rm BH} \approx 8 \times 10^8 {\rm M}_\odot$. This AGN is observed at a redshift of $z = 7.45$, at which point the age of the Universe was only 0.7 Gyr. Assuming this BH continuously accreted at its Eddington rate and that the seed formed at the time of the Big Bang - both quite unlikely - its seed mass was at least $\sim 670 {\rm M}_\odot$ - suggesting whatever seeds a SMBH is *much* heavier than a stellar remnant[12]. See the workshop for exercises on this aspect of BH growth.

10.3.3 Unification schemes

If mass accretion powers AGN, why is there such a variety of observational manifestations of the AGN activity? It is thought that the accretion disc is surrounded by a bigger structure in the shape of a doughnut - called a 'torus'. The presence of this torus may change our view of an AGN, depending on its orientation compared to our sightline to the AGN, as illustrated in Fig. 10.4. For example, when observing the AGN nearly perpendicular to the plane of the torus, we have a direct view of the accretion disc and hence can detect the optical/UV light emitted and infer the presence of a QSO. But if our sightline is more edge on, then the torus may obscure the optical light. However that still does not explain why some AGN have huge extended radio lobes and others don't. We still have much to understand!

10.4 Evidence for super massive black holes in galaxies

Only a small fraction of galaxies hosts an AGN. In the local Universe, $\approx 1\%$ or so of massive galaxies hosts a bright AGN, with around 10-20% showing some evidence for weak AGN activity. However, all these massive galaxies do host a SMBH. Clearly this requires that most SMBH are not active at any one time - the SMBH is not active because it is starved of gas. Feed the SMBH some gas and the SMBH will light up and become an AGN. It seems very likely that galaxies then go through relatively short phases where their SMBH is active and prolonged phases were it is not, with the active phase typically 100 times shorter than the inactive phase to obtain the 1% AGN duty cycle.

The evidence that galaxies host a SMBH even when the AGN is off is based on dynamics - basically observing large speeds for objects getting close to the centre, where they are accelerated by the gravitational attraction of the SMBH.

[11] Your lecturer used to hold the record for the discovery of the most distant known AGN, ULASJ112001.48+064124.3, at a redshift of $z = 7.09$.

[12] And if the BH accreted at less than its maximum rate or formed later in time, then the seed mass must, of course, have been even higher.

10.4.1 The Milky Way's SMBH

Evidence for the presence of a SMBH in the MW is exquisite. Recording the positions of stars in the centre of the MW over many years, it became possible to reconstruct the orbits of these stars, see Fig. 10.5. The observed speeds of several 1000 km s^{-1} and the large measured accelerations require the presence of a very massive object - a SMBH with mass $\sim 4 \times 10^6 M_\odot$. These observations were done in the IR, since otherwise the stellar light would be absorbed by the dust in the MW's disc - and hence not observable to us.

One of these stars - called 'S2' - ventured so close to the SMBH in 2018 that it was possible to measure gravitational redshift of H and He lines in its spectrum - that is, photons emitted by the star losing measurable amounts of energy and hence changing their wavelengths having to climb out of the gravitational potential of the SMBH. The loss of energy means that the wavelengths shift to longer wavelengths - to the red, hence 'gravitational redshift'.

10.4.2 The 'shadow' of the black hole in M87

Black holes are called black since 'not even light can escape from them'- more accurately, the escape speed from their event horizon is equal to the speed of light. Light emitted from within the event horizon cannot escape.

The presence of the SMBH distorts ('bends') space-time around it, so that from the point of view of a distant observer even the path of light appears curved. In the most extreme case, light can orbit the SMBH. This is an extreme instance of 'gravitational lensing' discussed in the next chapter. The scale at which extreme lensing occurs is of order of the Schwarzschild radius (event horizon) of the BH,

$$R_S = \frac{2GM}{c^2} \approx 126 \, \frac{M}{6.4 \times 10^9 M_\odot} \, \mathrm{AU} \, , \tag{10.7}$$

where the numerical value uses the mass of the SMBH in M87. At M87's distance, $d \approx 16.4$ Mpc, the angle[13] under which we see R_S is $\theta = R_s/d \approx 10^{-5}$ arc sec.

Given this the tiny angular extent, it would seem impossible to resolve M87's event horizon. However, this is exactly what the 'Event Horizon Telescope' managed, using Very Long Baseline Interferometry combining several

[13] Or approximately maximal angular extent of Durham cathedral - as seen from Saturn!

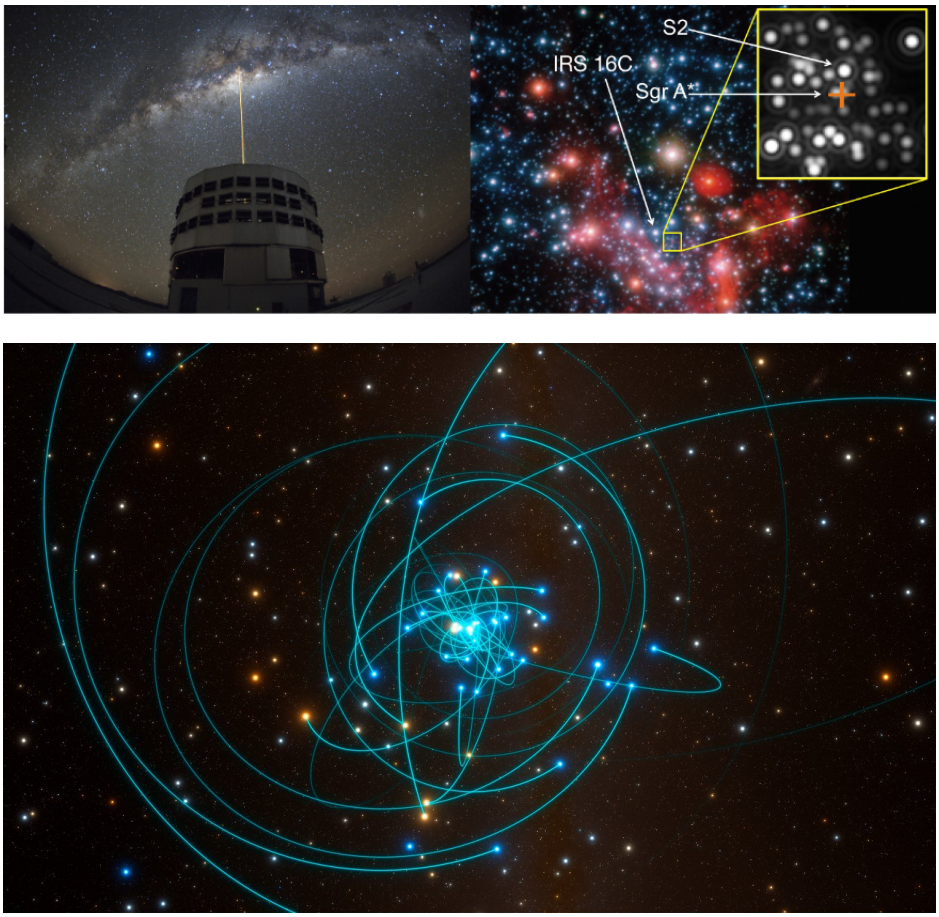

Figure 10.5: *Top left panel:* ESO's VLT telescope firing its laser to enable adaptive optics to correct for atmospheric seeing. *Top right panel:* VLT's IR image of the centre of the MW. The object labelled 'Sgr A' is the MWs SMBH - it is barely detectable in the IR. The bright sources in the inset are massive stars orbiting Sgr A; star 'S2' is also indicated. *Bottom panel:* reconstructed orbits of stars around Sgr A over the past 26 years. From ESO's *Messenger*.

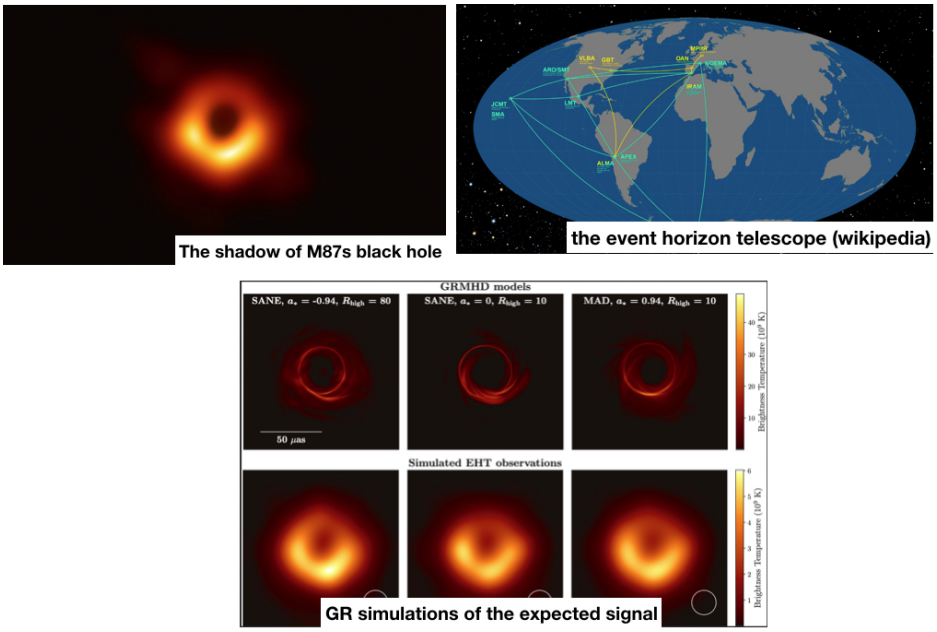

Figure 10.6: *Top left panel:* light with wavelength ≈ 1 mm detected from the immediate vicinity of M87's SMBH by the Event Horizon Telescope. *Top right panel:* radio dishes that make-up the telescope. *Bottom panel:* simulations of the expected light taking into account relativistic effects close to the event horizon, for different values of the BH's spin (left to right). The intrinsic signal is shown in the top row, the signal taking into account the telescope resolution and sensitivity is shown in the bottom row. From The event horizon telescope.

radio telescopes across earth. Observing at a wavelength of $\lambda \approx 1$ mm, they detected a ring of light around M87, see Fig. 10.6. Taking the earth's diameter, D, as the baseline yields an angular resolution of $1.2\lambda/D \approx 2 \times 10^{-5}$ arc seconds at a wavelength of $\lambda = 1$ mm, meaning the telescope can resolve scales of order R_S. This astonishing feat shows convincingly that the object in the centre of M87 is really a black hole - that is, a relativistic object with an event horizon [14].

10.4.3 SMBHs in other galaxies

The techniques used to identify the SMBH in the MW and M87 cannot be applied to more distant galaxies. Evidence for SMBHs in them is of course less convincing and comes from other observations.

- The gravity of the BH affects orbits of stars close to it. Even if individual orbits cannot be measured, this does lead to an enhanced stellar density close to the SMBH - a 'cusp' (rapid increase) of the surface brightness. This can be detected in nearby galaxies using HST imaging.
- *Maser* emission. A maser is the microwave equivalent of a laser[15]. It may already be surprising that maser emission occurs naturally in the ISM of galaxies. Sometimes, the maser emission originates close to the centre, allowing us to measure the mass of the central concentration (SMBH) orbited by the masering gas.
- Similar to maser emission, it is possible to estimate the mass of the BH from the line-widths of emission lines originating close to the BH. Assuming the large widths of these lines is due to the orbital motion of the emitting gas allows us to infer the black hole mass[16]

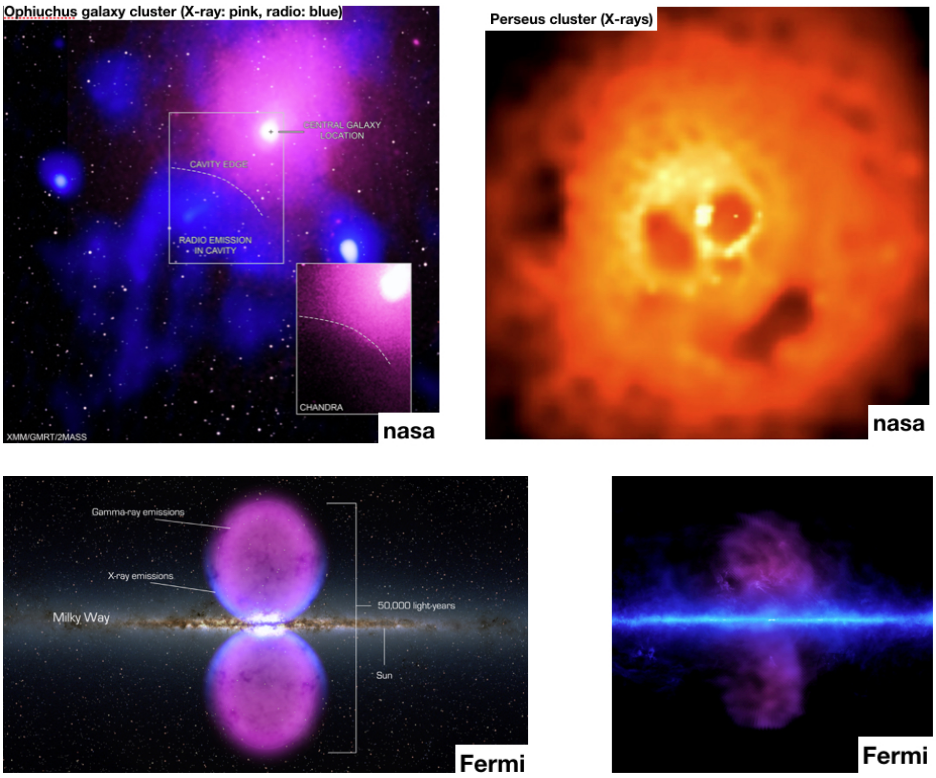

Figure 10.7: *Top left panel:* giant shocks detected in the Ophiuchus galaxy cluster. The energy required to create these shock is estimated at 5×10^{61} ergs equivalent to 5×10^{10} simultaneously supernova explosions - motivating the press to call this the biggest explosion since the Big Bang. *Top right panel:* X-ray image of the Persues cluster of galaxies. The large dark patches in the X-rays (resembling eye sockets of a skull) are due to the hot X-ray gas being displaced by relativistic particles injected by the AGN. *Bottom left panel:* cartoon of the gamma ray bubbles detected above and below the plane of the Milky Way by the Fermi satellite. These 'Fermi' bubbles are likely relics of past AGN activity of the MWs SMBH. *Bottom right panel:* actual Fermi data.

10.5 Impact of AGN on their host galaxy

The large amounts of energy that AGN inject into their surroundings is thought to affect the host galaxy. Exactly how this happens is currently not well understood - probably the jet inflates radio bubbles, filling them with relativistic particles, and these bubbles prevent gas from cooling. With no more gas able to cool, the galaxy cannot make any more stars. Examples of giant cavities created by AGN in the hot gas of cluster of galaxies - where they are visible due to the *absence* of the hot X-ray gas- are shown in Fig. 10.7. With AGN activity being prevalent in galaxies with mass $\geq 10^{10.5}\mathrm{M}_{\odot}$, this might explain the rapid drop in galaxies more massive than this that we noticed in Fig. 9.3: when the SMBH can turn into an AGN, the galaxy stops making stars. More speculative, this is probably also why massive galaxies tend to be elliptical: the SMBH they host prevents star formation when it switches to the 'AGN-on' state.

The SMBH in the Milky Way is thought to be a puny version of the energetic beasts seen in the top panels of Fig. 10.7. The bottom panels shows the striking Fermi bubbles, detected in gamma rays, that are thought to be relics of past activity of Sgr A - the SMBH in the MW.

[14]The other convincing evidence for the existence of black holes comes from the detection of gravitational waves emitted during the merging of stellar mass black holes. A discovery too exciting not to at least mention here.

[15]Light amplification by stimulated emission radiation

[16]Note unlike measuring the enclosed mass of the Milky Way from the observed rotation curve.

10.6 Summary

After having studied this lecture, you should be able to

- describe observational manifestations of AGN.
- describe briefly the current paradigm for energy generation in AGN (mass accretion through a disc onto a SMBH)
- relate the mass accretion rate to luminosity for AGN, and apply the concept of Eddington limited accretion to the maximum growth rate of black holes
- describe three arguments that suggest the presence of SMBHs in galaxies.

Chapter 11

Gravitational lensing

CO §28.4

Any object with mass causes the distortion ('warping') of space-time around it leading to the deflection of light rays - this phenomenon predicted by the theory of General Relativity (GR) is called gravitational lensing (GL). In addition to **light deflection**, GL may also create **image distortions, changes in brightness and multiple images of a lensed object** - all of which have been observed. Typically the effects are small meaning that our view of the Universe is not greatly distorted by GL. On the largest scales, GL of the Cosmic Microwave Background (CMB) has been used to constrain cosmology. It has also been used to measure the masses of galaxies and clusters of galaxies from the distortion introduced by GL of background galaxies, and to study faint distant galaxies (taking advantage of the increase in brightness). On smaller scales, GL can constrain the particle nature of the dark matter and has been used to image the event horizon of a black hole (Chapter 10). GL in the solar system provided the first verification of GR.

11.1 The lens equation

We first attempt to compute the deflection angle in Newtonian mechanics. The observed deflection of light rays grazing the Sun agrees with the GR prediction but disagrees with our 'Newtonian' estimate, providing the first confirmation of GR.

11.1.1 Bending of light

The shape of the orbit of an unbound test particle around a mass M is hyperbolic[1]: the incoming particle is 'deflected' by the encounter with the more massive object by an angle α which depends on M, the initial speed, v, and impact parameter, b. Provided the deflection angle is small,

$$\alpha = \frac{2GM}{b\,v^2}\,, \tag{11.1}$$

as derived in the Appendix (Eq. 11.10). Notice that this does not depend on the mass of the test particle. We now make the rather bold assumption that this applies to light as well and simply substitute v by c to obtain the 'Newtonian' prediction for GL.

For starlight grazing the Sun, we take $M = M_\odot$, $b = R_\odot \approx 7 \times 10^5$km, the radius of the Sun, to find [2] $\alpha = 0.87$ arcsec. Our numerical value agrees with Einstein's 1911 estimate as well as the calculation by *Johann Soldner* almost a century earlier.

Curiously, the answer is wrong! After developing the theory of General Relativity (GR), Einstein recalculated the deflection angle for light as,

$$\alpha = \frac{4GM}{b\,c^2}\,, \tag{11.2}$$

exactly *twice* the Newtonian prediction[3] of Eq. (11.1). The difference in prediction between Newtonian physics and GR motivated Eddington to lead an expedition to measure α during 1919's total solar eclipse[4]. Eddington's measurements were revealed during a talk at the Royal Society, and confirmed GR's prediction[5].

[1] For example the motion of an unbound asteroid around the Sun.

[2] In radians, $\alpha = 4.2 \times 10^{-6} \ll 1$, validating our assumption.

[3] The intuitive reason for the larger deflection takes account of *time-dilation*: it is as if the photon spends more time near the lens, so that the gravitational force has more time to act. Accounting for gravitational time-dilation - 'clocks appear to run slower when in a gravitational potential' - is essential to obtain the required precision in Global Positioning Systems (GPS).

[4] The expedition measured the apparent movement of stars on the sky for sight lines grazing the Sun compared to more distant sight lines. This is (only) possible during a solar eclipse, when the moon blocks most of the Sun light, allowing the observation of stars close in angle to the Sun.

[5] But subsequent measurements found less convincingly in favour of GR - probably due

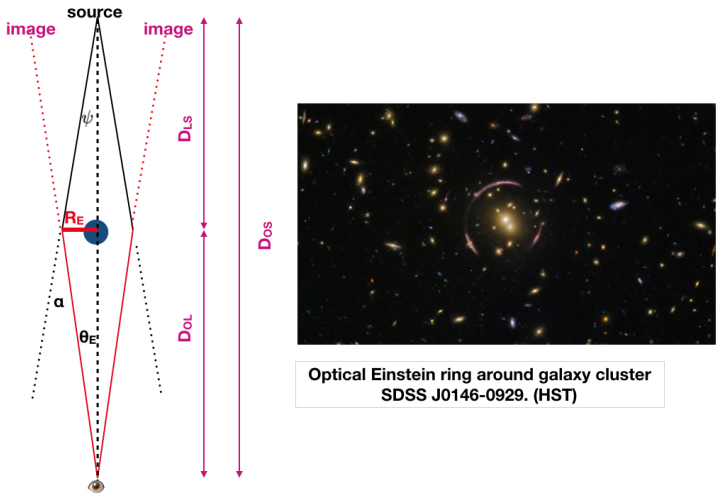

Figure 11.1: *Left panel:* Lensing configuration in a case in which observer, lens and source are co-linear. The source is imaged in a ring, called Einstein ring. *Right panel:* (Approximate) Einstein ring, when a distant galaxies is lensed by a foreground cluster of galaxies, SDSS J0146-0929..

11.2 Deflection through gravitational lensing

We can compute the image formed by GL for some simple configurations.

11.2.1 Point-like lens and source

Co-linear

Consider the special case when observer, lens and source are co-linear as illustrated in Fig. 11.1. Introducing the distances observer-lens, observer-source, and source lens, as in the figure, we find

$$\theta_E = \frac{R_E}{D_{OL}}; \quad \psi = \frac{R_E}{D_{LS}}; \quad \psi + \theta_E = \alpha; \quad \alpha = \frac{4GM}{R_E\, c^2}\,, \tag{11.3}$$

to measurement errors. Of the other two predictions that Einstein made - gravitational redshift was not detected and some astronomers were not convinced by the explanation for the perihelion precession of Mercury. As a consequence, the theory of GR was initially not as widely accepted as you may have thought, see this recent review.

where M is the mass of the lens. The first three relations are simple geometry[6], the last one is the lensing equation for the deflection angle from Eq. (11.2). Combining these yields

$$\theta_E = \left(\frac{4GM}{c^2}\right)^{1/2} \left(\frac{D_{LS}}{D_{OS}D_{OL}}\right)^{1/2} . \tag{11.4}$$

Notice that in this case the observer sees the source lensed in a **ring** called an *Einstein ring*[7]. *If* we can measure the distances observer-lens and observer-source, then the angular extent of the Einstein ring yields the mass of the lens, M.

This is a very powerful way of measuring masses of astrophysical objects, because we need not make any assumptions about what form this mass takes (baryonic or dark, for example), nor do we need to assume anything about its dynamical state.

The panel to the right shows an *approximate* optical Einstein ring, where a background galaxy is lensed by a foreground cluster. Neither the galaxy nor the cluster are point-masses: as far as lensing is concerned, the Einstein ring is a measure of the mass enclosed by R_E, the 'Einstein radius'.

Not co-linear

Figure 11.2 illustrates the configuration. Starting with simple geometry we find under the assumption of small angles, that

$$\theta_S = \theta_I - \beta = \theta_I - \frac{D_{SL}}{D_{OS}}\alpha; \quad R = \theta_I\, D_{OL}\,. \tag{11.5}$$

Combined with the lensing equation, $\alpha = 4GM/(Rc^2)$, yields a quadratic equation for the observable angle, θ_I, between the directions lens-image,

$$\theta_I^2 - \theta_I\theta_S = \frac{4GM}{c^2}\frac{D_{LS}}{D_{OS}D_{OL}} \equiv \theta_E^2\,, \tag{11.6}$$

[6]These relation are valid in the small-angle approximation. Notice that this approximation does not apply to the *cartoon* illustration, but works very well in real applications.

[7]In the cartoon, the observer sees *two images* in the plane of the paper. However, now rotate the plane along the axis observer-source: the observer sees the two images *in each of these planes* - i.e. they see a ring.

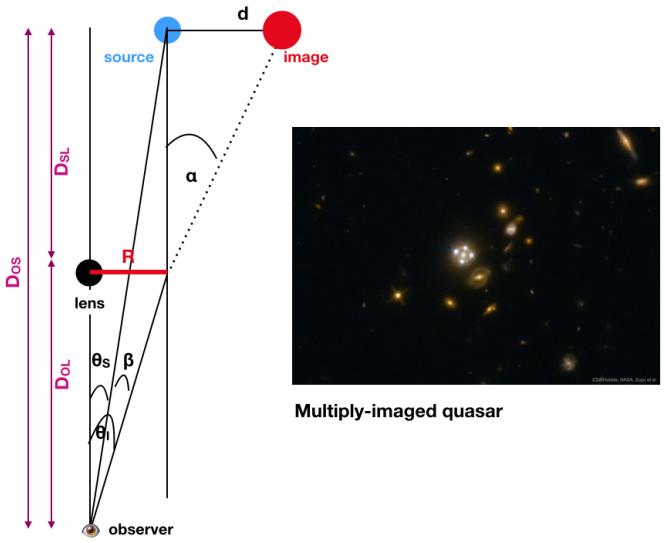

Figure 11.2: *Left panel:* Lensing configuration in a case in which observer, lens and source are not co-linear. *Right panel:* 4 images of a quasar lensed by an intervening galaxy.

using Eq. (11.4) for θ_E. This equation has two roots[8], $\theta_{I,\pm}$, which can be used to determine the two unknowns, θ_S and θ_E,

$$\theta_{I,+} + \theta_{I,-} = \theta_S; \quad (\theta_{I,+} - \theta_{I,-})^2 = (\theta_{I,+} + \theta_{I,-})^2 + 4\theta_E^2 \,. \tag{11.7}$$

Measuring the directions to the two images yields θ_E which yields the mass of the lens, M, provided all distances can be measured.

11.2.2 More complex lens-source configurations

More complex configurations occur when the source or the lens are spatially extended (for example the source is a galaxy, the lens a cluster of galaxies). The fact that the deflection angle may vary across a spatially extended source leads to image distortions. An extended *lens* may generate more than two images of a source, as illustrated by the 4 images[9] produced when a back-

[8] One root is positive, one is negative meaning the image is seen to the left of the source, under the convention that angles are positive when measured clockwise from the direction to the lens.

[9] The fifth central object is the lensing galaxy itself.

ground quasar is lensed by (the extended dark matter halo of) a foreground galaxy, shown in the right panel of Fig. 11.2.

11.3 Applications of gravitational lensing

Previously we only discussed how GL deflects light. In addition to displacing the image, GL may enlarge the image as compared to the source, make it look brighter, and create multiple images of the same source. Here we will show examples of these effects and discuss their applications.

11.3.1 GL in clusters of galaxies

GL is strong for sight lines passing close to the centre of a cluster of galaxies, causing the appearance of multiple images of the same galaxy, stretched in the form of tangential (and sometimes radial) arcs. The number of images, how strongly they are stretched and even their relative brightness, all depend on the projected mass distribution. These effects are referred to as **strong gravitational lensing**. Strong GL makes it possible to *infer* the mass distribution of the lens - i.e. the cluster - by modelling all of these effects. Fig. 11.3 shows a cluster (left) and its projected mass distribution (right) in the top panel. Gratifyingly, estimates of cluster masses using GL are in reasonable agreement with those based on the dynamics of galaxies or as inferred from the hot X-ray gas in clusters discussed in Chapter 8.

The effects of GL are weaker for background galaxies with lines-of-sight at larger impact parameters from the cluster. However, GL will still distort images slightly, imaging an intrinsically round source into a slightly elongated image. For any individual galaxy, it is not possible to blame the image elongation and its orientation on GL, since the galaxy that is lensed may be intrinsically elongated. However, GL will induce similar elongations and image orientations for galaxies that are close to each other on the sky - meaning distortions are correlated. This is called **weak gravitational lensing**: measuring the correlated distortions of galaxies on the sky, to infer the required lensing mass distribution.

Back to strong GL. Cluster mass distributions have characteristic lines where the effects of GL are very strong, with images of background galaxies factors of several brighter than they would be in the absence of lensing. These characteristic lines are called *caustics* and the reason they appear is similar

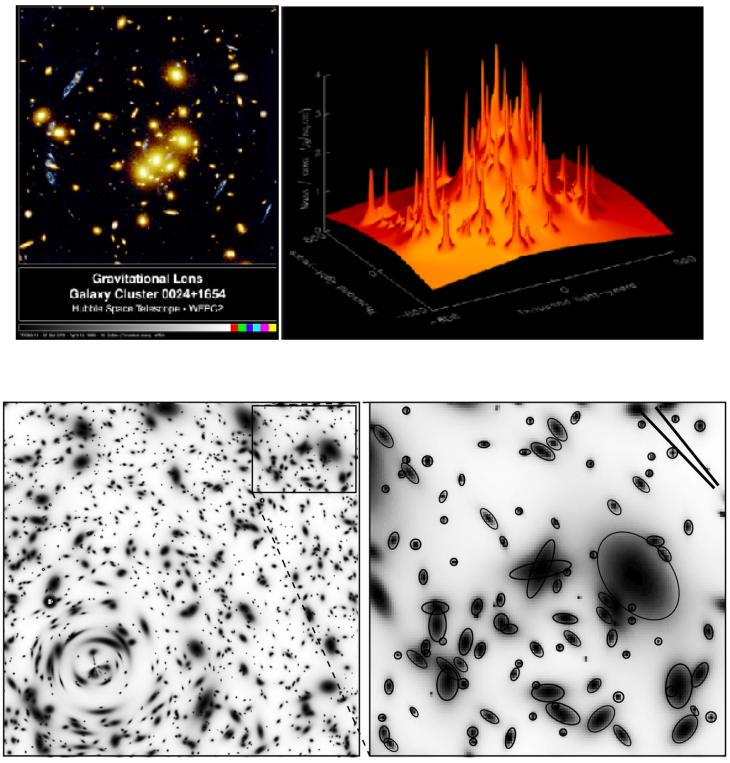

Figure 11.3: *Top left panel:* Optical image of a cluster of galaxies. Modelling of the multiple images of background galaxies due to GL makes it possible to infer the project mass distribution, which is shown in the *top right panel*, from Freese 2009. *Bottom panel:* At large impact parameter, GL does not produce multiply-imaged galaxies any more. However, it is still possible to infer statistically small stretchings in the shapes of background galaxies caused by GL, since these are correlated on the sky. This effect is exaggerated in this cartoon: close to the cluster's centre (bottom left), GL strongly deforms images of background galaxies into tangential arcs. Further away from the centre, distortions are much less but can be detected because they are *corrrelated* on the sky - galaxies close on the sky are stretched in similar directions.

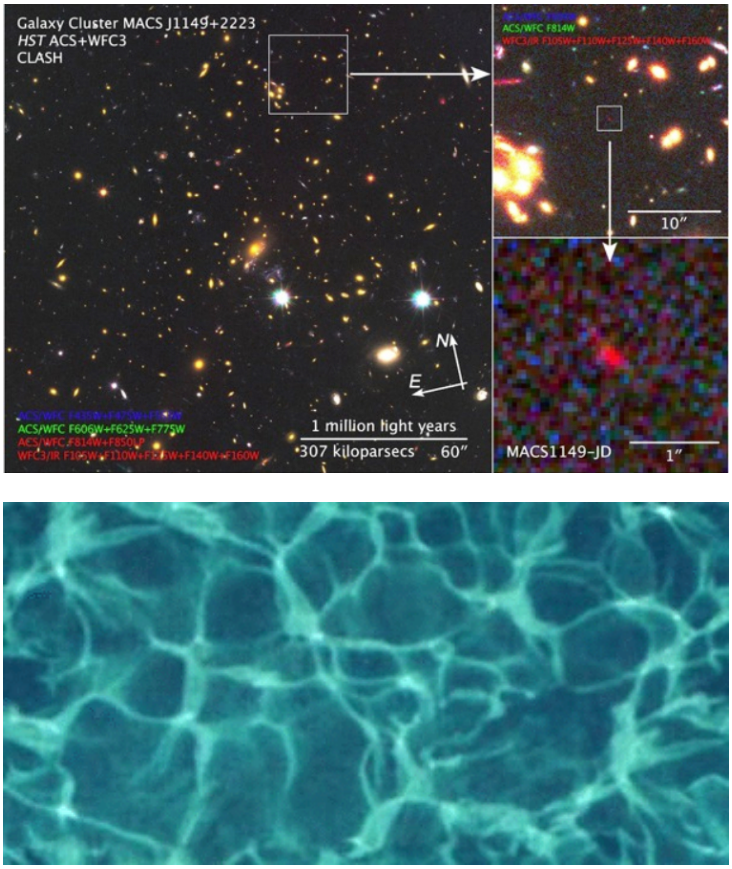

Figure 11.4: The physics of lensing caustics. *Bottom panel:* caustics - lines where sun light gets focussed on the bottom of a swimming pool. *Top panels:* Identifying very distant galaxies using GL near caustics in galaxy cluster MACS J1149-2223 (see text for details). From Heap 2015.

to the appearance of bright lines at the bottom of an outdoor swimming pool due to the focussing of sun light. This is illustrated in Fig. 11.4. Once a mass model for a cluster is constructed, it is possible to find where the caustics are - that is, where the lines are along which background galaxies will be strongly lensed with their image much brighter than it would be without GL. With flux enhancements by factors of 10-30, this enables us to study very distant galaxies that would be too faint to study without the help of GL. The top right panel identifies a faint red galaxy which may be one of the most distant galaxies ever identified. Without GL by the cluster, it would be impossible to find, let alone study, such infant galaxies.

11.3.2 GL and the nature of the dark matter

The nature of the dark matter (DM) is currently unknown with some type of elementary particle not part of the Standard Model of particle physics leading the pack of candidates. Because GL is agnostic about the nature the mass that causes lensing, it might be possible to constrain the DM's nature.

Dark matter in the Milky Way's halo

The MACHO collaboration proposed the following ingenious method to probe the MW's dark matter halo. Assume that the DM takes the form of Massive Astrophysical Compact Halo Objects - MACHO's - for example compact objects with masses of the order or the mass of a star or planet. Such MACHO's in the MW's halo would GL stars in other galaxies. The effect is generally weak, but occasionally, a MACHO might be passing *almost exactly* in front of a background star, and we may be able to detect the star becoming brighter by a factor of a few for a short time - due to GL. However, even if the whole dark matter halo of the MW was composed of such hypothetical objects, then the chance that any one background star is lensed is disappointingly small, 1 chance in 10^7. However, the collaboration realised that, by observing stars in the Large Magellanic Cloud (LMC) - a satellite of the MW (see chapter 5 on the local group) - it would be possible to observe of order 10^7 stars each night so we could expect of order 1 star to be lensed by a MACHO at anyone time. Notice that the brighting is time-dependent, because the lens moves - as does the observer and the source.

However, how can we conclude that a star gets brighter due to GL rather than simply because it is a variable star? An important aspect of GL is that

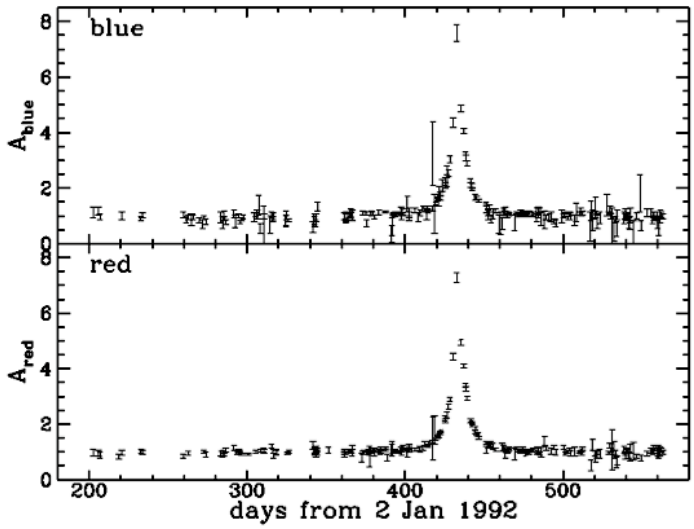

Figure 11.5: Flux of a star in the Large Magellanic Cloud (LMC, a satellite galaxy of the Milky Way) as a function of time, for a blue filter (top) and a red filter (bottom). The flux is normalised to its value at the start of the observations. Around day 420, the flux of the star increases by a factor of ~ 8 in both filters, due to gravitational lensing caused by an object moving across the sight line to the star.

the lensing properties are *independent* of the wavelength of light. Therefore if a star gets lensed, its brightness in a red and in a blue filter will increase by the exact same amount - whereas if it were an intrinsically variable star that would generally not be the case. Secondly, a variation due to lensing has a particular shape **and** a given star will vary *only once*. Combining all this should make it possible to recognise whether the flux of a star varies because it is intrinsically a variable star or because it is undergoing GL.

The collaboration detected several lensing events, one of which is shown in Fig. 11.5. However, in each case the lens was also detected in the image: it was not a dark matter MACHO but rather another star in the LMC. After years of observing, no lensing by MACHO's was detected. This negative result proved that the MW's dark matter halo*cannot* be dominated by MACHO's in the mass range $2.5 \times 10^{-7} \leq m/\mathrm{M}_{\odot} \leq 10^{-1}$.

Dark matter in other halos

The dark matter may clump on small scales producing DM halos with substructures. This is the case in the popular 'cold dark matter' (CDM) model, illustrated in the *left panel* of Fig. 11.6, which shows a numerical simulation of such a halo. Some of these substructures may host satellites of the main galaxy in such a DM halo. However, the CDM model predicts the existence of far more substructures than there are observed satellites around galaxies. Therefore many DM halos are dark: they do not host a visible galaxy.

How can we test whether real DM halos also contain so many dark substructures? Consider a case where a background galaxy is lensed into several tangential arcs due to GL by a massive intervening galaxy, as in the *right panel* of Fig. 11.6. The blue arcs are images of the *same* background galaxy, galaxies labelled G1-G3 are part of the same halo that causes the GL. In this particular image, there is another galaxy, labelled G4, that happens to fall on top of the right arc. As a consequence, the right arc is distorted by GL from G4 - but not the left arc.

This situation illustrates how we could detect dark DM substructures: examine in great detail the arcs produced by strong GL, and test whether we detect deformations of one image of the lensed galaxy but not the other(s). If such distortion is not caused by a detectable galaxy, then we may have found the first dark DM structure. No such cases have been found (yet). If you were to detect this, then drop everything you are doing to announce your discovery to the world!

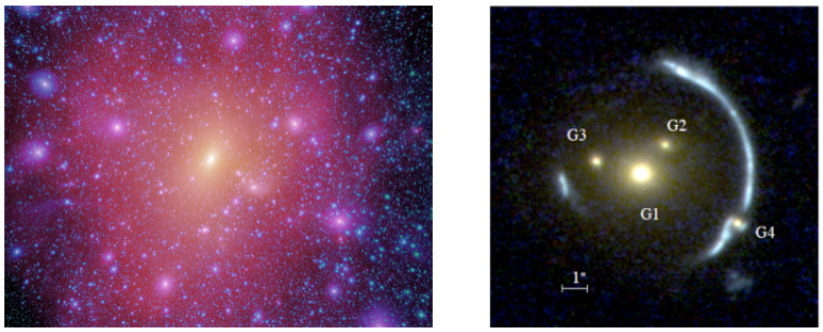

Figure 11.6: *Left panel:* theoretical density structure of a halo in case the DM is of the 'cold dark matter' type, from the Aquarius project. Regions of higher DM density are bright and yellow. This DM model predicts that the halo has a general smooth spherically symmetric density profile of the form $\rho(r) \propto 1/r^2$ around a centre that is close to the centre of the image, but on top of this contains many thousands of 'substructures' (the blobby density enhancements). Some of these may host satellite galaxies, but most must be completely dark since there are far fewer satellites observed compared to the number of substructures that are predicted. *Right panel:* Observed HST image of a background galaxy doubly imaged in the form of tangential arcs (the blue arcs) due to GL, lensing galaxies are labelled G1-G3. If one of the arcs happened to fall on top of a DM substructure, it would be additionally deformed. In this particular case, extra lensing is due to galaxy labelled G4, which distorts the right arc but not the left arc. Image from Lin 2009.

11.4 Summary

After having studied this lecture, you should be able to

- derive the Newtonian lensing equation (11.1)
- 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.

Appendix: Gravitational deflection of a test mass

Consider the case were an unbound test mass (mass m) is gravitationally deflected by a more massive object (mass $M \gg m$). The orbit of m is of course hyperbolic. Here we examine the case where the deflection is small and we want to compute the *deflection angle* α - the angle between the initial and the final velocity. The geometry is illustrated in Fig. 11.7.

The deflection is caused by the component of the gravitational force perpendicular to the orbit, $F_\perp$. According to Newtonian mechanics

$$m\frac{dv_\perp}{dt} = F_\perp = \frac{GMm}{r^2}\cos(\theta) = \frac{GMmb}{r^3}\,. \tag{11.8}$$

Notice that the the solution $v_\perp(t)$ does not depend on m .

A first order-of-magnitude estimate follows by evaluating the force at closest approach, $F_\perp \approx GmM/b^2$, and estimating that the force acts over a time scale of order $\Delta t \approx b/v$. The induced velocity to the left is then $v_\perp \approx F_\perp \Delta t/m = GM/(bv)$, so that the deflection angle $\alpha = v_\perp/v \approx GM/(bv^2)$.

We can do better by simply integrating Eq. (11.8):

$$v_\perp = \int_{-\infty}^{\infty} \frac{GMb}{v(b^2+x^2)^{3/2}}\,vdt = \frac{GM}{bv}\int_{-\infty}^{\infty}\frac{dx'}{(1+x'^2)^{3/2}} = \frac{2GM}{bv}\,, \tag{11.9}$$

so that the deflection angle

$$\alpha = \frac{2GM}{bv^2}\,. \tag{11.10}$$

In this derivation I used the fact that $x = vt$ in the first step, and I set $x' = x/b$ in the second. Notice that the derivation assumes that v is constant, consistent with the assumption of a small deflection. To obtain the numerical value of the definite integral, you may want to change variables once more setting $x' = \tan(\phi)$.

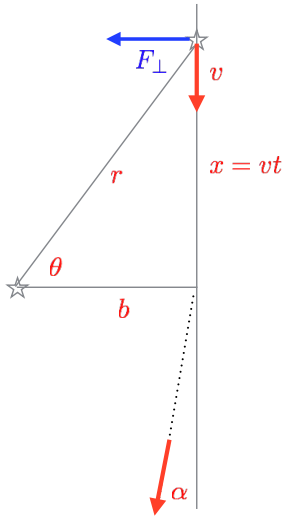

Figure 11.7: Geometry illustrating the deflection of a test mass by a more massive deflector. The impact parameter is b and the deflection angle α.

Contents