

# Astrophysics ASTR3415

## Introduction to Cosmology

### 1 Section I: galaxies and their properties

The Sun inhabits a vast system of stars which we call the **Milky Way Galaxy**. From the Earth, the Milky Way appears to the unaided eye as a thin band of diffuse light stretching across the sky. Telescopes resolve this band of light into individual stars. The Milky Way appears as a thin strip because most of the stars in the Galaxy are found in a thin, flat disc with a central bulge. From our position embedded *within* the disc we see it edge-on.

The visible Milky Way is just our own galaxy seen 'edge-on'.

#### 1.1 The Milky Way: facts and figures

mass *	$1.4 \times 10^{11} M_{\odot}$
radius of Sun's orbit	8.7 kpc
disc diameter	30 kpc
number of stars	$4 \times 10^{11}$
bulge diameter	10 kpc
luminosity	$2 \times 10^{10} L_{\odot}$
halo diameter	100 kpc (approx)
absolute magnitude	-20.5 ( $M_V$ )
disc thickness **	1 kpc
rotation period **	$2.5 \times 10^8$ y

\* inside Sun's distance from centre of the Galaxy.

\*\* at Sun's distance from centre of the Galaxy.

#### 1.2 Distance indicators

We can determine the distances to stars in the Galaxy using **distance indicators**, such as certain types of **variable stars**, or the **annual parallax** of stars (their change in apparent position over the year as seen from Earth). If we know the luminosity of a star, and measure its flux at Earth, we can estimate its distance because the flux drops off as the **inverse-square** of the distance.

If we express this idea in terms of magnitudes we get the **distance modulus** formula. This relates the apparent magnitude ( $m$ ) and absolute magnitude ( $M$ ) of a star or galaxy to its distance modulus ( $\mu$ ) which is a simple function of its distance in parsecs ( $r$ ) (see A1Y Stellar Astrophysics course and handout):

$$m - M \equiv \mu = 5 \log r - 5. \quad (1)$$

We can rewrite this equation as

$$r = 10^{0.2(m-M+5)}, \quad (2)$$

so that if we know  $m$  and  $M$  we can calculate the distance,  $r$ . Certain variable stars provide us with an *estimate* of  $M$  (something we cannot measure directly, without travelling to the star!), so these distance indicators are basically **luminosity indicators**.

The two most commonly used variable star distance indicators are

**RR Lyrae stars** These are A and F type giants<sup>a</sup>, which are pulsating. Their mean  $M_V$  is approximately constant, so they are a good example of a **standard candle** which is simply a class of object (star, galaxy etc.) assumed to have a predictable intrinsic luminosity. RR Lyrae stars are often found in globular clusters.

**Cepheid variable stars** These are F and G type supergiants, with  $-6 < M_V < -2$ , which pulsate with a period of  $\sim 1$  day to  $\sim 50$  days. Because of the existence of the **period luminosity relation** (see later) their absolute magnitude can be accurately estimated from their period. There are two types of Cepheids: type I and type II. Type II Cepheids are also known as W Virginis stars and are about 2 magnitudes fainter than a type I Cepheid of the same period.

We can measure distances in the Galaxy from stellar parallax (and geometry) or by measuring the apparent magnitudes of 'standard candles', such as RR Lyraes and Cepheids.

### 1.3 Where are we in the Milky Way?

There were several early attempts to map out the overall structure of the Milky Way. In the early 1900s Kapteyn studied how the number of stars of a given apparent magnitude – which could be

<sup>a</sup>The letters here denote the *spectral type* of a star, which is in turn related to its temperature and age. The sequence of spectral types, which runs O B A F G K M, is discussed in detail in the A1 stellar astrophysics course. For now all we need to remember is that the sequence gets cooler and redder from O to M. O type stars are very hot, blue stars which shine for only a few tens of millions of years; G type stars are much cooler, yellowish stars like the Sun, and shine for about 10 billion years.

related to the actual *space density* of stars – varied with direction. He found that the density seemed to fall off in all directions, suggesting that the Sun was located in the centre of the Galaxy. Kapetyn, however, took no account of **extinction** – the absorption of starlight by interstellar dust grains which makes stars appear dimmer. The apparent density drop-off was not a real effect but was due to extinction. Also, extinction restricted Kapetyn’s survey volume to a very small part of the disc – too small to reveal its true shape. Shapley (1917) analysed the distribution of galactic **globular clusters** (GCs), measuring their distances using RR Lyraes. He found that GCs appeared to be roughly spherically distributed, centred on a point about 10 kpc from the Sun, which he argued was the centre of the galaxy. Oort and Linblad, in the 1920s, studied the *motions* of stars in the solar neighbourhood – revealing that the Sun was in a circular orbit around a point approximately coincident with the centre of Shapley’s GC distribution.

The concept of the Galaxy (and other galaxies) is relatively recent – only since the 1920s.

## 1.4 Rotation of the Galaxy

Oort and Linblad showed the the Galaxy does not rotate as a rigid body, but *differentially* – i.e., the angular speed of stars around the galactic centre depends on their distances from it. The inner part of the disc rotates like a rigid body: the speeds of stars are proportional to their distances from the galactic centre. The outer part of the disc is known as the **Keplerian** part, since the orbits approximately obey Kepler’s laws. The transition from rigid-body to Keplerian motion occurs at a distance just inside the Sun’s distance from the galactic centre.

A **rotation curve** is a plot of rotation speed as a function of distance from the centre of the galactic disc. Fig. 1 shows a schematic rotation curve for the Galaxy.

The total mass of the Galaxy *interior* to the Sun’s distance from the galactic centre can be estimated using **Kepler’s third law**<sup>b</sup>,

$$GM_{\text{Gal}}P^2 = 4\pi^2 a^3, \quad (3)$$

where  $M_{\text{Gal}}$  is the mass of the Galaxy interior to  $a$  (the Sun’s distance from the centre), and  $P$  is the Sun’s orbital period – about  $2.5 \times 10^8$  yr. The mass outside the distance of the Sun has no effect on the Sun’s orbit. **Taking  $a \simeq 8.5$  kpc we get  $M_{\text{Gal}} \simeq 10^{11} M_{\odot}$ .**

The stars in our Galaxy orbit around the galactic centre, either with an approximately constant *period* (inner stars) or with an approximately constant *speed* (outer stars).

<sup>b</sup>This is only an approximation because the mass interior to the Sun’s distance includes the central bulge and the disc. Since the distribution of the latter isn’t spherically symmetric about the galactic centre this complicates things somewhat compared with the simple prediction of Kepler III, but we ignore this complication here.

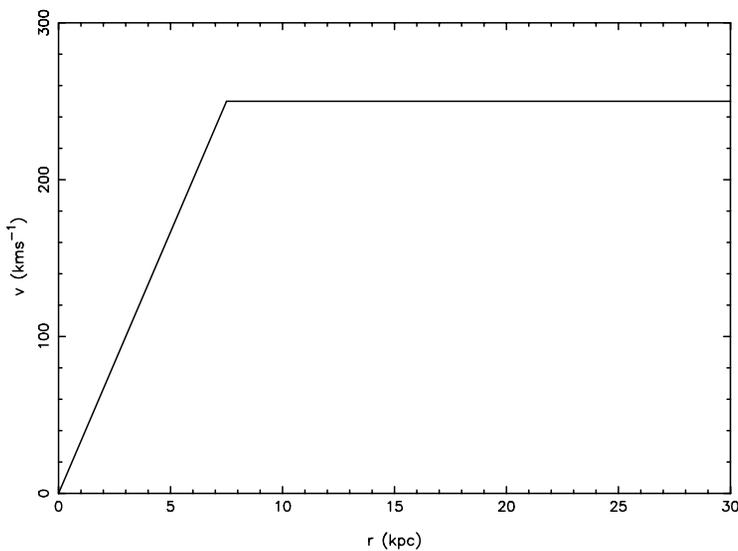


Figure 1: A schematic galaxy rotation curve.

## 1.5 The spiral structure of the Galaxy

The stars in the disc of the Milky Way Galaxy are not uniformly distributed but are found to lie along spiral arms, which wind tightly around the galactic bulge. The spiral arms are populated by large number of young O and B type stars and lanes of dust and molecular clouds – i.e., they are the sites of recent and ongoing **star formation**. The spiral structure of the galaxy can be mapped by measuring the emission of neutral hydrogen (‘HI’) throughout the disc. This radiation is emitted at a radio wavelength of  $\lambda \simeq 21$  cm and is largely unaffected by extinction. Since the interstellar medium comprises primarily hydrogen, and is concentrated along the spiral arms, measurement of HI emission traces the spiral structure very well.

Because of the relative motion of the galactic hydrogen with respect to the Sun, the HI emission will be **Doppler shifted** to wavelengths other than 21 cm; the amount of the shift tells us how fast the HI clouds are moving. One can measure the amount of HI emission as a function of wavelength in different directions on the sky, and interpret the ‘spread’ in wavelengths around 21 cm in terms of the spatial distribution and differential rotation of the spiral arms along each line-of-sight. 21 cm maps of the Milky Way show the spiral structure to be somewhat fragmented and disjointed (see Fig. 2).

Observations of 21 cm radio emission from hydrogen atoms in the interstellar medium help us to map the shape of the Galaxy.

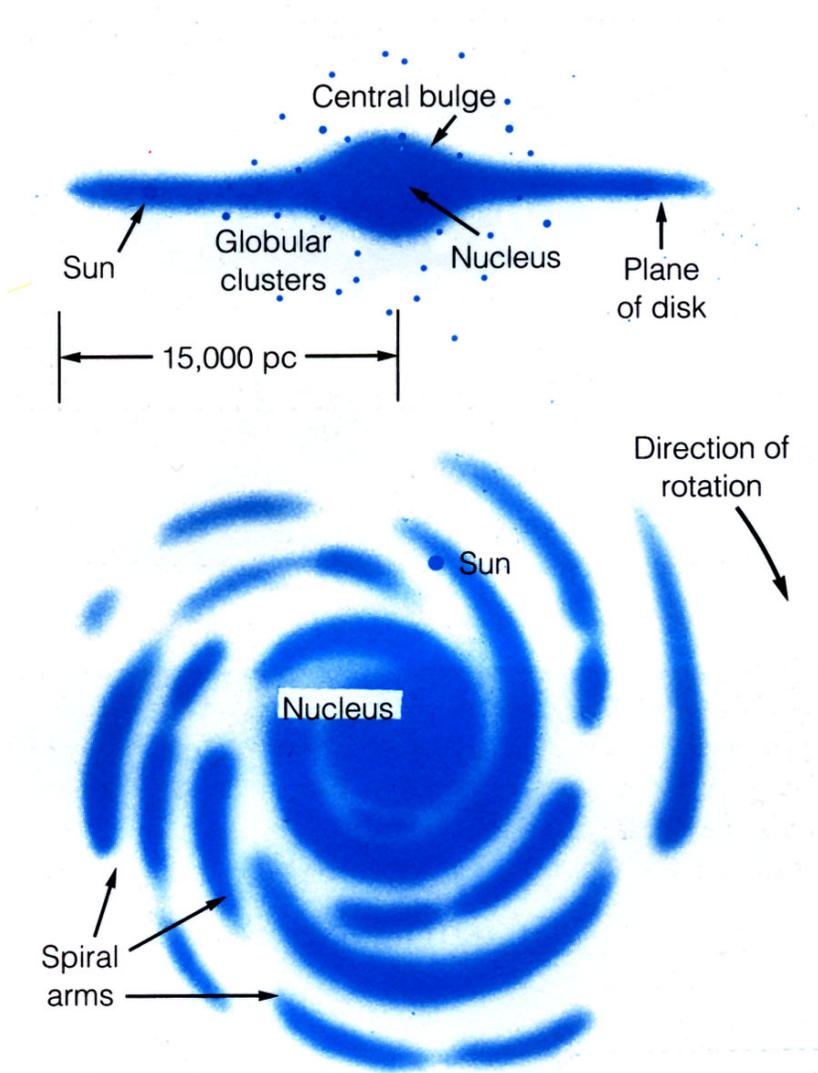


Figure 2: The structure of the Milky Way Galaxy.

## 1.6 The galactic halo

By plotting the rotation curve (from radio observations) out to several tens of kiloparsecs (kpc), astronomers have deduced that the galactic disc appears to be embedded in a roughly spherical **halo** of **dark matter**. The evidence for this halo comes from the fact that the rotation curve does not fall off as rapidly as one would expect if only the luminous stars in the disc were contributing to the gravity of the Milky Way (see Fig. 3.)

If we assume the mass in the Galaxy to be spherically symmetric rather than in a plane, and equate the gravitational attraction it exerts to the centripetal force necessary for circular motion around it, we get

$$\frac{mv^2}{r} = \frac{GM(r)m}{r^2}, \quad (4)$$

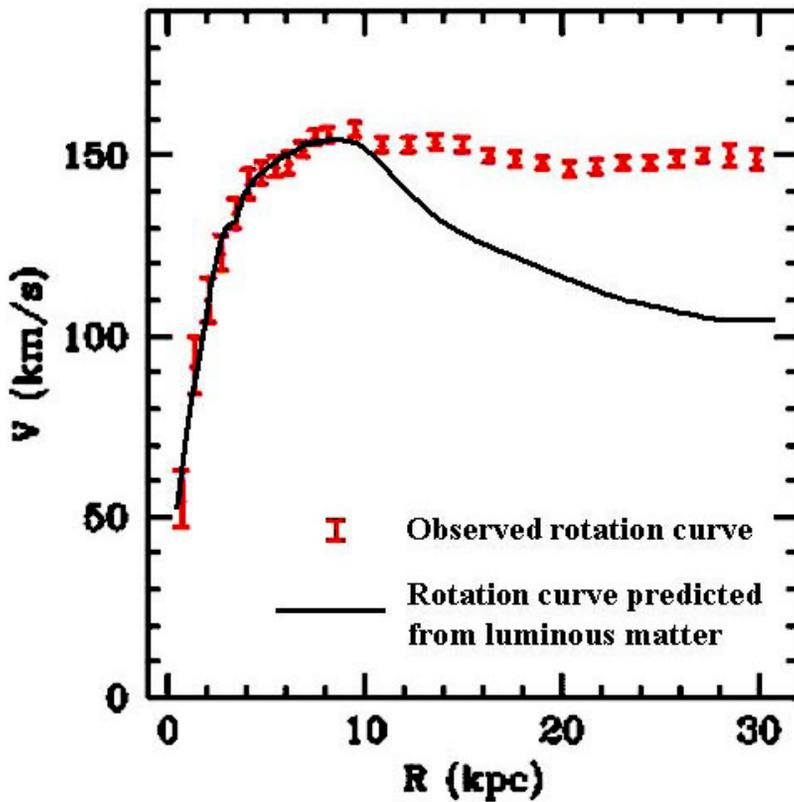


Figure 3: Typical rotation curve for a spiral galaxy like the Milky Way.

where  $m$  is the mass of an orbiting star,  $r$  is its distance from the galactic centre and  $M(r)$  is the mass of the Galaxy within  $r$ . For stars outside the bulk of the Galaxy we have

$$v^2 = \frac{GM_{\text{Gal}}}{r}, \quad (5)$$

$$\text{so that } v(r) \propto r^{-\frac{1}{2}}. \quad (6)$$

We would therefore expect their speeds to fall off inversely with the square root of the distance. The shape of the galactic disc complicates the form of the rotation curve, but even taking this into account, the speeds are still predicted to fall off at large distances. In fact, the observed rotation curve is almost *flat* to distances well beyond the extent of the luminous disc. The presence of a nearly spherical **dark matter halo** extending well beyond the disc would explain this observation. (We return to the question of dark matter, and its implications for cosmology, in more detail later in the course).

The orbits of the outer stars are not consistent with the gravitational attraction of just the visible Galaxy. They imply that the Galaxy contains a lot of dark (i.e., not visible) matter in a spherical halo.

## 1.7 Formation and maintenance of the spiral arms

A **density wave** theory for the formation and maintenance of the spiral arms was proposed in 1960. This theory supposes the existence of a spiral-shaped wave pattern of high and low density regions, centred on the galactic bulge. These density waves cause gas and dust to ‘pile up’ in regions of higher density, causing stars in turn to pile up and become more concentrated in the spiral arms. The density wave theory *predicts* that the inside edge of the spiral arms are the most active star-forming regions.

In the absence of the density wave the spiral structure of a disc galaxy would be much more chaotic and disordered. (In fact, given the fragmented appearance of the spiral arms in Fig. 2, it is thought that the role of the density waves in forming and maintaining the spiral arms was less pronounced for the Milky Way than for other spirals).

The spirals in some galaxies trace out regions of low and high density. The spiral arms are ‘bottlenecks’ of stellar congestion, which individual stars move into and through.

## 1.8 The nature of the “nebulae”

The Messier catalogue contains many galaxies. Previously these were thought to be **nebulae** – i.e., gas clouds *within* the Milky Way. Examples include the Andromeda Spiral, M31. In 1924 Edwin Hubble measured the distance to M31 using Cepheid variables. He found that M31 was much too distant to be inside the Milky Way and, by deducing its intrinsic diameter from its apparent angular diameter, he found that M31 was in fact comparable in size to the Milky Way. Hubble then embarked on a systematic survey and classification of nearby galaxies. He identified three main types of normal galaxies: **spirals**, **ellipticals** and **irregulars**.

---

Elliptical	E0-E7	spheroidal; the number is defined as $n = 10(1 - b/a)$
Dwarf elliptical	dE	spheroidal; very low mass, luminosity
Lenticular	S0	disc-like; no spiral structure
Spiral	Sa-Sc	disc-like; spiral arms
Barred Spiral	SBa-SBc	disc-like; elongated, bar-like nucleus
Irregular I	Ir I	disc-like; spiral structure, but poorly organised
Irregular II	Ir II	‘misfits’

---

An Sa galaxy has a large central bulge and small, tightly wound spiral arms. An Sc galaxy has a small central bulge and wide, open spiral arms.

## 1.9 Properties of normal galaxies

**Spiral galaxies** have diameters in the range 10 to 100 kpc. The mass of the disc is  $10^{11} - 10^{12} M_{\odot}$ . The spiral arms contain OB stars, dust and molecular clouds. The disc rotates around the centre of the galaxy.

e.g., M31, M51, M100

**Elliptical galaxies** have diameters in the range 1 to 100 kpc, and masses in the range  $10^7 - 10^{13} M_{\odot}$ . They are spheroidal in shape, with a smooth brightness profile. They have little interstellar gas. There is a large population of dwarf ellipticals.

e.g., M32, M87

**Irregulars** are irregular in shape – possibly due to recent collisions or mergers with other galaxies.

e.g., the Large Magellanic Cloud.

Ellipticals are old systems: since they have little interstellar gas and dust, they have very little current star formation. In spirals, on the other hand, star formation is still going on – particularly in the spiral arms. (This is why the spiral arms contain O and B stars).

The visible **mass-to-light ratio** is higher for ellipticals than spirals. This is consistent with there being very little current star formation in ellipticals: they contain a smaller proportion of young, massive stars than do spirals.

Stars are still being formed in spiral galaxies, making them relatively rather luminous.

## 1.10 The Hubble tuning fork diagram

Hubble's classification is often represented via a **tuning fork diagram** (see Fig. 4). For many years the prevailing belief was that ellipticals evolve into spirals, from left to right in the tuning fork (although it should be pointed out that Hubble did not argue for the tuning fork diagram as an evolutionary sequence). More recently, the discovery that spirals contain many newly-formed stars led to speculation that, conversely, spirals evolve into ellipticals. However, spirals contain many old stars too, and it is now generally accepted that the spirals and ellipticals which we observe today evolved separately, as part of a much more complex overall pattern of galaxy formation.

The Hubble tuning fork diagram is not an evolutionary sequence.

## 1.11 Active galaxies

Galaxies whose luminosity is greater than that solely due to the stars which they contain are known as **active galaxies**. Their cores are known as **active galactic nuclei** (AGN). All types of active galaxies are observed predominantly at high “redshifts” (see later) indicating they are very distant. Light travels at a finite speed,

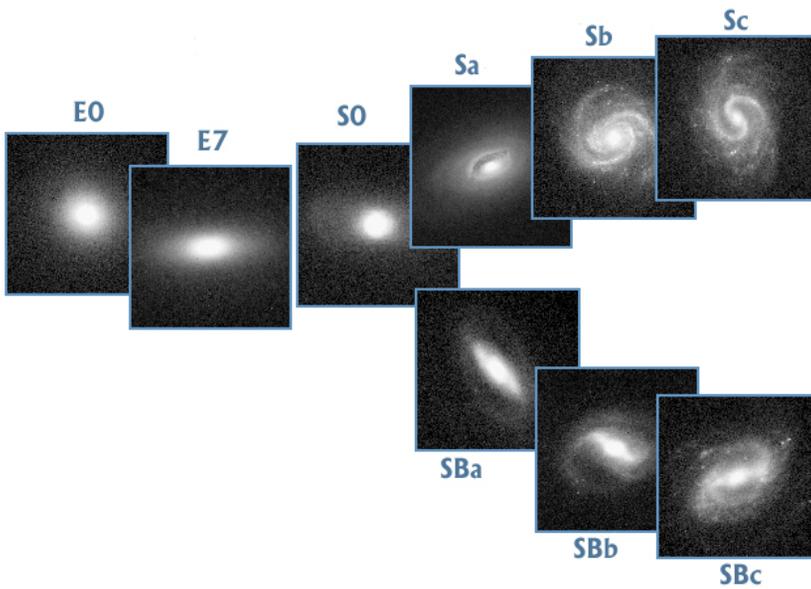


Figure 4: Hubble's tuning fork galaxy classification.

so we therefore see them as they were in the distant past. This indicates that they represent a phase in the early history of galaxy formation, now over. We will consider three types of active galaxy: radio galaxies, Seyfert galaxies and quasars.

### 1.11.1 Properties of radio galaxies

- Galaxy type = **elliptical or giant elliptical (cD) galaxy**
- Ratio of radio to optical luminosity in the range 0.1 – 10.
- Radio source shape is double-lobed (e.g., Cygnus A) or compact central source, often with a *jet* (e.g., M87).
- Compact sources often vary on timescales of days, implying that the size of the emitting region is no more than a few light-days across.
- The radio source spectrum is usually **synchrotron** radiation, indicating the presence of a strong energy source and intense magnetic field capable of accelerating particles (e.g., electrons) close to the velocity of light. The radio lobes are created as jets material, travelling close to the speed of light, drive through the tenuous intergalactic medium.

### 1.11.2 Properties of Seyfert galaxies

- Seyfert galaxies are spiral galaxies with unusually luminous, blue nuclei.
- About 10% of Seyfert galaxies show strong radio emission; some also have jets.
- Optical spectra show strong **emission lines**, formed in a highly ionised gas. Both narrow and broad emission lines are observed (broadening is assumed due to Doppler motions of the gas); broad lines are thought to be formed close to the core of the nucleus, in gas moving at several thousand  $\text{km s}^{-1}$ .
- Short exposure images of Seyfert galaxies reveal only the nucleus – galaxies appear star-like.
- Few nearby spirals show the features of Seyfert galaxies, suggesting that they are an evolutionary phase in the early life of a galaxy.

### 1.11.3 Properties of quasars

- Quasars were first discovered in 1960. They were identified as star-like in appearance but with radio emission and optical spectra which matched no known stars (hence the name, short for “quasi-stellar object”). Their spectra contained strong emission lines, eventually identified as Balmer lines from atomic hydrogen, but redshifted to much longer wavelengths than in the laboratory.
- The conventional interpretation of this redshift is that it is due to the **Hubble expansion** (see later), implying that quasars are at very large distances, and are observed to have very large recession velocities.
- Quasar spectra contain **highly ionised** emission lines of H, He, and often C, N, O, indicating a very intense, hot radiation field. Lines are often very broad, indicating very rapid motions of the hot gas in the emitting region. Many spectra also show weak **absorption lines**.
- Quasar optical luminosities are up to **10 to 100** times that of a normal galaxy.
- Many quasars vary in luminosity over timescales of days or weeks, indicating a very compact emitting region of only 10 to 100 AU in diameter.

- About 10% of quasars are strong radio sources – emission due to synchrotron radiation. Some quasars have observed optical or radio jets.

## 1.12 Quasars: galaxies in infancy

The evidence for quasars being at cosmological distances now appears conclusive, particularly since the Hubble Space Telescope (HST) has recently observed quasar host galaxies at the same redshift. If quasars are so distant however, how can such a high luminosity be produced in such a small volume? The accepted answer to this puzzle is that a quasar is powered by a supermassive black hole at its core. Infalling material releases large amounts of energy as it is swallowed up by the black hole; there is no other satisfactory model which can provide a sufficiently luminous source of energy.

In the standard model, a quasar is thought to be the core of a very young galaxy. The black hole forms at its centre during the chaotic early collapse of the protogalactic gas cloud. Its mass may be as much as  $10^8 M_{\odot}$ . Infalling matter then forms an **accretion disc** around the black hole. The energy released by this infalling matter produces two accelerated jets of particles moving at relativistic velocities, which stream out from the accretion disc, producing beams of synchrotron radiation (see Fig. 5).

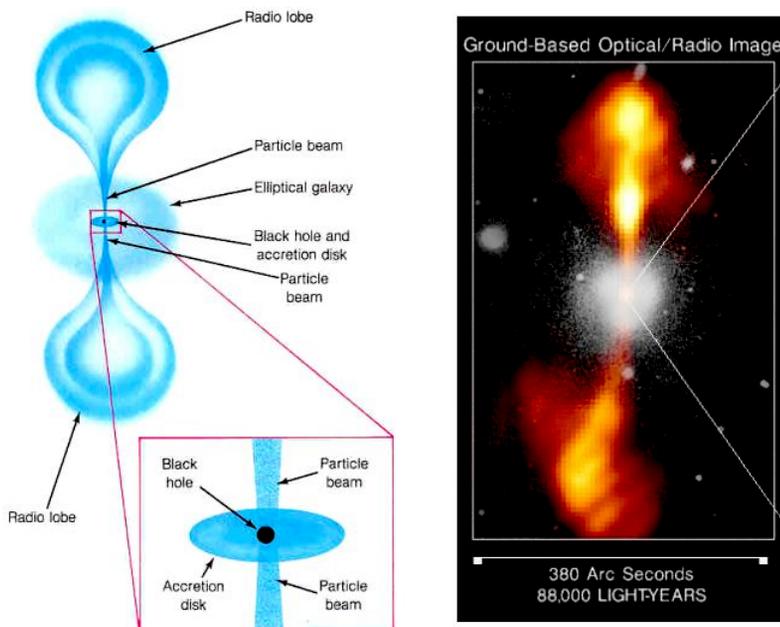


Figure 5: Quasar standard model.

Active galaxies are probably all galaxies in their early stages of formation, seen at different times and different orientations.

Seyfert galaxies and radio galaxies can be described by a similar model to that for quasars, but with somewhat diminished intensity – i.e. these are *less active descendants* of quasars.

### 1.13 Quasar absorption lines

Absorption lines are very common in quasar spectra: usually at a different redshift (almost always lower than that of the quasar). These lines are thought to be due to the absorption of light from the quasar by the extended halos of intervening galaxies. Quasar absorption spectra therefore provide useful information on the environment of newly-formed galaxies.

Some absorption spectra show only H absorption lines indicating that the light from the quasar has passed through intervening clouds which have not yet had time to undergo stellar processing of heavier elements (again, see A1 stellar astrophysics for further discussion). These lines are due to absorption by **pre-galactic clouds**. Quasar absorption spectra can, therefore, constrain the abundance of **primordial elements** (see Sec. 4).

## 2 Section II: large-scale structure of the Universe

### 2.1 Evidence for galaxy clustering: redshift surveys

As we survey the local Universe, we see that the spatial distribution of galaxies is not uniform: galaxies appear to be **clustered**. We see evidence of galaxy clustering in the projected (i.e., 2-dimensional) distribution of galaxies on the sky. Surveys such as the APM Galaxy Survey, showing galaxies visible towards our south galactic pole, reveal a tangled filamentary distribution of galaxies through the local Universe. Of course to get a better idea of the 3-dimensional distribution of these galaxies we also need to know their distances.

In 1936 Edwin Hubble plotted the observed (radial) velocities of nearby galaxies, deduced from the Doppler shift of their spectral lines, against their distances, derived from Cepheid variables within the galaxies. The Doppler shift of a line is just

$$z = \frac{\lambda_{\text{observed}} - \lambda_{\text{emitted}}}{\lambda_{\text{emitted}}}, \quad (7)$$

where  $\lambda$  stands for wavelength. By assuming the observed Doppler shift was related to the velocity of the galaxy by  $v = cz$ , where  $c$  is the speed of light<sup>c</sup>, he found that the galaxies were nearly all moving away from us (that is they were ‘redshifted’) and that their recession velocities,  $v_{\text{rec}}$ , were approximately proportional to their distances,  $d$ , so that

$$v_{\text{rec}} = H_0 d. \quad (8)$$

This is **Hubble’s Law**, and  $H_0$  is known as the **Hubble constant**, usually measured in units<sup>d</sup> of  $\text{km s}^{-1} \text{Mpc}^{-1}$  (we will discuss the cosmological significance of Hubble’s law, and methods of measuring  $H_0$  in Section III). So provided Hubble’s law holds for more distant galaxies, the measured recession velocity of a galaxy gives us an accurate estimate of its distance, assuming we know the value of the Hubble constant.

For example, suppose we observe a galaxy with a recession velocity  $v_{\text{rec}} = 13\,000 \text{ km s}^{-1}$ . If we assume  $H_0 = 71 \text{ km s}^{-1} \text{Mpc}^{-1}$ , then the galaxy is at distance  $d = 13\,000/71 = 183 \text{ Mpc}$ .

You could be worried that because we don’t know  $H_0$  exactly, using recession velocity (or equivalently redshift) as an indicator of a galaxy’s distance is unreliable. In fact we believe we know  $H_0$

Galaxies are clustered in regions of the sky.

Hubble’s law says that distant galaxies recede from us with speeds that increase linearly with distance – the first indications of an expanding Universe.

<sup>c</sup>This is the standard Doppler shift formula for velocities much less than  $c$ . In cosmology, the exact formula depends on the cosmological model used, but all of them approximate to  $v = cz$  at low redshift.

<sup>d</sup>Remember a megaparsec (Mpc) is a million parsecs, which is 3.3 million light years or about  $3 \times 10^{22} \text{ m}$ .

rather well now, but *whatever* its value, Hubble's law tells us that  $v_{\text{rec}}$  at least provides an accurate measure of the *relative distances* of galaxies. For example, if galaxy A is found to have  $v_{\text{rec}} = 12\,000 \text{ km s}^{-1}$  and galaxy B is found to have  $v_{\text{rec}} = 18\,000 \text{ km s}^{-1}$  then, regardless of the value of  $H_0$ , we can say

$$\frac{d_B}{d_A} = \frac{18\,000}{12\,000} = 1.5. \quad (9)$$

We can therefore make accurate maps of the galaxy distribution on large scales using the measured redshift to indicate the relative separation of galaxies. We call such a map a **redshift survey**. Currently about a million galaxy redshifts have been measured, and in the next few years several new redshift surveys will significantly increase this number.

Redshift surveys reveal patterns in the galaxy distribution. In particular we see

- galaxy clusters,
- sheets and filamentary structure,
- voids (i.e., regions which are empty of galaxies).

Fig. 6 shows a famous 'slice' from the Harvard CfA redshift survey and results from the more recent, and deeper, Sloan survey. The largest of these features we can presently see is the Sloan Great Wall. This shows a recession velocity range of about  $30\,000 \text{ km s}^{-1}$  which, from Hubble's law, corresponds to a size of about 430 Mpc (1.4 billion light years). Cosmologists use various different statistical methods to quantify the degree of structure and clustering in redshift surveys; these methods lie beyond the scope of this course (and will be discussed in the honours cosmology course), but when applied to the most recent redshift surveys they show that on scales larger than about  $30\,000 \text{ km s}^{-1}$  the Universe begins to look uniform and **homogeneous**.

## 2.2 Galaxy clustering hierarchy: summary

On small scales galaxies are grouped together in **clusters**. Small clusters may contain about 10 galaxies; the largest rich clusters contain several thousand galaxies. Unlike constellations, galaxy clusters are not simply chance occurrences or 'line of sight' effects: the galaxies in a cluster are believed to have been formed together at the same epoch<sup>e</sup>, and are gravitationally bound together – a property which allows the mass of a galaxy cluster to be

Galaxy redshift surveys reveal the 3-dimensional structure of the Universe on the largest scales.

The clustering of galaxies reveals their gravitational attraction and therefore their masses.

<sup>e</sup>Cosmologists often use the word *epoch* to mean a common moment in time.

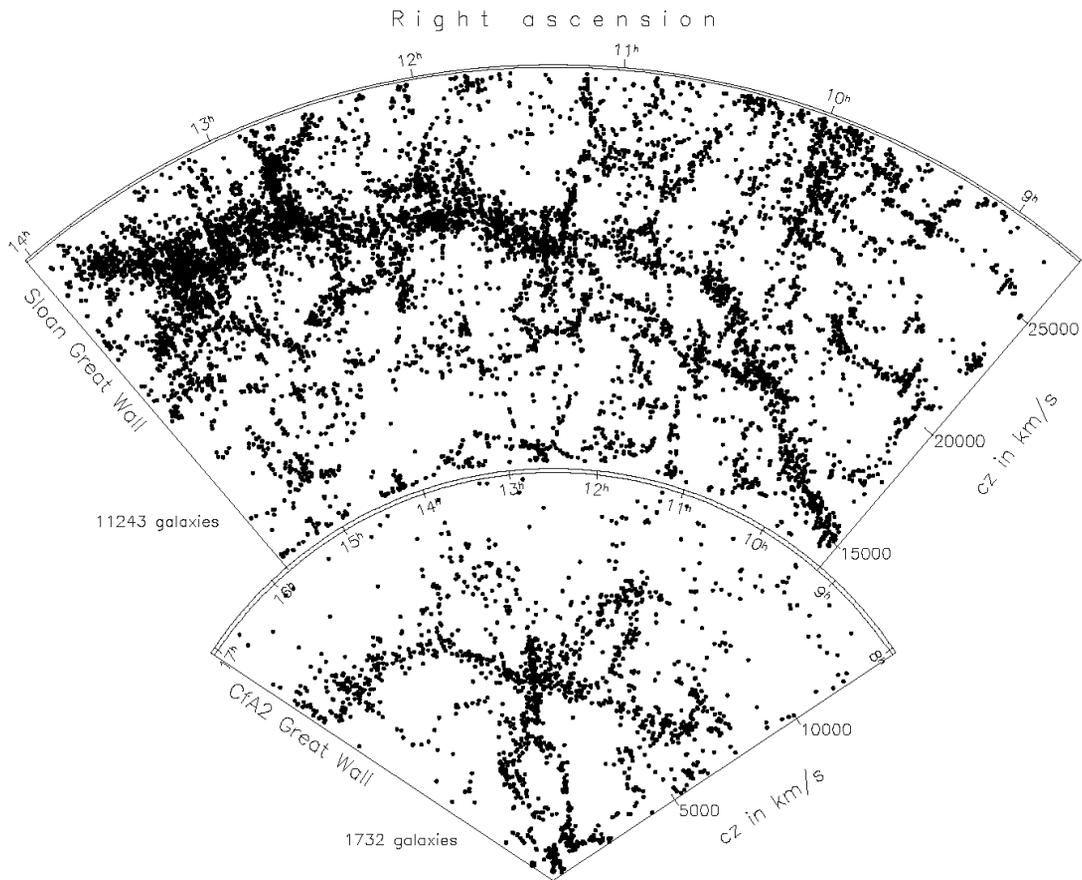


Figure 6: Slices from the CfA and Sloan galaxy redshift surveys. Note the ‘walls’ of galaxies in each, indicating strong clustering (from astro-ph/0310571).

estimated (see Section III). *Within* galaxy clusters, galaxies may have a large **peculiar motion**, or speed, that differs slightly from their recession velocity given by Hubble’s law. This is caused by their gravitational interaction with the other cluster members. The effects of this are most pronounced for galaxies that are reasonably close, and have relatively low recession velocities due to the Hubble flow.

The distribution of galaxy clusters is also non-uniform. Galaxy clusters are themselves clustered, and are organised into larger-scale structures which we refer to as **superclusters**. The filamentary structures observed in redshift surveys (such as CfA, Las Campanas or 2dF) delineate superclusters.

The Milky Way is part of a small cluster of about 30 galaxies which is called the **Local Group**. The Local Group is roughly disc-shaped and about 2 Mpc in diameter. The dominant members of the Local Group are the Milky Way and the Andromeda galaxy, M31, which lies at a distance of about 2.2 million light years, or 0.67 Mpc. These two spirals dominate the dynamics of the

Galaxy clusters themselves cluster to form superclusters.

Local Group and contain the majority of the luminous matter. The remaining galaxies are mainly **dwarf ellipticals** and **irregulars**. The irregular galaxies include the **Large** and **Small Magellanic Clouds**, which are satellite galaxies of the Milky Way at a distance of about 50 kpc.

The nearest rich cluster of galaxies is the **Virgo cluster**, which contains  $\sim 2\,500$  galaxies. Both the mean distance and the mean recession velocity of Virgo cluster galaxies are somewhat uncertain, due to the large peculiar motions of galaxies in the cluster (this has contributed to the uncertainty in the value of  $H_0$  measured using Cepheids). This problem is compounded by the fact that the Virgo cluster is thought to be highly elongated along the line of sight. Recent Hubble Space Telescope data suggests that the core of the Virgo cluster lies at a distance of about 18 Mpc.

The Local Group and Virgo cluster are part of the **Local Supercluster**, which is a planar concentration of many rich galaxy clusters within about  $5\,000\text{ km s}^{-1}$ . The scale of the Local Supercluster is of the order of the largest scales over which structure is observed in galaxy redshift surveys.

The LMC and SMC are the nearest galaxies to the Milky Way.

<i>type</i>	<i>typical scale</i>	<i>typical <math>N_{\text{gal}}</math></i>	<i>examples</i>
galaxy group, small cluster	1 Mpc	10 - 100	Local Group, Fornax cluster
rich cluster	up to 10 Mpc	$\sim 1\,000$	Virgo and Coma clusters
supercluster	50 - 100 Mpc	many thousands	Local Supercluster

### 2.3 Morphological segregation

Statistical analysis of galaxy redshift surveys reveals that elliptical galaxies are preferentially found in the cores of rich clusters, while spirals are generally *not* found there. This *morphological segregation* is thought to be a consequence of the galaxy formation process. It is believed that spirals existed briefly in galaxy clusters shortly after the clusters formed, but their discs could not survive the strong gravitational tidal forces in the cores of clusters. This early population of cluster spirals was, therefore, torn apart and many may have been ‘cannibalised’ by the giant elliptical (cD) galaxies in the centre of the clusters.

Elliptical and spiral galaxies are found in different locations in clusters.

### 2.4 Redshift-independent galaxy distance indicators

As we have seen, for the closest galaxies the Hubble expansion law is *distorted* by peculiar motions, due to the gravitational pull of nearby galaxies, i.e.,

$$v_{\text{obs}} = H_0 d + v_{\text{pec}}. \quad (10)$$

Typical magnitudes for  $v_{\text{pec}}$  are about  $300 \text{ km s}^{-1}$ , although in rich clusters some galaxy peculiar motions may be as much as several thousand  $\text{km s}^{-1}$ . Note that if  $v_{\text{pec}} = 300 \text{ km s}^{-1}$  and  $H_0 = 71 \text{ km s}^{-1} \text{ Mpc}^{-1}$ , then for a galaxy at distance,  $d > 100 \text{ Mpc}$ , then  $v_{\text{pec}}$  is less than 5% of the **cosmic expansion velocity**,  $H_0 d$  (i.e., the recession velocity predicted by Hubble's Law). Therefore:

- We can expect that Hubble's law will hold to within a few percent provided we are *not* considering galaxies in our immediate neighbourhood, where peculiar velocities significantly affect the observed recession velocities.
- If we want to measure the distance to nearby galaxies, we cannot rely on using Hubble's law to do so, i.e., we need to use galaxy distance indicators which are *independent* of redshift.
- Such redshift-independent distance indicators can also be *combined* with measured recession velocities of more distant galaxies (where Hubble's law holds accurately) in order to estimate the value of  $H_0$ .

Peculiar velocities confuse Hubble's law for nearby galaxies.

## 2.5 Standard candle distance indicators

We introduced the idea of a **standard candle** in Section I, and many galaxy distance indicators are based on the standard candle principle. To recap, a standard candle is any class of object, such as a type of galaxy or star, whose luminosity (or equivalently absolute magnitude) can be assumed to be constant.

Suppose we observe a standard candle, of absolute magnitude  $M = M_*$ , to have an apparent magnitude,  $m = m_{\text{obs}}$ . Then (neglecting absorption)

$$m_{\text{obs}} = M_* + 5 \log r + 25, \quad (11)$$

where the distance,  $r$ , of the standard candle is measured in Mpc.<sup>f</sup> Hence an *estimate* of the distance to the object is

$$r_{\text{est}} = 10^{0.2(m_{\text{obs}} - M_* - 25)}. \quad (12)$$

<sup>f</sup>You can prove this if you remember the definition of apparent magnitude and absolute magnitude. Two stars of fluxes  $S_1$  and  $S_2$  at Earth have apparent magnitudes related by  $m_1 - m_2 = -2.5 \log(S_1/S_2)$  by definition. For the same star at two distances, its flux is just proportional to one over the distance squared, so

$$m_1 - m_2 = -2.5 \log(r_2^2/r_1^2) = -5 \log(r_2/r_1).$$

The absolute magnitude of a star is its apparent magnitude at a distance of 10 pc, or  $10^{-5} \text{ Mpc}$ . So working in Mpc we have

$$M_* - m_{\text{obs}} = -5 \log(r/10^{-5}) = -5 \log r - 25.$$

Of course in practice the standard candle assumption is only an approximation (not all our standard candle objects have absolute magnitude exactly equal to  $M_*$ ) but the standard candle is still useful as a distance indicator provided that the *spread* in absolute magnitude from object to object is small.

If the difference between the *assumed* absolute magnitude and the *true* absolute magnitude is  $\Delta M$ , then the *fractional error* in the estimated distance is

$$\epsilon = \frac{r_{\text{est}} - r_{\text{true}}}{r_{\text{true}}} = 10^{0.2\Delta M} - 1. \quad (13)$$

So to be practically useful the standard candle should have  $\Delta M \ll 1$ , giving  $\epsilon \ll 1$ . For example, suppose  $m_{\text{obs}} = 15$  and  $M_*$  is assumed to be  $-20$ , then

$$r_{\text{est}} = 10^{0.2(15+20-25)} \quad (14)$$

$$= 100 \text{ Mpc}. \quad (15)$$

But if  $M_*$  is *actually*  $-21$ , then

$$r_{\text{true}} = 10^{0.2(15+21-25)} \quad (16)$$

$$= 158 \text{ Mpc}, \quad (17)$$

which represents a significant error ( $\epsilon \simeq -0.37$ ).

In summary, a good standard candle distance indicator should

- have a small spread in absolute magnitude,
- be observable to large distances.

Standard candles are best if they truly have a known luminosity.

Some examples of standard candles in common use are

1. Sc spiral galaxies,
2. brightest cluster elliptical (cD) galaxies,
3. type Ia supernovae (see Section III).

## 2.6 Primary and secondary distance indicators

Distance indicators can also be divided into two categories:

**Primary indicators** can be calibrated from theory or from distances measured within our immediate neighbourhood, the Local Group. These include Cepheid variable and RR Lyrae stars (via their respective period-luminosity relations), measurements of annular stellar parallax and main sequence fitting (see stellar astrophysics course)

**Secondary indicators** must be calibrated using a sample of galaxies beyond the Local Group whose distances have been determined by other methods (using primary indicators). These include type Ia supernovae (as standard candles), the Tully-Fisher relation (see later) and a variety of other indicators such as galaxy luminosity class.

Here is a brief list of some good distance indicators. We will look at some of them more carefully in Section III (see also Fig. 7).

<i>method</i>	<i>type</i>	<i>range</i>	<i>calibration</i>
MS-fitting	primary	200 pc	Pleiades
parallax	primary	500 pc	absolute (Hipparcos satellite)
RR Lyraes	primary	1 Mpc (HST)	LMC, Milky Way RR Lyraes
Cepheids	primary	20 Mpc (HST)	LMC, Milky Way Cepheids
GCLF <sup>a</sup>	secondary	50 Mpc	ellipticals in nearby rich clusters
Tully-Fisher <sup>b</sup>	secondary	200 Mpc	spiral galaxies in local supercluster
standard galaxies	secondary	< 1 000 Mpc	galaxies in nearby rich clusters
supernovae	secondary	> 1 000 Mpc	SNIa host galaxies in local supercluster

<sup>a</sup>Globular clusters luminosity functions: GCs have standard candle-like luminosities.

<sup>b</sup>A relationship between galactic rotation speed and luminosity.

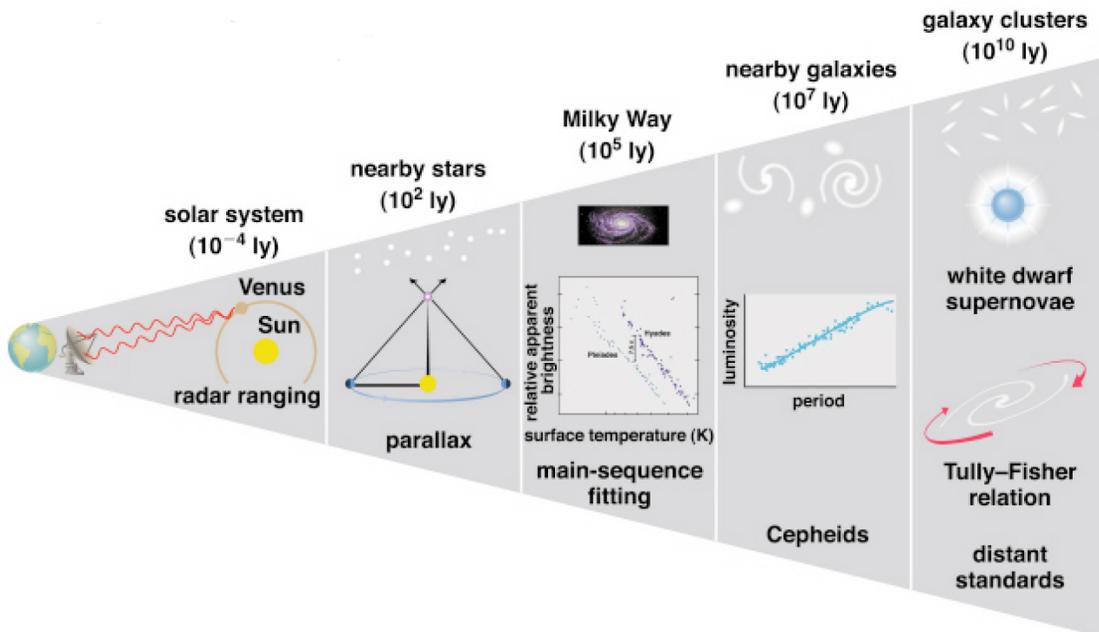


Figure 7: The cosmological distance ladder, showing how different overlapping measuring techniques allow us to measure distances out to Gpc scales (1 Gpc = 1 000 Mpc).

## 2.7 Improving the standard candle assumption

For some distance indicators we do not need to assume that they are standard candles, we can get an excellent indication of their luminosity by using some other directly measurable quantity which is **correlated** with absolute magnitude (which of course we cannot measure directly).

The best-known examples of such an indicator are **Cepheid variable stars**, introduced earlier. For Cepheids, a linear relationship exists between the mean absolute magnitude (averaged over the pulsation cycle of the star) and the logarithm of the pulsation period. We call this the **period-luminosity relation** (See also A1 stellar astrophysics). Therefore, if we can measure the apparent magnitude and the pulsation period of a Cepheid in a distant galaxy, we can use the PL relation to estimate the absolute magnitude of the Cepheid, and hence deduce an estimate of its distance.

The Cepheid PL relations were discovered by Henrietta Leavitt, from a plot of mean apparent magnitude against  $\log(\text{period})$  of Cepheids in the Large Magellanic Cloud. Because these Cepheids were approximately all equidistant from us, her plot translated directly into a linear relationship between absolute magnitude and  $\log(\text{period})$ . Cepheid PL relations exist at all wavebands<sup>g</sup> between U and K. All are well fitted by the linear form

$$\overline{M} = a \log P + b, \quad (18)$$

where  $a$  and  $b$  are constants. Since the relations may be calibrated using Cepheids in the LMC and SMC (see Fig. 8) Cepheids are *primary* distance indicators.

For example, LMC data show that in the V band the PL relation is

$$\overline{M}_V = -2.76 \log P - 1.40.$$

If the HST observes a Cepheid in another galaxy with  $\overline{m}_V = 25$  and  $\log P = 1.5$ , then the prediction for the star's absolute magnitude

<sup>g</sup>The letters U, B, V, R, I, J, H, K denote well-defined wavelength ranges, called the *Johnson wavebands*, used in manufacturing astronomical filters. The sequence follows increasing wavelength: U is in the ultra-violet and K is in the infra-red part of the spectrum. Stars and galaxies will give out different amounts of light at different wavelengths, so that measuring their apparent magnitudes through different Johnson filters will give different results. See A1 stellar astrophysics and observational astrophysics for more details on apparent magnitude filters.

Cepheids are some of the best primary distance indicators – locally calibrated and visible to great distances.

is

$$\begin{aligned}
 \overline{M}_v &= -2.76 \times 1.5 - 1.40 \\
 &= -5.54, \\
 \text{so that } d &= 10^{0.2(\overline{m}_v - \overline{M}_v - 25)} \text{ Mpc} \\
 &= 10^{0.2(5.54)} \\
 &= 12.8 \text{ Mpc}.
 \end{aligned}$$

The Cepheid PL relations are not perfect, but have some statistical scatter (the true absolute magnitude at a given period does not always lie exactly on best-fit straight line). Consequently, Cepheid distance estimates are also not perfect, but are generally subject to a random error of about 10%. A more serious problem with Cepheids is the presence of *systematic* errors due to extinction (the stars can look dimmer than they should, because their light is absorbed slightly by the gas and dust along the line of sight to us). Cepheids are found in spiral arms, where extinction due to dust may be considerable. Ignoring extinction leads to an overestimate of distance. Recent HST and ground-based observations have overcome this problem by using multicolour observations giving B, V, R, I Cepheid PL relations. The amount of extinction varies with wavelength, and so may be estimated and corrected.

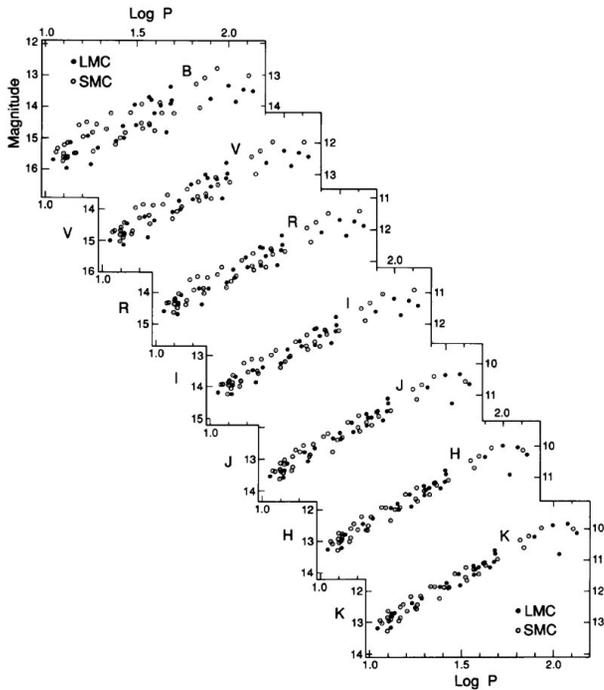


Figure 8: Cepheid period-luminosity relations in the LMC and SMC at different wavebands. Note the small variations in apparent magnitude between the different bands.

## 3 Section III: The expanding Universe

### 3.1 The Hubble expansion

In Section II we introduced **Hubble's law**. It says that, based on observational evidence, the recession velocity of a galaxy is proportional to its distance from us. The standard interpretation of Hubble's law is that the Universe is **expanding**, carrying distant galaxies away from us. A uniform expansion would indeed result in all objects receding from each other at a rate proportional to their separation, like ants on a stretching rubber rope. Therefore, Hubble's constant measures the **expansion rate** of the Universe. Recent determinations of the Hubble constant,  $H_0$ , suggest that it has a value of about  $70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ .

We see the Universe as expanding at a rate determined by Hubble's constant,  $H_0$ .

### 3.2 Measuring $H_0$

In the 1920s and 1930s when Edwin Hubble measured the redshifts and distances to nearby galaxies, he estimated  $H_0$  from the gradient of the **best-fit straight line** drawn through a plot of recession velocity against distance (see Fig. 9). This gives a more reliable estimate of  $H_0$  than simply dividing velocity by distance for a single galaxy. From Hubble's original data he estimated a value of  $H_0 \simeq 500 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . In the light of current measurements this was a gross over-estimate. There were a number of reasons for this:

Hubble's original data were flawed, but his basic conclusions were sound.

- He only measured velocities out to about  $1\,000 \text{ km s}^{-1}$ . Within this distance peculiar velocities are dominant and he was *not* measuring the true cosmological expansion velocity.
- He grossly underestimated the distances to his calibrating galaxies, partly due to using the wrong absolute magnitude for Cepheid variables, making the wrong correction for extinction, and (even worse!) misclassifying as Cepheids objects which were not Cepheids at all.

Over the course of the next 50 years or so, many of these problems were resolved, but by the 1980s there was still much disagreement over the value of  $H_0$ . In particular, one camp argued for a value close to  $50 \text{ km s}^{-1} \text{ Mpc}^{-1}$  and another camp argued for a value close to  $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . It was (and still is, in many contexts) common to write  $H_0 = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}$ , so that uncertainty over the value of  $H_0$  could be recast as uncertainty over the value of the dimensionless number  $h$ . Therefore, the disagreement was between those who favoured  $h \simeq 0.5$  and those who favoured  $h \simeq 1.0$ . Much of this disagreement involved disputes over the distance to the Virgo galaxy cluster.

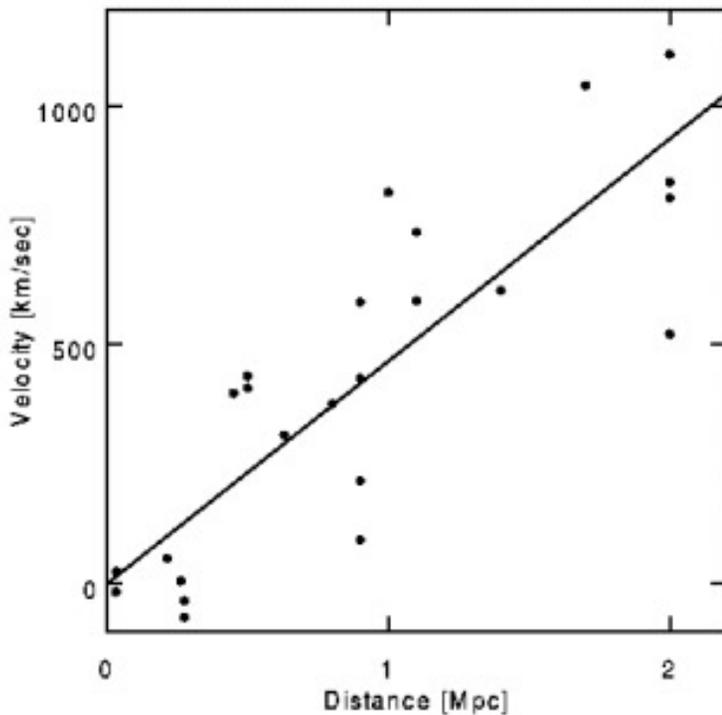


Figure 9: Hubble's original data. (Note that the slope Hubble measured was  $\simeq 500 \text{ km s}^{-1} \text{ Mpc}^{-1}$ ).

### 3.3 The cosmological distance ladder

To determine  $H_0$  we need to combine primary and secondary distance indicators, because

- $H_0$  estimates require **both** accurate distances and recession velocities,
- primary distance indicators only extend out to about 20 Mpc (and before HST they extended only to about 4 Mpc from ground based observations),
- for  $d \simeq 20 \text{ Mpc}$ , the observed radial velocities of galaxies are still seriously affected by peculiar motions, so that  $v_{\text{rec}} \neq H_0 d$ ,
- we require secondary distance indicators to extend from  $\sim 20 \text{ Mpc}$  to  $\geq 100 \text{ Mpc}$ , where Hubble's law holds more accurately, so that  $v_{\text{rec}} = H_0 d$ .

No single distance indicator works well on all scales, and no primary indicators work on the largest scales, so we need overlapping methods.

We call this combination of two or more primary and secondary distance steps the **cosmological distance ladder**.

## 3.4 Examples of secondary distance indicators

### 3.4.1 Type Ia supernovae

A type Ia Supernova (SNIa) is believed to occur when a white dwarf star (see A1 stellar astrophysics) has accreted sufficient matter from a binary companion to push itself over the Chandrasekhar mass limit, causing a **thermonuclear explosion**.

A SNIa brightens by many magnitudes over a few days. At **maximum light**, they are almost as luminous as an entire galaxy, and the supernova then fades over several months. By plotting the SNIa's **light curve** we can determine the apparent magnitude at maximum light. For some time SNIas have been known to be good standard candles, because their **Hubble diagram** is linear, at least out to distances of a few hundred Mpc. The Hubble diagram of a SNIa is a plot of the maximum apparent magnitude,  $m_{\max}$ , versus the log of the recession velocity. Neglecting extinction and peculiar motions, this should be linear for a nearby standard candle because

$$\begin{aligned} m_{\max} &= M_{\max} + 5 \log_{10} d + 25 \\ &= M_{\max} + 5 \log_{10} \left( \frac{v_{\text{rec}}}{H_0} \right) + 25 \\ &= 5 \log_{10} v_{\text{rec}} + M_{\max} - 5 \log_{10} H_0 + 25. \end{aligned} \quad (19)$$

If  $M_{\max}$  is constant then  $M_{\max} - 5 \log_{10} H_0 + 25$  is constant, so that the SNIa will lie along a straight line in the Hubble diagram. By measuring  $M_{\max}$  independently (e.g., by determining Cepheid distances to some SNIa host galaxies) we can go on to estimate  $H_0$  from more distant SNIas.

### 3.4.2 The Tully-Fisher relation

The Tully-Fisher relation is a linear relationship between the absolute magnitude and the log of rotation velocity of spiral galaxies. The rotation velocity is usually the velocity in the flat part of the rotation curve (recall Fig. 1), often measured from the width of the HI 21 cm line.

The Tully-Fisher relation is a secondary distance indicator, because it requires to be calibrated using a set of nearby galaxies, usually in clusters, whose distance (and therefore absolute magnitude) has been determined using primary distance indicators. Unfortunately, there are not enough suitable spiral galaxies in the Local Group to calibrate Tully-Fisher using *only* Local Group galaxies.

An equivalent relation for elliptical galaxies exists between the intrinsic diameter of the galaxy and the range in velocity of its central

Because they all have about the same mass, type 1 supernovae have very similar peak luminosities.

We use the maximum apparent magnitude of the supernova to measure its distance.

Both the Tully-Fisher relation and the  $D_n$ - $\sigma$  relation use the velocities of stars in a galaxy to infer its absolute magnitude.

stars. Note that elliptical galaxies do not undergo large-scale rotation. Instead, their stars have random motions, and the larger the galaxy the larger the random motions. We call this relation the  $D_n$ - $\sigma$  relation, and it is a special case of the so-called *Fundamental Plane* relation for ellipticals. Calibrating the  $D_n$ - $\sigma$  relation (and several other secondary indicators which use elliptical galaxies) is problematic because there are no suitable large elliptical galaxies within the Local Group. We need to extend our distance ladder at least to the Virgo Cluster, the core of which contains many suitable ellipticals, to calibrate them.

### 3.5 The distance ladder after HST

Before the launch of the Hubble Space Telescope, Cepheids could be observed only within the Local Group. This was still inadequate to calibrate all secondary distance indicators, and therefore make the jump to distances where Hubble's law is valid. Reasons for this include:

- the lack of elliptical galaxies in the Local Group to calibrate elliptical based methods,
- the lack of Local Group spirals, suitable to calibrate the Tully-Fisher relation,
- the lack of Local Group SNIa hosts, suitable to calibrate the Hubble diagram.

After the launch of HST, Cepheids became directly observable within nearby clusters. This allowed the *direct* calibration of secondary distance indicators, including Tully-Fisher and those involving ellipticals or SNIa host galaxies, and provided a link to more distant clusters, such as the Coma cluster, where Hubble's law could be assumed to hold to within a few percent.

HST Cepheid observations therefore allowed the cosmic distance ladder to  $H_0$  to be cut to just two steps. This has greatly improved the accuracy of  $H_0$  estimates, has resolved the disputes over the distance to the Virgo cluster and has largely settled the question whether  $H_0$  lies close to 50 or 100. The answer is that *neither* value is correct! A value between 60 and 80  $\text{km s}^{-1} \text{Mpc}^{-1}$  is now almost universally accepted and new microwave background evidence (see later) has led many cosmologists to favour a value of about 70  $\text{km s}^{-1} \text{Mpc}^{-1}$ .

A sketch illustrating the cosmological distance ladder, and in particular showing how HST Cepheids have linked the Magellanic Clouds directly to SNIa host galaxies and the Virgo and Leo clusters, is shown in Fig. 10. The properties of different primary and secondary distance indicators were summarised in Section II.

The HST allowed us to miss out the 'Local Group' rung on the distance ladder, and so get to  $H_0$  is only two steps.



of looking at it is that if each star has some finite volume of space allocated to it, and all of space is allocated in this way, then the volumes will fill up with light over time.

Olbers's Paradox is easily resolved. We now know that

1. stars have finite lifetimes, and can't fill their portion of space with light forever,
2. the speed of light is finite, so only stars within a finite distance can be observed, i.e., only those born long enough ago to allow time for their light to reach us,
3. above all, the Universe almost certainly has a *finite age*.

Olbers's Paradox tells us that the Universe is *not* infinite, eternal and unchanging.

Olbers's Paradox is still of interest today because it reminds us that even apparently simple questions can lead to profound insights. Note also that only point (3) above uses any property of the Universe as a whole (any *cosmological* property). Therefore, just using what we know about the properties of stars and the speed of light, we can say something about the nature of the Universe itself (it can't be infinite and eternal without introducing some way to replenish the matter turned into energy by stars). Point (3) is at the heart of modern Big-Bang models.

### 3.6.2 The Cosmological Principle

The standard model for the origin and evolution of the Universe is called the **Hot Big Bang model**. This says that the Universe began sometime between 10 and 20 billion years ago and has been **expanding** ever since.<sup>1</sup> Although the Universe is evolving over time, it is assumed in the Big Bang model that at any time  $t$  the Universe is **homogeneous** and **isotropic**:

1 billion years in a 'gigayear' (Gyr).

Universe homogeneous = Universe looks the same no matter where you are in it,

Universe isotropic = Universe looks the same no matter what direction you look in.

We call these two assumptions the **Cosmological Principle**. Clearly the Cosmological Principle is not valid on small scales, since we have seen that galaxies are clustered, but it is assumed to hold on sufficiently large scales, that is on scales *larger* than  $\sim 10\,000\text{ km s}^{-1}$ , the characteristic size of the largest observed structure in galaxy redshift surveys. (We will see in Section IV that

The Cosmological Principle states that there are no 'special places' in the Universe, such as a centre or an edge.

<sup>1</sup>Expanding into what? Remember that the word *expansion* here means an increasing separation between galaxies over time. That's true for all galaxies, no matter where they are. This isn't an idea that needs some empty space to expand into, as there is no space that does not contain galaxies. Rather it demands extra space be created *between* galaxies.

good evidence for the validity of the Cosmological Principle also comes from the smoothness of the cosmic microwave background radiation).

### 3.6.3 The expansion of the Universe

We can think of galaxies and clusters that we observe in the Universe as embedded within, and expanding with, the **underlying structure of the Universe**. Fig. 11 shows how this underlying structure expands over time. Note that the sizes of galaxies themselves don't change, only the distances between them. This underlying structure is usually assumed to satisfy the Cosmological Principle on *all* scales, so galaxies can be thought of as local disturbances in an otherwise perfectly homogeneous and isotropic Universe. The evolution of the Universe can then be described by

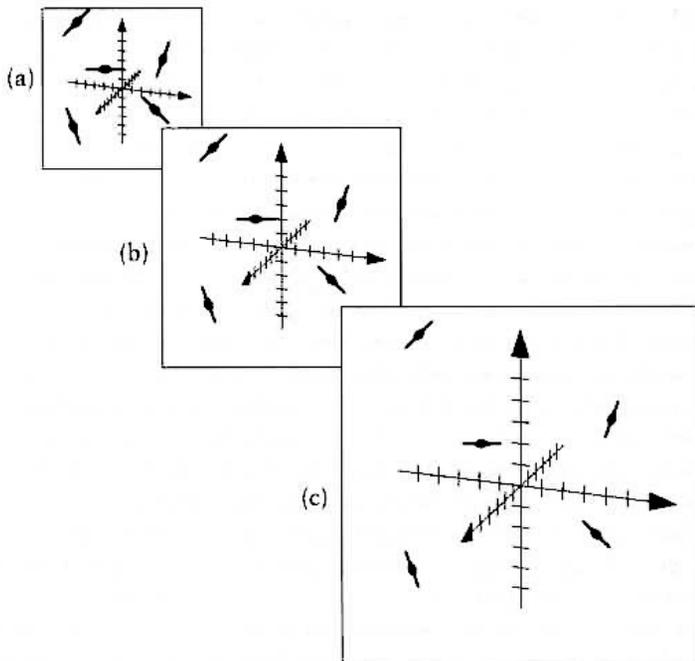


Figure 11: Cartoon picture of universal expansion. Note that both the axes and tick-marks grow, but the galaxies do not.

the size of a dimensionless number which we call the **cosmic scale factor**, and is usually written  $R(t)$ . The scale factor measures the characteristic size of the Universe at time  $t$ . More specifically, it allows one to determine by how far galaxies (embedded in the homogeneous and isotropic underlying structure) have been carried apart by the expansion of the underlying space.

We express this idea mathematically by introducing the **proper distance** between two galaxies at time  $t$ , which is their actual

Proper distance is what metre rulers laid end-to-end would measure. Co-moving separation is measured with respect to an expanding coordinate system.

separation (measured in Mpc perhaps) and their **co-moving separation**, which is their separation expressed in terms of a coordinate system which expands along with the background space. Their co-moving separation is not changed by the expansion of the Universe. The proper distance,  $r(t)$ , between two galaxies with co-moving separation  $s$  is

$$r(t) = R(t) \times s. \quad (20)$$

To repeat, the co-moving separation of galaxies is constant in time. As a useful example of co-moving coordinates think of latitude and longitude on the surface of a spherical balloon. The latitude and longitude of a point on the surface does not change as the balloon is inflated. The proper distance between galaxies, on the other hand, continually changes as the scale factor,  $R(t)$ , changes.

### 3.6.4 Cosmological redshift

We denote the present day value of the scale factor by  $R_0$ , and express other values of  $R$  in units of  $R_0$ . We can give **another interpretation of the redshift** of light from a distant object in terms of the amount by which the Universe has expanded since the light from the object was emitted. The wavelength of light emitted by a distant object will be ‘stretched’ by the expansion of the Universe. If light from a distant object was emitted at time  $t$ , when the scale factor was  $R(t)$ , and is observed at time  $t_0$  (the present day), when the scale factor is  $R_0$ , then

$$\frac{\lambda_{\text{obs}}}{\lambda_{\text{emit}}} = \frac{R_0}{R(t)}, \quad (21)$$

$$\text{i.e., } 1 + z = \frac{R_0}{R(t)} \quad (22)$$

where  $z$  is the apparent redshift defined in Eq. (7). Although the cosmological redshift takes the same mathematical form as the familiar Doppler formula, strictly it is *not* the same effect. **Cosmological redshifts are not due to the *motions* of distant objects but are the result of the stretching of the wavelength of their light as it propagates through expanding space.** We still use much of the vocabulary of motion though, so we talk about galaxies having recession velocities, but the cause of that recession is not the motions of the galaxies within a static fabric of space, but rather the stretching of the fabric itself (in which they are embedded).

Cosmological redshifts are caused by the the expansion of space.

### 3.6.5 Hubble's law and the Cosmological Principle

Consider a galaxy at proper distance  $r$  from us. Its **proper velocity**,  $v$ , is the rate of change of its proper distance, i.e.,

$$v = \frac{dr}{dt} = \frac{d}{dt}(Rs) = \dot{R} \cdot s = \frac{\dot{R}}{R} \times (Rs) = \frac{\dot{R}}{R}r. \quad (23)$$

This is just Hubble's law. In a homogeneously and isotropically expanding universe (that is, one obeying the Cosmological Principle), an observer in *any* galaxy would observe neighbouring galaxies to obey Hubble's law and have proper velocities proportional to their proper distances.

Hubble's constant therefore measures the rate of change of the scale factor,  $R(t)$ . We see that Hubble's constant is not in fact a constant in time, but is a constant in space at any given time, (since  $R(t)$  is independent of position). So in fact

$$H(t) = \frac{\dot{R}}{R}, \quad (24)$$

and the present day value of the Hubble constant is

$$H_0 = \frac{\dot{R}_0}{R_0}. \quad (25)$$

### 3.6.6 When was the Big Bang?

We define the Big Bang by the condition that  $R(t) \rightarrow 0$  at time  $t = 0$ , so that it is the time in the past when the proper distance between galaxies tended to zero. (Clearly this is a simplistic treatment since, as we will see in Section IV, in the very early Universe there were no galaxies!)

We can estimate the time elapsed since the Big Bang by the following simple argument. If we assume a constant expansion rate, so that  $H(t) = H_0$  for all  $t$ , then

$$v = H_0 r = \frac{\text{distance}}{\text{time}} = \frac{r}{t}. \quad (26)$$

If we denote the age of the Universe by  $\tau$ , then it follows from Equation (26) that

$$\tau = H_0^{-1}, \quad (27)$$

or, expressing  $H_0$  in  $\text{km s}^{-1} \text{Mpc}^{-1}$  and  $\tau$  in years,

$$\tau = \frac{978}{H_0} \times 10^9 \text{ yr}. \quad (28)$$

We call  $\tau$  the **Hubble time**, and it sets a timescale for the expansion

$H_0$  will change over time if the rate of expansion of the Universe is changing.

Hubble's constant has dimensions of  $\text{time}^{-1}$ , and the reciprocal of Hubble's constant is called the Hubble time – approximately the age of the Universe.

of the Universe. This simple treatment ignores the effect of gravity however, which will slow down the expansion so that  $H(t)$  was larger in the past. Therefore, including the effects of gravity should give an age of the Universe which is smaller than the Hubble time, so that

$$t_0 < \tau. \quad (29)$$

A more precise determination of the age of the Universe requires solving for  $R(t)$  incorporating the effects of gravity (and possibly also the Cosmological Constant – see later). A rigorous treatment of this requires Einstein’s **General Theory of Relativity** (GR), and lies well beyond the scope of this course, but we can semi-derive an equation for the evolution of  $R(t)$  using only **Newtonian concepts**. We call this Friedmann’s Equation.

### 3.7 Friedmann’s Equation: a simple derivation

Consider a galaxy of mass  $m$ , a proper distance  $r$  from the centre of a sphere containing many other galaxies. The galaxy is gravitationally attracted by the other galaxies within a sphere and this force is equivalent to that from a point mass at the centre, equal to the mass of the sphere. Let the mass of the sphere be  $M$ , and the (uniform) density of the sphere be  $\rho$ . Then

$$M = \frac{4}{3}\pi r^3 \rho. \quad (30)$$

The **kinetic energy** of the galaxy is given by

$$\text{KE} = \frac{1}{2}m\dot{r}^2 = \frac{1}{2}m\dot{R}^2 s^2 \quad (31)$$

and the **potential energy** is given by

$$\text{PE} = -\frac{GMm}{r} = -\frac{4}{3}\pi R^2 s^2 G\rho m. \quad (32)$$

Since the total energy of the galaxy is constant, we have

$$\text{total energy} = \frac{1}{2}ms^2 \left[ \dot{R}^2 - \frac{8\pi G\rho R^2}{3} \right] = \text{constant}, \quad (33)$$

or equivalently

$$\frac{\dot{R}^2}{R^2} - \frac{8\pi G\rho}{3} = -\frac{k}{R^2}, \quad (34)$$

where  $k$  is a constant. Equation (34) is called **Friedmann’s equation**, and describes how gravity slows the rate of expansion of the Universe. We have derived it for a simple Newtonian universe, consisting of a sphere of matter, but a proper relativistic treatment gives the same result.

### 3.8 The Cosmological Constant: Einstein's greatest blunder?

In fact a rigorous General Relativistic treatment yields a second equation for the evolution of the scale factor,  $R(t)$ , which is

$$\frac{\ddot{R}}{R} = -4\pi G \left( \rho + \frac{3P}{c^2} \right), \quad (35)$$

where  $c$  is the speed of light and  $P$  is the mean pressure of the Universe.<sup>j</sup> (Note that you will not be required to know Eq. (35) for the A1Y exam). Einstein realised an important consequence of Eq. (35): for normal matter (for which  $\rho$  and  $P \geq 0$ ) one cannot have a static universe as that would require  $\ddot{R} = 0$ . This was a big problem since, when Einstein was publishing GR, the prevailing belief was that the Universe *was* static.

Einstein fixed this problem by introducing an extra constant,  $\Lambda$  ('lambda'), into Eq. (34) and Eq. (35). This is known as the **Cosmological Constant** and we can think of it as an integration constant in the equations of General Relativity. Eq. (34) and Eq. (35) are the special case where  $\Lambda = 0$ . By choosing the appropriate value for  $\Lambda$ , Einstein could obtain a static solution with  $R(t) = \text{constant}$ . Of course, Hubble's discovery of the expanding Universe did away with the need for a non-zero Cosmological Constant, and Einstein supposedly later described it as his "greatest blunder".

For many decades cosmologists generally assumed that  $\Lambda = 0$ , partly because it simplified the solution of Friedmann's Equation and partly because it avoided the difficult physical problem of explaining what the cosmological constant actually *is*. A positive value of  $\Lambda$  behaves rather like 'anti-gravity': a repulsive force which overcomes the attraction of gravity on very large scales.

Since the late 1990s a mounting body of evidence (e.g. from Type Ia supernovae, the cosmic microwave background radiation and the pattern of galaxy clustering in the Universe – see later) suggests that we do indeed live in a Universe with  $\Lambda > 0$ . We discuss some implications of this startling result later in this section. It remains for physicists and cosmologists to explain more fully what  $\Lambda$  is. The most popular idea is that its origin lies in the so-called 'zero point energy' of the vacuum of empty space. This idea is now spawning even more exotic  $\Lambda$  theories such as 'dark energy' or 'quintessence'. This is a very exciting new field in cosmology but further discussion of it lies well beyond the scope of this course. Some popular books and weblinks are listed on the website.

We now return to Friedmann's Equation in the form of Eq. (34). To keep things simple we will only consider the case of  $\Lambda = 0$ ,

We need to integrate our equations to determine  $R$ , so we are allowed a constant of integration, called the Cosmological Constant.

Although it was long assumed  $\Lambda = 0$ , there is now observational evidence that it *is* needed to describe our Universe.

<sup>j</sup>In General Relativity the pressure of a gas contributes to its gravity, in addition to the mass of the atoms in the gas.

although we should keep in mind that the current cosmological data suggest that in fact  $\Lambda > 0$ .

### 3.9 The curvature of the Universe

The constant,  $k$ , in Eq. (34) defines the geometry, or **curvature**, of the Universe. We can define  $R$  so that  $k$  has three possible values:

$k = 1$  implies that the Universe is **closed**, with **positive curvature**,

$k = -1$  implies that the Universe is **open**, with **negative curvature**,

$k = 0$  implies that the Universe is **flat**, with **zero curvature**.

We can visualise the curvature of the Universe by analogy with the curvature of a 2-D surface, as illustrated in Fig. 12. Changing the curvature of a surface affects the behaviour of initially parallel lines drawn on it. In the Universe, these ‘lines’ can be thought of as light rays: changing the curvature of the Universe distorts these paths and affects the apparent size and brightness of distant objects (this is how distant Type Ia supernovae can be used to measure the curvature of the Universe).

The curvature of the Universe tells us how images are distorted as the light propagates through space.

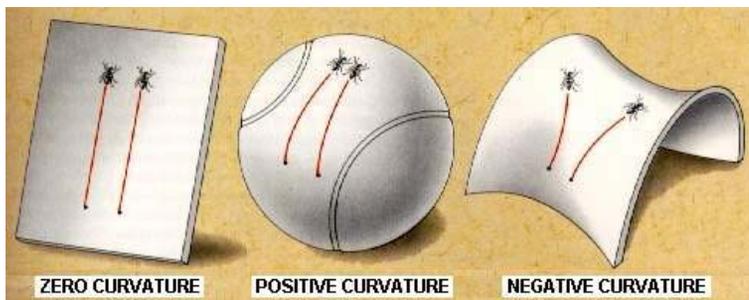


Figure 12: 2-D representation of surfaces of different curvature.

If  $\Lambda = 0$  then  $k$  also determines the long-term behaviour of the scale factor:

$k = 1$  implies  $KE < PE$ ; Universe **bounded**: it expands, then recollapses,

$k = -1$  implies  $KE > PE$ ; Universe **unbounded**: it expands, indefinitely,

$k = 0$  implies  $KE = PE$ ; Universe **just unbounded**: it slows to  $\dot{R} = 0$  as  $R \rightarrow \infty$ .

(The relation between curvature and boundedness is more complicated if  $\Lambda \neq 0$ , but that need not concern us in this course).

### 3.10 Solution of Friedmann's equation for a flat universe with $\Lambda = 0$

The analytic solution of Friedmann's equation is straightforward only for the case of a flat universe ( $k = 0$ ), so that

$$\left(\frac{dR}{dt}\right)^2 = \frac{8\pi G\rho R^2}{3}. \quad (36)$$

If we assume the Universe is **matter dominated**, and mass is conserved, then mass (density  $\times$  volume)  $\propto \rho R^3$  is constant, so

$$\left(\frac{dR}{dt}\right)^2 = \frac{A}{R}, \quad (37)$$

where  $A$  is a constant. It is easy to show that  $R(t) = at^{2/3}$  is a solution to this equation, where  $a$  is another constant. If  $t_0$  is the present age of the Universe, then

$$\frac{R}{R_0} = \left(\frac{t}{t_0}\right)^{2/3}. \quad (38)$$

Equations (22) and (38) give a relation between redshift and time:

$$1 + z = \left(\frac{t}{t_0}\right)^{-2/3}, \quad (39)$$

so that if we observe e.g. a quasar at redshift  $z = 3$ , its light was emitted when the Universe was one eighth of its present age.

For  $R(t) = at^{2/3}$ , differentiating and dividing by  $R$  we obtain

$$\frac{\dot{R}}{R} = \frac{2}{3}t^{-1}, \quad (40)$$

i.e.,

$$t_0 = \frac{2}{3}H_0^{-1}. \quad (41)$$

Comparing with the *Hubble time*, we see that  $t_0 = (2/3)\tau$ . If  $H_0$  is expressed in  $\text{km s}^{-1} \text{Mpc}^{-1}$ , and  $\tau$  in years, the age of the Universe (for  $k = 0$ ) is

$$t_0 = \frac{652}{H_0} \times 10^9 \text{ yr}. \quad (42)$$

Therefore, for  $H_0 = 70 \text{ km s}^{-1} \text{Mpc}^{-1}$ ,  $t_0 \simeq 9 \times 10^9 \text{ yr}$ . This is an uncomfortably low age, compared with the estimated ages of Globular Clusters. By considering the **main sequence turn-off** on the colour-magnitude diagram of globular clusters<sup>k</sup>, astronomers

Stars must be younger than the Universe, and the oldest globular clusters are uncomfortably close to  $2/3\tau$ . However, if the expansion of the Universe is accelerating, it may be somewhat older than  $2/3\tau$ .

<sup>k</sup>See A1 Stellar astrophysics course for details about colour-magnitude diagrams and the Main Sequence.

have recently estimated the age of the oldest globular clusters to be around 11 Gyr. There are many possible sources of uncertainty in these calculations, but most astronomers would accept that the ages of globular clusters cannot be ‘squeezed’ much lower than 10 Gyr, and of course one has to allow a little time between the Big Bang and the formation of the globular cluster. This ‘age paradox’ is now seen as less problematic, since models with a positive  $\Lambda$  now appear to be supported by cosmological data. The effect of  $\Lambda > 0$  is to increase the age of the Universe, for a given value of  $H_0$ .

### 3.11 The critical density of the Universe

As stated in Eq. 36, when  $k = 0$ , Eq. (34) reduces to

$$\left(\frac{\dot{R}}{R}\right)^2 = \frac{8\pi G\rho}{3}, \quad (43)$$

or

$$\rho = \frac{3H^2}{8\pi G}. \quad (44)$$

We call this value of  $\rho$ , the **critical density**, and denote it by  $\rho_{\text{crit}}$ . Its present day value is

$$\rho_{\text{crit}} = \frac{3H_0^2}{8\pi G}. \quad (45)$$

$\rho_{\text{crit}}$  is the density required to just close the Universe. If  $\rho > \rho_{\text{crit}}$  the universe recollapses, but if  $\rho < \rho_{\text{crit}}$ , the Universe expands indefinitely. The present day value of  $\rho_{\text{crit}}$  is equivalent to approximately 5 hydrogen atoms per cubic metre.

Cosmologists often use the dimensionless **density parameter**,  $\Omega(t)$ , where

$$\Omega(t) = \frac{\rho(t)}{\rho_{\text{crit}}(t)}. \quad (46)$$

Therefore

- $\Omega > 1$  implies **Universe closed,**
- $\Omega < 1$  implies **Universe open,**
- $\Omega = 1$  implies **Universe flat.**

Cosmologists denote the present-day value of  $\Omega$  by  $\Omega_0$ . Although  $\Omega$  can change with time, it can be shown that its state of being closed, open or flat cannot change.

Fig. 13 sketches the solution of Friedmann’s equation for  $R(t)$ , for different values of  $\Omega_0$ . We see that  $R(t)$  shows three distinct types of behaviour, depending on whether the geometry of the Universe is open, closed or flat.

$\Omega$  tells us whether there is enough matter in the Universe for gravity to overcome its expansion.

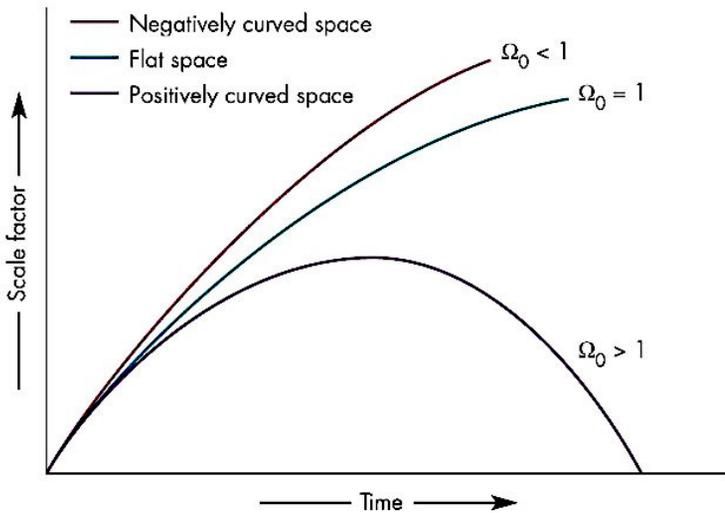


Figure 13: Behaviour of the scale factor in different cosmologies.

### 3.12 Weighing the Universe: methods for measuring $\Omega_0$

So which universe do we live in? To answer that question we need to know  $\Omega_0$ , which depends on the matter density of the Universe,  $\rho$ . We can determine estimates of the matter density in several different ways:

**Visible Stars in the Milky Way** If we assume that all stars in the galaxy are of one solar mass,  $M_\odot$ , then we can estimate the matter density to be

$$\rho = \frac{N_{\text{stars}} \times M_\odot}{\text{volume of Milky Way}}. \quad (47)$$

**Galaxy rotation curves** By measuring the rotation velocity of clouds of neutral hydrogen gas within the disc of spiral galaxies as a function of their radial distance from the centre, we can deduce the amount of mass inside that radius. Recall from Section I that the observed rotation velocities are *greater* than those expected from the gravitational influence of the luminous stars alone, indicating the presence of a **dark matter halo** surrounding the galaxy.

**Galaxy clusters** By assuming that a galaxy cluster is **virialised**, which means that the cluster has ‘settled down’ into a state of equilibrium, there should exist a relation between the mass,  $M$ , and radius,  $r$ , of the cluster and the *mean square peculiar velocity*,  $\langle v^2 \rangle$ , of the cluster galaxies. We call this relation the **virial theorem** and it arises because in a virialised cluster the galaxies have reached a state of balance between their

kinetic energy and potential energy such that  $2\text{KE} + \text{PE} = 0$ . Taking  $\text{KE} = \frac{1}{2}M\langle v^2 \rangle$  and  $\text{PE} = -GM^2/r$ , a **virial mass estimate** for the cluster is then given by

$$M = \frac{\langle v^2 \rangle r}{G}. \quad (48)$$

Note that  $\langle v^2 \rangle$  is the 3-D mean square peculiar velocity, but in practice we measure only the *radial component* of the peculiar velocity (deduced from the galaxy redshift). Assuming a spherical cluster with an **isotropic** velocity distribution

$$\langle v^2 \rangle = 3\langle v_{\text{radial}}^2 \rangle. \quad (49)$$

**Gravitational lensing** The General Theory of Relativity predicts that light will be deflected in a strong gravitational field: we call this phenomenon **gravitational lensing**. We can use lensing to deduce estimates of the matter density in at least two ways:

1. *Weak lensing*: Here light from distant galaxies is distorted by passage through an intervening cluster. The amount of distortion allows the cluster mass density to be estimated.
2. *Microlensing*: Here, light from stars in the LMC and the bulge of the Milky Way is distorted by dark matter crossing our line-of-sight, giving a temporary rise in the brightness of the background stars. The shape of the microlensed star's light curve allows one to place constraints on the mass of the lensing object. Several monitoring programs (such as MACHO, EROS and OGLE) have checked the brightness of millions of LMC and bulge stars every day for a number of years, looking for evidence of microlensing. Hundreds of microlensing events have been found.

**Hubble diagram of standard candles** We have already seen how one may use the Hubble diagram of a standard candle distance indicator to estimate the value of  $H_0$ . For relatively nearby objects the relation between apparent magnitude and log redshift is *linear*. For more distant objects the relation begins to *curve*, and the amount of curvature indicates the curvature of the Universe, which depends of the value of the matter density ( $\Omega$ ) and the cosmological constant ( $\Lambda$ ). The values of these parameters in turn indicate whether the expansion of the Universe is *accelerating* or *decelerating*.

For many years cosmologists have tried to estimate the curvature using the magnitude-redshift relation for quasars or

distant galaxy clusters. However, neither of these objects make good standard candles at very large redshift, however, since **evolutionary effects** become important: at  $z \sim 1$ , we are looking back to sufficiently early times that the luminosity and number density of galaxy clusters and quasars has changed significantly compared with their present day values. It is very difficult to correct for these effects, which one *must* do first before estimating the curvature of the Hubble diagram.

Recently the Hubble diagram of type Ia supernovae, instead of quasars or galaxy clusters, has been used to estimate the matter density and cosmological constant,  $\Lambda$ . The conclusion of these studies is that  $\Lambda > 0$ , showing the expansion of the Universe is **accelerating** (i.e., the Universe is expanding faster now than it was in the past). This also means that the Universe will continue to expand indefinitely and there will be no re-collapse to a ‘Big Crunch’. This startling conclusion is also supported by analysis of galaxy clustering and the cosmic microwave background radiation (see below and Section IV).

**Galaxy redshift and redshift-distance surveys** We saw in Section II that the large-scale distribution of galaxies in the Universe is far from uniform: galaxies are *clustered*, due to the influence of gravity causing structures in the galaxy distribution to grow as the Universe evolves. By studying the patterns of these structures in galaxy redshift surveys we can place limits on the value of  $\Omega_0$ : the higher the matter density the stronger the pattern of galaxy clustering. If we also have redshift-independent information about the *distance* of the galaxies then we can directly measure their line-of-sight peculiar velocity, which results from the gravitational pull of the surrounding matter distribution. (Note that galaxies will experience the gravitational pull *not* just of the luminous matter, but also of the dark matter around them). By studying the patterns of **galaxy peculiar velocities** we can also estimate  $\Omega_0$ .

We can also place limits on the matter density by two methods which we will mainly discuss in Section IV

1. considering the relative amounts of the lightest elements, which we believe were manufactured during the first few minutes after the Big Bang. We call this process **nucleosynthesis**. Note that nucleosynthesis constraints the density of what are called **baryons** – see below.

2. studying the pattern of temperature variations in the **cosmic microwave background radiation**, the relic radiation from the Big Bang itself. Recent CMBR measurements constrain very precisely *both* the baryonic and non-baryonic matter density (see Section IV).

So in summary:

method	limits
nucleosynthesis	$0.015 \leq \Omega_B h^2 \leq 0.026$
CMBR (baryons)	$\Omega_B h^2 = 0.0224 \pm 0.0009$
CMBR (baryons + non baryons)	$\Omega_0 = 0.27 \pm 0.04$
visible stars	$\Omega_B \simeq 0.002$
galaxy rotation curves	$\Omega_0 \simeq 0.01 \text{ to } 0.02$
galaxy clusters	$\Omega_0 \simeq 0.15 \text{ to } 0.3$
large scale motions	$\Omega_0 \geq 0.2$
gravitational lensing	$\Omega_0 \geq 0.2$

### 3.13 Evidence for the existence and nature of dark matter

Note that some of the methods listed above measure the density of **baryonic matter**. Baryonic matter is matter made up of neutrons, protons (and electrons) – the normal matter we might perhaps think the whole universe is made of. The remaining methods measure the gravitational effect of *all* matter – whether baryonic or non-baryonic. Here we have again written  $H_0 = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}$ .

Estimates of  $\Omega_0$  from visible stars are a factor of about 100 smaller than estimates from galaxy clusters, large scale motions and gravitational lensing. **This provides conclusive evidence for the existence of dark matter.** Dark matter is simply matter that cannot be seen through telescopes, and it can be baryonic or non-baryonic.

Since the values of  $\Omega_0$  from galaxy clusters are a factor of  $\sim 10$  greater than from galaxy rotation curves, it seems that dark matter is not only in galaxy halos, but also *between* galaxies. X-ray observations indicate a smooth distribution of intra-cluster gas in galaxy clusters. Cluster gas is baryonic however, while the constraints on  $\Omega_B$  from nucleosynthesis indicate that a substantial fraction of the dark matter in the Universe is **non-baryonic**.

The limits on  $\Omega_B h^2$  are deduced from the dependence of reaction rates on density and temperature in the early Universe. If  $h = 0.70$ , then

$$0.031 \leq \Omega_B \leq 0.053. \quad (50)$$

These limits are compatible with  $\Omega_0$  from galaxy rotation curves, but fall well short of the limits on  $\Omega_0$  from larger scales, and from e.g. the CMBR (which also independently confirm the nucleosynthesis measurements of the baryon density). Therefore, it seems very likely that a substantial fraction of the dark matter in the Universe is non-baryonic. For example, if  $H_0 \simeq 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ , this implies that some of the dark matter is baryonic. Moreover, if  $\Omega_0 = 1$ ,  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ , then  $> 90\%$  of dark matter would be non-baryonic.

Matter may be baryonic or non-baryonic and bright or dark. It appears that most of the matter in the Universe is both dark and non-baryonic.

### 3.14 Dark matter candidates

#### Baryonic

- Gas clumps in galaxy halos and clusters.
- MACHOs : Massive compact halo objects, such as ‘Brown dwarfs’ (failed stars), ‘Jupiters’ (cold planet-like objects) and undetected white dwarfs (now unlikely, after HST).
- Low surface brightness galaxies.

#### Non-Baryonic

- WIMPs: Weakly interacting massive particles, such as massive neutrinos, exotic particles (axions, photinos, magnetic monopoles, wimpzillas, ...) and primordial black holes.

Note that if primordial black holes form before nucleosynthesis then they don’t affect the limits on  $\Omega_B$ , so they are effectively non-baryonic.

### 3.15 Hot or cold dark matter?

Non-baryonic dark matter interacts weakly with baryons and photons *now*, but interacted more strongly (i.e., was more strongly **coupled** to them) in the early Universe, which was hotter and denser. If non-baryonic dark matter was moving **relativistically** ( $v \simeq c$ ) at the time of decoupling from baryonic matter, we call it **hot**. Examples include neutrinos and photinos. If non-baryonic dark matter was moving **non-relativistically** ( $v \ll c$ ) at decoupling, we call it **cold**. Examples include axions and monopoles.

## 4 Section IV: The early Universe

### 4.1 The Hot Big Bang

In the **hot big bang model** the Universe began with a fireball about 14 billion years ago. Consequently the early phase of the Universe was much hotter and denser than the present day. We could also see this from our derivation of Friedmann's equation in Section III, where we assumed mass conservation, which implies that the density,  $\rho$ , increases as the scale factor,  $R$ , decreases.

At time  $t \rightarrow 0$  we have  $R \rightarrow 0$ , and the density of the Universe approaches infinity.

### 4.2 The cosmic microwave background radiation

In 1946 Gamow predicted the existence of relic radiation from the 'Hot Big Bang'. Peebles and Dicke (1964) developed this theoretical work further, and proposed the building of an antenna to detect the radiation. The radiation was predicted to be **isotropic** and to be **blackbody** in form, fully specified by a single temperature  $T$  (see Fig. 14 and also A1Y Stellar Astrophysics). **Wien's Law** states

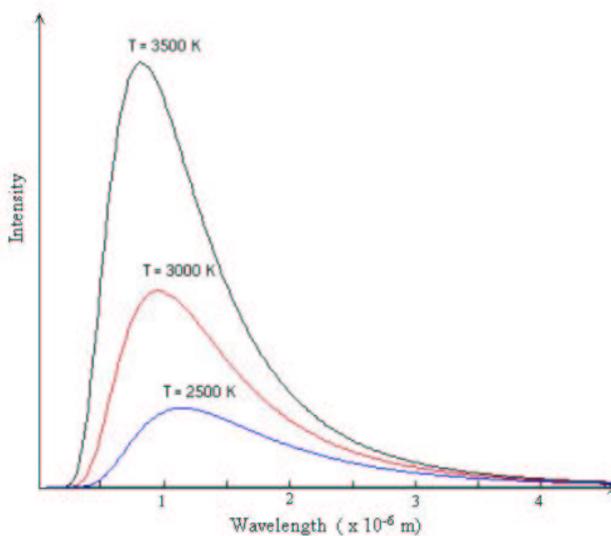


Figure 14: Black body radiation curves for different temperatures. Vertical axis shows the radiation intensity as a function of wavelength.

that the wavelength,  $\lambda_{\text{max}}$ , at which the blackbody radiation curve

is a maximum is inversely proportional to its temperature<sup>1</sup>. Since the wavelength of photons in this relic radiation should increase with the expansion of the Universe, proportionally to the scale factor  $R(t)$ , the temperature of the radiation should be *inversely* proportional to the scale factor, i.e.,

$$T \propto R^{-1} \quad \text{and} \quad \frac{T(t)}{T_0} = 1 + z, \quad (51)$$

where  $T(t)$  is the temperature of the relic radiation a time  $t$  after the Big Bang (corresponding to redshift  $z$ ) and  $T_0$  is the temperature of the radiation today.

In 1965 Penzias and Wilson discovered the Cosmic Background Radiation by accident as they tried to account for excess noise in their receiving equipment. They found that the radiation was highly isotropic, and appeared to have a uniform temperature of about 3 K. This corresponds to a  $\lambda_{\text{max}}$  in the microwave region of the spectrum, and the radiation is now known as the **cosmic microwave background radiation** (CMBR).

Much later, the CoBE (Cosmic Background Explorer) satellite (launched in 1989) confirmed that the spectrum of the background radiation is blackbody to a remarkable level of precision (see Fig. 15) and refined the mean temperature of the radiation to  $T_0 = 2.725$  K. The CoBE results showed that the CMBR is isotropic to better than one part in  $10^4$  so, **the CMBR provides excellent support for the cosmological principle and the Big Bang model.**<sup>m</sup>

The Universe is filled with microwave radiation (the CMBR) that is the highly redshifted glow of the early Universe.

### 4.3 Where does the CMBR come from?

Gamow's prediction that the Universe should be filled with relic blackbody radiation from the Big Bang was based on the idea that, in the Hot Big Bang model, the Universe was hotter and denser in the past. In particular, the very early Universe was much too hot for neutral atoms to exist – it was **fully ionised**<sup>n</sup>, and consisted of a dense mix of free protons and electrons, bathed in blackbody radiation. The free electrons scattered the photons to such an extent, however, that the Universe was effectively **opaque** – analogous to being in a fog. We say that matter and radiation

<sup>1</sup>See A1X notes and A1Y stellar astrophysics handout on blackbody radiation.

<sup>m</sup>More recently (2003) the WMAP satellite has made even more precise measurements of the CMBR, but these are primarily of the *fluctuations* in the radiation (see below)

<sup>n</sup>We will see later that, for a tiny fraction of a second after the Big Bang, the Universe was too hot for even sub-atomic particles to exist.

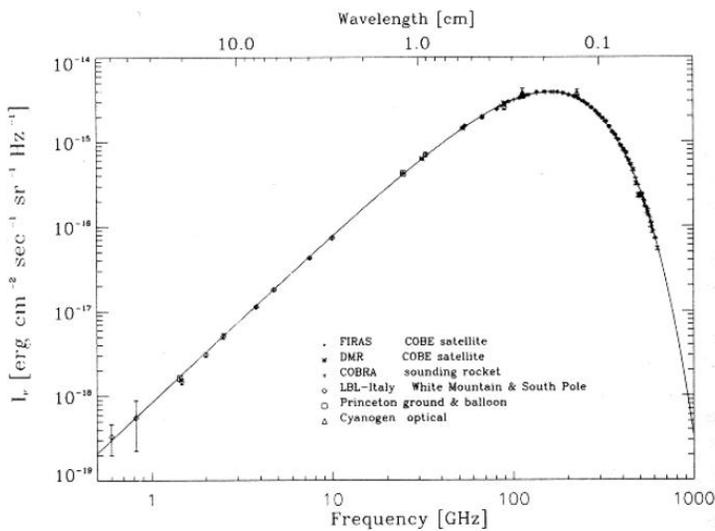


Figure 15: Measurements of the CMBR. The fit to a blackbody curve is very good.

were **coupled**, since the photons interacted so strongly with the free electrons.

As the Universe expanded and cooled, eventually its mean temperature dropped to about 3 000 K. At this temperature the free protons and electrons could combine to form neutral hydrogen, which was hugely less effective at scattering the blackbody photons. As a result the ‘fog’ cleared, and photons could propagate freely through space without any further scattering by the matter in the Universe (see Fig. 16). We call this time the **epoch of recombination**; at this epoch matter and radiation are said to have **decoupled**. The CMBR consists of photons which were emitted

The CMBR comes to us from the time when the Universe had cooled to about 3 000 K and became transparent to the radiation.

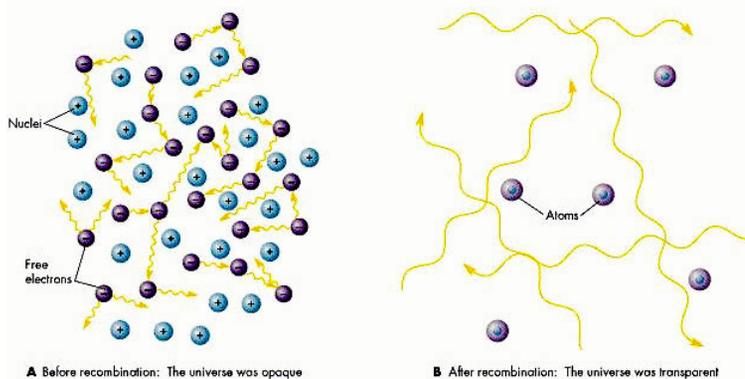


Figure 16: The Universe at recombination.

ted at the epoch of recombination, and which have been travelling towards us ever since – cooling as the Universe continued to expand, but no longer being scattered. For this reason the CMBR is

often referred to as coming from the **surface of last scattering**. The CMBR photons were emitted about  $3.8 \times 10^5$  yr after the Big Bang, although the exact time depends on details of the cosmological model. So the CMBR gives us a glimpse of the Universe when it was only about 0.003% of its present age. The redshift of the CMBR is  $z_{\text{CMBR}} \simeq 1000$  (following from Eq. 51, above).

#### 4.4 Why does the CMBR have a temperature of 3000 K?

The typical energy of a black body photon of temperature  $T$  is given by  $E_{\text{typical}} = kT$ , where  $k$  is Boltzmann's constant. Taking  $T \simeq 3000$  K we find that  $E_{\text{typical}} \simeq 4.14 \times 10^{-20}$  J, or  $E_{\text{typical}} \simeq 0.26$  eV. However, we know from A1Y Stellar Astrophysics that the **ionisation energy** of hydrogen is  $E_{\text{ion}} = 13.6$  eV. This presents us with something of a puzzle: we see that, for  $T \simeq 3000$  K,  $E_{\text{typical}} \ll E_{\text{ion}}$ .

Expressing this puzzle another way, suppose we ask at what black body temperature does the the *mean* energy of a photon equal the ionisation energy of hydrogen? Plugging in the numbers we see that  $T \simeq 15800$  K. Hence, for all  $T < 15800$  K, the mean photon energy is *less* than the ionisation energy of hydrogen. Why, then, wasn't the CMBR emitted when the Universe had a mean temperature of 15800 K?

An answer follows from the fact that black body photons have a *distribution* of energies, as discussed in A1Y Stellar Astrophysics. In particular, at any black body temperature there is a long 'tail' of photons with energies  $E > kT$ . This tail falls off rapidly at large  $E$  – i.e. the fraction of photons with energy  $E \gg kT$  is very small. However, in the very early Universe there were about  $10^9$  photons for every proton and electron. This means that for  $T \leq 15800$  K, although the *mean* photon energy was insufficient to ionise a hydrogen atom, there were enough high energy photons in the tail of the distribution to keep the Universe ionised.

It was only once the temperature of the Universe reached about 3000 K that the fraction of high energy photons, with  $kT \geq 13.6$  eV, became sufficiently small that effectively *none* of the hydrogen in the Universe was left in ionised form.

#### 4.5 Matter-dominated and radiation-dominated epochs

From Einstein's famous formula,  $E = mc^2$ , we can define the **energy density** of matter in the Universe to be

$$u_{\text{matter}} = \rho_{\text{matter}}c^2 \quad (52)$$

where  $\rho_{\text{matter}}$  is the matter density of the Universe. The energy density of blackbody radiation of temperature  $T$  is given by the equation

$$u_{\text{radiation}} = \frac{4\sigma}{c} T^4 \quad (53)$$

where  $\sigma$  is the Stefan-Boltzmann constant (see A1Y Stellar astrophysics).

The Universe is currently **matter dominated**, meaning simply that  $u_{\text{matter}} \gg u_{\text{radiation}}$ . The dependence of the matter and radiation energy density on the scale factor is different however. From mass conservation it follows that  $\rho_{\text{matter}} \propto R^{-3}$  and therefore  $u_{\text{matter}} \propto R^{-3}$ , while from  $T \propto R^{-1}$  it follows that  $u_{\text{radiation}} \propto R^{-4}$ . Therefore

$$\eta = \frac{u_{\text{radiation}}}{u_{\text{matter}}} \propto R(t)^{-1}. \quad (54)$$

The present-day value of  $\eta$  is roughly  $3.2 \times 10^{-4}$ , so  $\eta = 1$  for  $R \simeq 3.2 \times 10^{-4} R_0$ . We call that the epoch of **matter-radiation equality**. It is the epoch at which the mean energy densities of matter and radiation are equal. It follows that

$$\frac{R_0}{R_{\text{eq}}} \simeq 3.2 \times 10^3 \quad \text{i.e.,} \quad z_{\text{eq}} \simeq 3.2 \times 10^3. \quad (55)$$

Also

$$T_{\text{eq}} \simeq 3.2 \times 10^3 T_0 \simeq 9000 \text{ K} \quad (56)$$

For  $T > T_{\text{eq}}$  the Universe is **radiation dominated**. Therefore, when the CMBR photons were emitted the Universe was matter dominated but only about 160,000 years earlier<sup>o</sup> the Universe had been radiation dominated. It is something of a coincidence that both the epochs of recombination and matter-radiation equality occur within a short space of time, relative to the current age of the Universe.

The energy in matter is presently greater than the energy in radiation (the CMBR). In the early Universe, radiation energy dominated.

## 4.6 The Primordial Universe

Although the Universe is opaque to photons beyond the CMBR, the success of the Big Bang model does not stop at recombination. The Big Bang model remains an accurate description of how matter and radiation evolve at much earlier times, and therefore at much higher densities and temperatures. Our ideas about the *primordial Universe* lie at the interface between cosmology and particle physics. Formally, as  $t \rightarrow 0$ , the density and temperature of the Universe in the Big Bang model tend to infinity – an

<sup>o</sup>As with recombination, the exact time depends on details of the cosmological model.

indication that the Big Bang model (and indeed what is known as the ‘Standard Model’ of particle physics, which describes all elementary particles) breaks down. Both models are, nonetheless, valid from  $t \simeq 10^{-40}$  seconds after the Big Bang, when the temperature of the Universe was  $T \simeq 10^{27}$  K. **The Big Bang can, therefore, be thought of as a natural particle accelerator, able to test particle physics theories beyond terrestrial limits.**

The physics of the primordial Universe lies mainly beyond the scope of this course, but we will pick out a few highlights.

Present theories of the Universe can only get to within about  $10^{-40}$  s of the Big Bang.

#### 4.6.1 GUT scale: $T \simeq 10^{27}$ K; $t \simeq 10^{-35}$ s

This is the approximate energy scale above which **grand unified theories** hold, so that three of the four fundamental forces of nature (the electromagnetic force and the strong and weak nuclear forces) behave as if they are a single unified force. Only gravity does not fit within this unified picture. (To unify gravity too would require a so-called ‘Theory of Everything’ or full theory of quantum gravity, which still eludes us – although we have some clues as to what that theory might be).

At this time the Universe consists of a ‘soup’ of fundamental particles: **quarks** and **leptons**, and radiation. There are six different types of quarks, and these are believed to be the real indivisible building blocks of all matter. Quarks can combine together to form **hadrons** such as protons and neutrons which each consist of a triplet of quarks, but above the GUT scale the Universe is still too hot for this to happen. Leptons are light, elementary particles including electrons and positrons.

Above the GUT scale there is spontaneous conversion of particle antiparticle pairs into radiation and vice versa. The Universe contains equal amounts of matter and antimatter during this time. Below the GUT scale, the strong nuclear force breaks off from the unified picture, and a tiny imbalance between matter and antimatter arises. The remaining, equal, amounts of matter and antimatter annihilate each other, leaving the Universe consisting of matter that we see today.

#### 4.6.2 ‘Electroweak transition’: $T \simeq 10^{15}$ K; $t \simeq 10^{-12}$ s

Below this energy scale the Universe has cooled sufficiently that the weak nuclear and electromagnetic forces are no longer unified. From now on there are four distinct fundamental forces (as today!).

### 4.6.3 ‘Quark-hadron transition’: $T \simeq 10^{12} \text{ K}$ ; $t \simeq 10^{-6} \text{ s}$

At this energy scale the ‘quark soup’ condenses. The Universe has cooled enough to form stable hadrons such as protons and neutrons. The quarks are now confined inside hadrons, and no longer exist as free particles. The Universe now consists of hadrons and leptons, bathed in a background of blackbody radiation (as today, only much hotter!).

### 4.6.4 Primordial nucleosynthesis: $T \simeq 10^9 - 10^8 \text{ K}$ ; $t \simeq 1 - 100 \text{ s}$

At this time the Universe has cooled sufficiently to allow protons and neutrons to combine together and form stable light atomic nuclei: deuterium and tritium (isotopes of hydrogen), helium and lithium. As we remarked in Section III, the relative amounts of these elements ‘cooked’ during the first 100 s or so after the Big Bang depends strongly on the density of baryons ( $\rho_B$ ) since this determines the rates of the nuclear reactions which take place. We call the formation of these light atoms **primordial nucleosynthesis**.

Note that since we can write  $\rho_B = \Omega_B \rho_{\text{crit}}$  and  $\rho_{\text{crit}} \propto H_0^2$  (see Section III), it follows that  $\rho_B \propto \Omega_B h^2$ .

We can compute the *relative abundances* of helium, deuterium and lithium to hydrogen predicted in the Big Bang model as a function of  $\Omega_B h^2$ , and compare each ratio with the observed limits, based on the abundances measured today. Fig. 17 shows schematically the current limits on the baryon density from primordial element abundances. Each light element provides an independent check of the Big Bang model. The solid curves denote the theoretically predicted abundances, for different values of  $\Omega_B h^2$ . The dotted lines denote the observational limits on the relative abundances.

There is consistent agreement between observations and predictions *for all the light elements* over a narrow range of values of  $\Omega_B h^2$  (as shown by the vertical grey bands in Fig. 17). This is one of the major successes of the Big Bang and places strong constraints on the value of  $\rho_B$ .

The relative abundance of helium is about 25%; other light elements are very rare. **All heavier elements are manufactured inside stars and supernovae – see A1Y Stellar Astrophysics.** It is interesting to note that we can now also check the nucleosynthesis constraints at high redshift, by deducing the light element abundances at (for example)  $z \sim 1$  from quasar absorption spectra – we mentioned this point at the end of Section I.

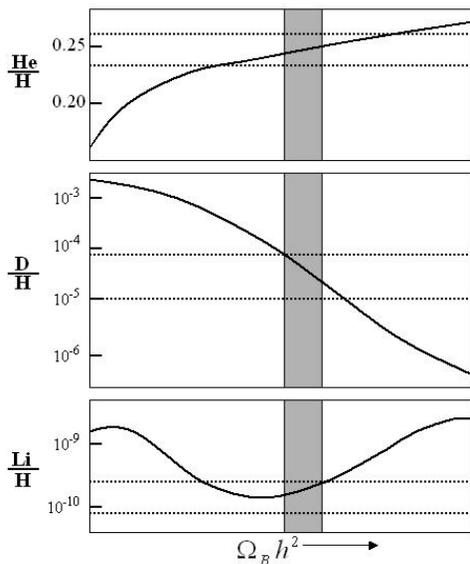


Figure 17: Primordial nucleosynthesis and the baryon density. The grey band shows the value of  $\Omega_B h^2$  consistent with the observations.

## 4.7 Microwave background anisotropies and galaxy formation

The CMBR is not *perfectly* smooth. In 1992, CoBE first detected temperature variations of about  $10^{-5}$  K on angular scales of about  $10^\circ$  (see Fig. 18). Previously, CoBE had confirmed the existence of what we call a **dipole** temperature variation in the CMBR, of the form

$$T(\theta) = T_0(1 + \delta T \cos \theta), \quad (57)$$

where  $T_0$  is the mean temperature of the CMBR, averaged over the whole sky and where  $T(\theta)$  is the temperature of the CMBR at an angle  $\theta$  away from the ‘CMBR hotspot’ direction – the direction in which the temperature is  $T_0(1 + \delta T)$ . CoBE measured

$$\Delta T = 3.35 \times 10^{-3} \text{ K}. \quad (58)$$

This dipole anisotropy is *not* believed to be intrinsic to the CMBR, but instead is due to our peculiar motion with respect to the CMBR, which causes a Doppler Shift of the radiation of an amount which varies with direction according to the above dipole formula. The Doppler formula gives

$$\frac{\Delta T}{T_0} = \frac{v_{\text{pec}}}{c}. \quad (59)$$

Taking  $T_0 = 3 \text{ K}$  gives  $v_{\text{pec}} = 330 \text{ km s}^{-1}$ . In 2003 the WMAP satellite confirmed the earlier CoBE measurement of the CMBR dipole temperature and direction.

After removing the dipole anisotropy, and also the contaminating effect of the microwave emission from stars and dust in our

own Milky Way galaxy, the CoBE map reveals intrinsic temperature fluctuations of about  $10^{-5}$  K. These variations have now been confirmed (and improved) by a large number of other experiments, which make measurements of the temperature variation on a range of different angular scales – most notably the new results from the WMAP (Wilkinson Microwave Anisotropy Probe) Satellite in 2003 and other ground-based results, which show fluctuations on much smaller angular scales than in the CoBE maps.

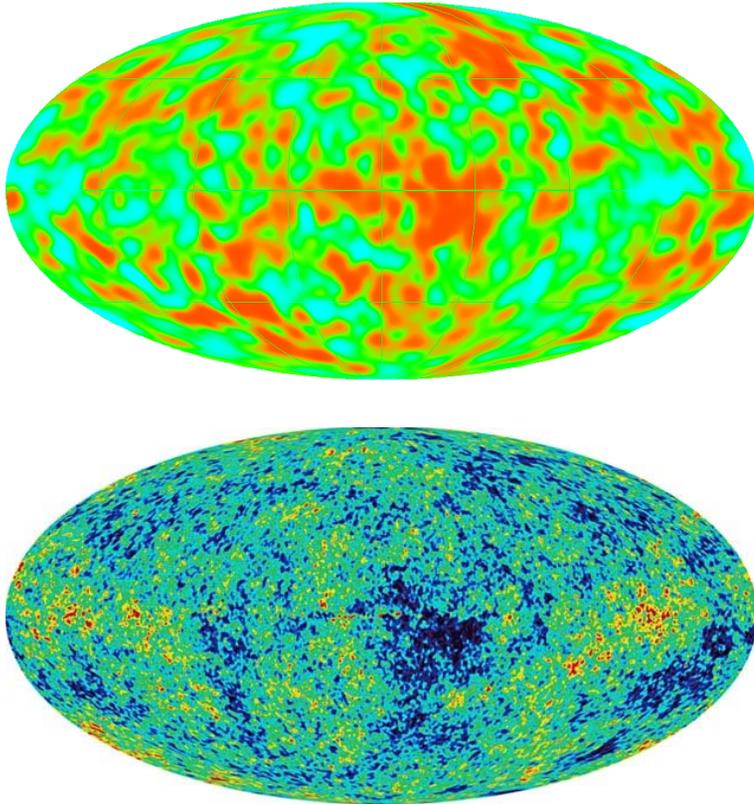


Figure 18: Maps of CMBR temperature fluctuations. The upper panel (from Jan 1996) shows results from the first four years of CoBE data. The lower panel (from Feb 2003) shows results from the first year of data from the WMAP satellite.

These intrinsic anisotropies are *very* important for cosmology, since they indicate density inhomogeneities in the Universe at  $t = 3.8 \times 10^5$  yr after the Big Bang. These density fluctuations are the seeds of the cosmic structure which we observe today, such as galaxies, clusters and superclusters. The fluctuations are believed to have been created very soon after the Big Bang itself, during a period known as **inflation** when the Universe expanded extremely rapidly.

The mechanism for structure formation is **gravity**, which causes the density inhomogeneities to grow as the Universe expands. Structure evolves under the influence of gravity, and the amount and nature of the dark matter determines how structure forms at

different times and on different scales. By comparing model predictions for different structure formation scenarios with the observed Universe, we can determine which model is correct. For example, **hot dark matter** smooths out clustering on small scales, so that in models with hot dark matter one expects large structures to form first and then later fragment. On the other hand, in models where the dark matter is cold structures form on both small and large scales from the outset. More specifically we can also determine very precisely how much dark matter there is and how it is distributed by careful study of the pattern of temperature variations in the CMBR and in the clustering of galaxies. Recall from Section III that we mentioned that these measurements can also reveal the **curvature** of the Universe; the results from WMAP seem to indicate that the Universe is **flat**.

## 4.8 Understanding the formation of galaxies and larger structures

This is one of the biggest challenges facing modern cosmology. The task is to explain how the Universe got from the (nearly) smooth CMBR which we observe at  $z = 1000$  to the ‘lumpy’ galaxies and clusters which we observe at  $z = 0$  (now). One can think of this task as trying to find the correct recipe for galaxy formation. The ingredients are known to be

- CMBR temperature fluctuations,
- dark and luminous matter, and
- gravity,

and the cooking time is about 14 billion years! While a definitive model for structure formation which matches the observations on *all* scales and at all epochs has yet to be found, it is likely that a solution to this problem will be found within the next decade.