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 $70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$.

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 $H_0 \simeq 500\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. 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 $1\,000\,\mathrm{km\,s^{-1}}$. 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 $50\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ and another camp argued for a value close to $100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. It was (and still is, in many contexts) common to write $H_0 = 100h\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, 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.

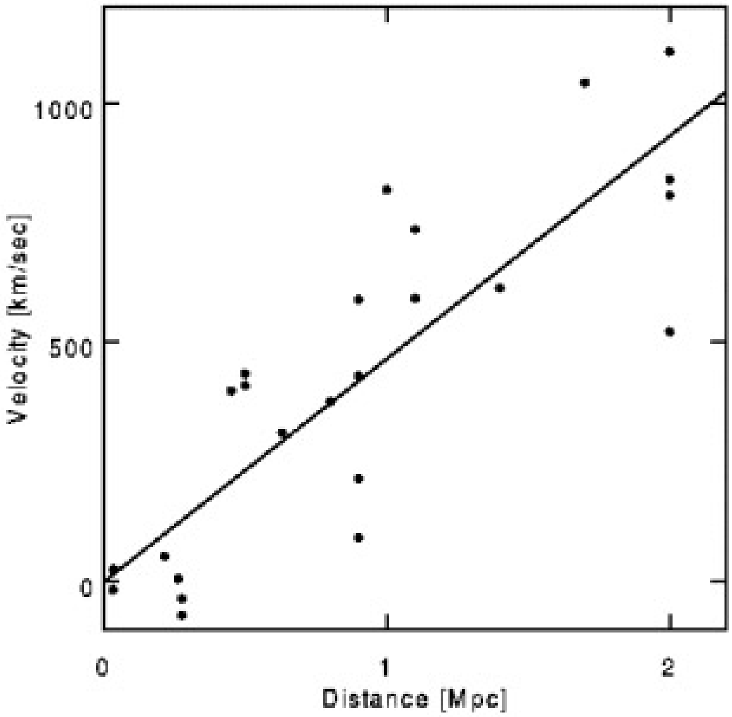

Figure 9: Hubble's original data. (Note that the slope Hubble measured was $\simeq 500\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$).

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 $d \simeq 20$ Mpc, the observed radial velocities of galaxies are still seriously affected by peculiar motions, so that $v_{\rm rec} \neq H_0 d$,
- we require secondary distance indicators to extend from ~ 20 Mpc to ≥ 100 Mpc, where Hubble's law holds more accurately, so that $v_{\rm rec} = H_0 d$.

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**.

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

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, $m_{\rm max}$, versus the log of the recession velocity. Neglecting extinction and peculiar motions, this should be linear for a nearby standard candle because

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

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

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

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.

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

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 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-σ relation, and it is a special case of the so-called *Fundamental Plane* relation for ellipticals. Calibrating the D_n-σ 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.

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.

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 kms^{-1} 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 km s^{-1} 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.

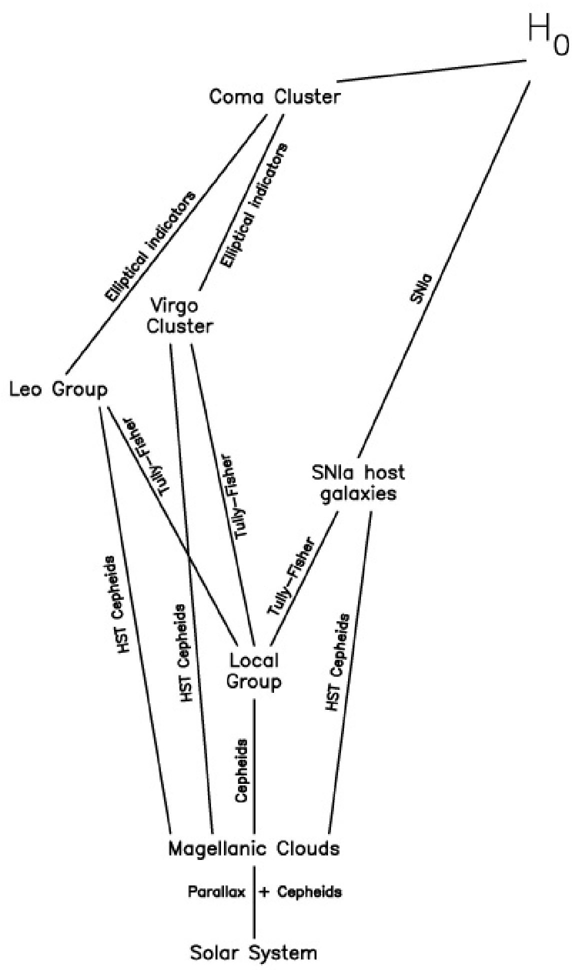

Figure 10: A schematic of the cosmological distance ladder.

3.6 The basis of cosmological models

3.6.1 Olbers's Paradox

Before we investigate the details of modern Big-Bang models, it is useful to consider what deductions we can make about the Universe[h] from much more elementary observations: specifically what can one deduce about the Universe from the simple fact that the night sky is dark?

Heinrich Olbers (1758-1840) posed the question "why is the sky dark at night?". He realised that a dark night sky is paradoxical *if* (as many philosophers of the time believed) the Universe is infinite in extent and eternal, with stars roughly uniformly distributed throughout space. This is because, in an infinite and eternal universe, eventually *every* line of sight will intercept a star, so that the whole night sky should be as bright as the surface of a star. This puzzle is known today as **Olbers's Paradox**. Another way

[h]We will talk about the Universe, but models of different universes.

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 tells us that the Universe is *not* infinite, eternal and unchanging.

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 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

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

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.[i] 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**:

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 $\sim 10\,000\ \mathrm{km\,s^{-1}}$, 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.

[i]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

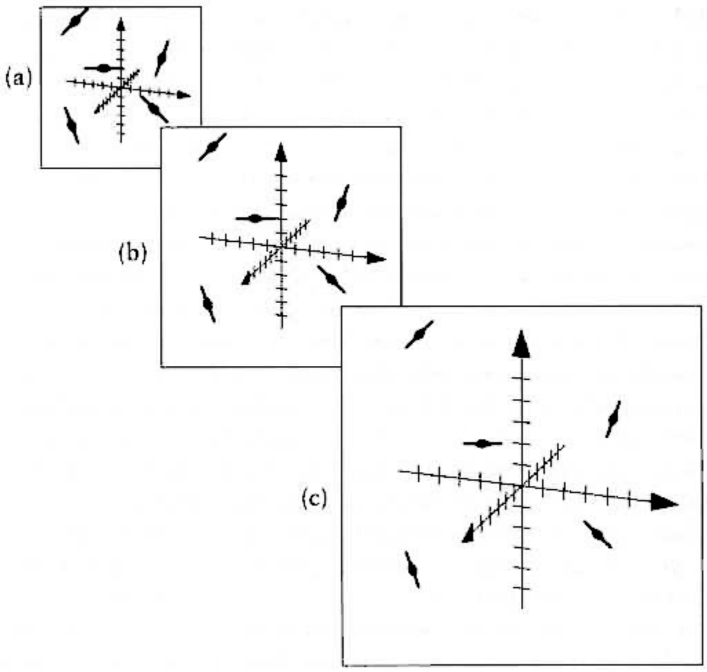

Figure 11: Cartoon picture of universal expansion. Note that both the axes and tick-marks grow, but the galaxies do not.

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

$$r(t) = R(t) \times s. \tag{20}$$

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

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

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

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.,

$$v = \frac{\mathrm{d}r}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}(Rs) = \dot{R} \cdot s = \frac{\dot{R}}{R} \times (Rs) = \frac{\dot{R}}{R} r. \tag{23}$$

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.

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

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

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

and the present day value of the Hubble constant is

$$H_0 = \frac{\dot{R}_0}{R_0}. \tag{25}$$

3.6.6 When was the Big Bang?

We define the Big Bang by the condition that $R(t) \to 0$ 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

$$v = H_0 r = \frac{\text{distance}}{\text{time}} = \frac{r}{t}. \tag{26}$$

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

$$\tau = H_0^{-1}, \tag{27}$$

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

$$\tau = \frac{978}{H_0} \times 10^9 \,\text{yr}. \tag{28}$$

We call τ the **Hubble time**, and it sets a timescale for the expansion

Hubble's constant has dimensions of 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

$$t_0 < \tau. \tag{29}$$

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 ρ. Then

$$M = \frac{4}{3}\pi r^3 \rho. \tag{30}$$

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

$$\mathrm{KE} = \frac{1}{2}m\dot{r}^2 = \frac{1}{2}m\dot{R}^2 s^2 \tag{31}$$

and the **potential energy** is given by

$$\mathrm{PE} = -\frac{GMm}{r} = -\frac{4}{3}\pi R^2 s^2 G\rho m. \tag{32}$$

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

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

or equivalently

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

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.[j] (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 ρ 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.

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

Einstein fixed this problem by introducing an extra constant, Λ ('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 Λ, Einstein could obtain a static solution with $R(t)$ = 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 Λ behaves rather like 'anti-gravity': a repulsive force which overcomes the attraction of gravity on very large scales.

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

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 Λ 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 Λ 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 $\Lambda = 0$,

[j] 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.

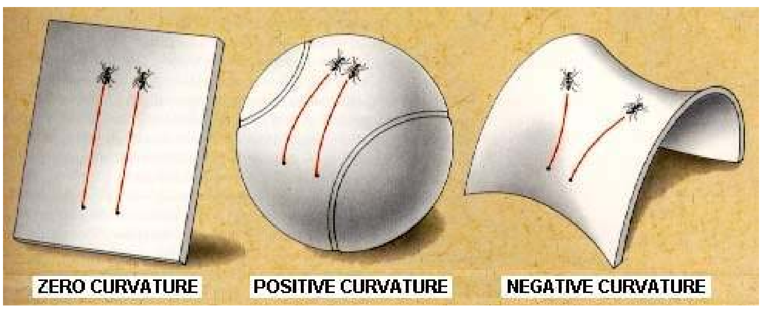

Figure 12: 2-D representation of surfaces of different curvature.

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 $R \to \infty$.

(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 the case of a flat universe ($k = 0$), so that

$$\left(\frac{\mathrm{d}R}{\mathrm{d}t}\right)^2 = \frac{8\pi G\rho R^2}{3}. \tag{36}$$

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

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

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

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

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

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

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

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

i.e.,

$$t_0 = \frac{2}{3}H_0^{-1}. \tag{41}$$

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

$$t_0 = \frac{652}{H_0} \times 10^9\ \mathrm{yr}. \tag{42}$$

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

Therefore, for $H_0 = 70\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$, $t_0 \simeq 9 \times 10^9$ 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[k], astronomers

[k] 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 Λ 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

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

or

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

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

$$\rho_{\rm crit} = \frac{3H_0^2}{8\pi G}. \tag{45}$$

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

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

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

$$\Omega(t) = \frac{\rho(t)}{\rho_{\rm crit}(t)}. \tag{46}$$

Therefore

$\Omega > 1$ implies Universe **closed**,

$\Omega < 1$ implies Universe **open**,

$\Omega = 1$ implies Universe **flat**.

Cosmologists denote the present-day value of Ω by Ω_0. Although Ω 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 Ω_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.

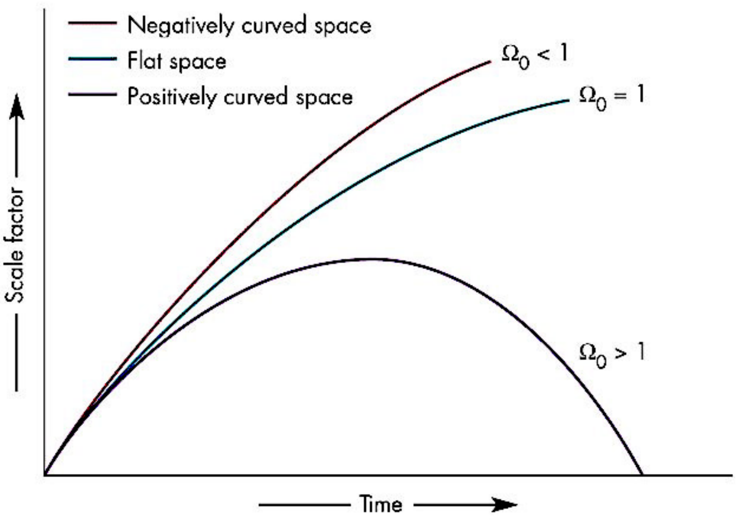

Figure 13: Behaviour of the scale factor in different cosmologies.

3.12 Weighing the Universe: methods for measuring Ω_0

So which universe do we live in? To answer that question we need to know Ω_0, which depends on the matter density of the Universe, ρ. 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

$$\rho = \frac{N_{\rm stars} \times M_\odot}{\text{volume of Milky Way}}. \tag{47}$$

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

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

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

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

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 (Ω) and the cosmological constant (Λ). 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, Λ. 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 Ω_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 Ω_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	$0.015 \le \Omega_{\rm B} h^2 \le 0.026$
CMBR (baryons)	$\Omega_B h^2 = 0.0224 \pm 0.0009$
CMBR (baryons + non baryons)	$\Omega_0 = 0.27 \pm 0.04$
visible stars	$\Omega_{\rm B} \simeq 0.002$
galaxy rotation curves	$\Omega_0 \simeq$ 0.01 to 0.02
galaxy clusters	$\Omega_0 \simeq$ 0.15 to 0.3
large scale motions	$\Omega_0 \ge 0.2$
gravitational lensing	$\Omega_0 \ge 0.2$

3.13 Evidence for the existence and nature of dark matter

Note thatsome 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 $H_0 = 100h$ km s^{-1} Mpc^{-1}.

Estimates of Ω_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 Ω_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 $\Omega_{\rm B}$ 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

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

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.

These limits are compatible with Ω_0 from galaxy rotation curves, but fall well short of the limits on Ω_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 $H_0 \simeq 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, this implies that some of the dark matter is baryonic. Moreover, if $\Omega_0 = 1$, $H_0 = 70\ \mathrm{kms^{-1}\ Mpc^{-1}}$, then $> 90\%$ of dark matter would be 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 Ω_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.