

STATISTICAL PHYSICS AND COSMOLOGY

Part IIA Mathematical Tripos

Prof. P.K. Townsend

DAMTP, CMS

Wilberforce Road, Cambridge CB3 0WA

E-mail: P.K.Townsend@damtp.cam.ac.uk

ABSTRACT:

- (I) The Expanding Universe (5)
- (II) Statistical Mechanics and Thermodynamics (4)
- (III) Stars and gravitational collapse (3)
- (IV) Thermal History of the Universe (4)

There are also three example sheets (which have corresponding answer sheets).
Starred sections may be ignored but could be helpful.

RECOMMENDED BOOKS:

A. Liddle, *An Introduction to Modern Cosmology* (Wiley 1999)

F. Mandl, *Statistical Physics*, **Second Edition** (Wiley 1988)

A.C. Phillips, *The Physics of Stars* (Wiley 1994)

S. Weinberg, *The First Three Minutes*, (Basic Books, 1993)

E.R. Harrison, *Darkness at Night*, (Harvard U. Press, 1987)

E.R. Harrison, *Cosmology: The Science of the Universe* **Second Edition** (C.U.P. 2000).

PREREQUISITES: Newtonian mechanics, including gravity; Elementary QM and SR.

Contents

1. The Expanding Universe	2
1.1. Isotropy and Homogeneity	2
1.2. Static vs Dynamic Universe	5
1.3. Olbers' paradox	7
1.4. Hubble's law	8
1.5. Scale factor of the Universe	10
1.6. Dynamics of expansion	11
1.7. The Big Bang	14
1.8. Flatness problem	16
1.9. Equation of State and Cosmological Models	17
1.9.1. Einstein-de Sitter universe	18
1.9.2. Tolman universe	18
1.10. Redshift formula	19
1.11. The cosmological horizon	20
1.12. The horizon problem	21
1.13. The Cosmological Constant	21
1.14. The inflationary universe *	23
2. Statistical Mechanics	24
2.1. Entropy and the laws of Thermodynamics	24
2.2. Chemical potential	27
2.3. Thermodynamic Ensembles	28
2.4. The Gibbs Distribution	29
2.5. Ideal Gases	30
2.6. Bosons and Fermions	32
2.7. Density of states	33
2.8. Ideal Gas Pressure	34
2.9. Classical NR Gas	36
2.10. Fermi Gas and degeneracy pressure	37
3. Stars and gravitational collapse	40
3.1. Thermal Pressure Support	40
3.2. Virial Theorem	42
3.3. Elementary particle interactions	44
3.4. Stellar evolution	45

3.5. White Dwarfs	47
3.6. Neutron Stars and Black Holes	50
4. Thermal History of the Universe	52
4.1. Photon gas	52
4.2. Planck's radiation formula and Wien's law	53
4.3. Photon pressure from Kinetic Theory *	55
4.4. Photon gas entropy	57
4.5. Adiabatic expansion and the CMBR	58
4.6. The baryon to photon ratio	59
4.7. Recombination	60
4.8. Decoupling	62
4.9. The early universe	63
4.10. Primordial nucleosynthesis	64
5. Example sheets	67
5.1. Example Sheet 1	67
5.2. Example Sheet 2	69
5.3. Example Sheet 3	72
6. Answers to Example sheets	74
6.1. Answers to Example Sheet 1	74
6.2. Answers to Example Sheet 2	76
6.3. Answers to Example Sheet 3	78

1. The Expanding Universe

1.1. Isotropy and Homogeneity

There is a long history of cosmological models. Those of interest to us have one or both of two features: *isotropy* (same in all directions, from some 'central' point) and *homogeneity* (same at all points).

If we ignore the planets and the Milky Way then the distribution of stars in the night sky is approximately isotropic because one is as likely to see a star in one direction as any other one. This feature, isotropy with the Earth as the centre, was a common one in pre-scientific models of the cosmos in which the stars were presumed to lie on a 'celestial sphere' centred on the Earth. Copernicus shifted the centre from the Earth to the Sun,

but the Copernican universe was not even approximately homogeneous because the Sun occupies a privileged position.

Fig. I.1

Copernicus had no reason to think that the Sun might be just another star, and there was an argument against this possibility. The apparent brightness of the Sun depends on how close one is to it, as does the solid angle it subtends in the sky. If one goes far enough away for the Sun to be as faint as a star then the solid angle it subtends would be correspondingly small—too small to measure. A star, on the other hand, appears to have a measurable angular size. Galileo demolished this argument by showing that the apparent size of a star as seen through a telescope is independent of the magnification; it follows that the apparent size cannot be an indication of actual size. This observation also shows that stars must be very far away, and hence very bright. By Newton’s day the best guess was that stars are indeed suns like our own. Stars play a key role in Newtonian mechanics because inertial frames are, mysteriously, those in which the stars appear to be at rest (Newton’s Bucket). Despite the existence of the Milky Way (observed by Galileo to be a band of closely packed stars) Newton supposed the stars to be distributed at random, but *uniformly*, throughout an infinite universe. This universe is one that is approximately homogeneous—it would look essentially the same from any point. It is also approximately isotropic—pick our Sun as the central point; then the distribution of stars in the sky must be approximately isotropic because it could not otherwise be uniform. Of course the same must be true whichever star we pick as the central one, by homogeneity: **a homogeneous universe that is isotropic about one point is isotropic about all points.**

However, a universe that is homogeneous need not be isotropic. The Sun is spinning and the axis of rotation defines a direction. The other stars will likely be spinning too. Suppose that the axes of rotation are all aligned; in this case the Newtonian universe is still homogeneous but it is no longer isotropic because the common axis of rotation defines a ‘preferred direction’.

Fig. I.2

If the axes of rotation are randomly distributed then we recover approximate isotropy, so homogeneity is consistent with isotropy but does not imply it.

Neither does isotropy imply homogeneity—we have already seen that the Copernican universe is not homogeneous but may be considered to be approximately isotropic. How-

ever, **isotropy about three non-colinear points implies homogeneity**. To see why, let C_1 and C_2 be two of the three ‘central’ points about which a given (infinite 3D Euclidean) universe is assumed to be isotropic, and let ℓ_{12} be the line joining them.

Fig. I.3

Consider any other line ℓ passing through C_1 . Any inhomogeneities along ℓ would show up as anisotropies when viewed from C_2 , so this universe must be homogeneous along all lines ℓ through C_1 except, possibly, along ℓ_{12} . This possibility can hardly be of physical relevance, partly because we only expect approximate homogeneity anyway, but if one wants a mathematical theorem one needs a third ‘central’ point C_3 that does not lie on ℓ_{12} . We can then use the same argument but with C_3 replacing C_2 , and since ℓ_{12} is no longer an exceptional line there can be no inhomogeneities along it.

The distribution of stars in the sky is actually neither isotropic nor homogeneous once one takes into account the Milky Way. Of course, once it has been realized that this is our own galaxy, and that there are other galaxies, it is possible that an approximate isotropy and/or homogeneity will re-emerge at the level of galaxies. The typical galactic separation is one Mega parsec (Mpc) but it turns out that their distribution is *not* homogeneous: they group into clusters of galaxies which themselves group into ‘superclusters’. However, on still larger scales, of about 100 Mpc, the Universe does appear to be both homogeneous and isotropic. The best evidence comes from the Cosmic Microwave Background Radiation (CMBR) which we shall consider closer to the end of the course. We shall therefore assume that the Universe obeys

The Cosmological Principle: *On the largest scales, the universe is both homogeneous and isotropic*

1.2. Static vs Dynamic Universe

Newton’s cosmological ideas appear in a letter in which he attempts to answer a possible objection to his theory of gravity. If stars are suns like our own, subject to the attractive

force of gravity, what prevents them from all collapsing on to us? Newton's answer was to claim that a uniform distribution of self-gravitating point masses in an *infinite* universe would be in static, although unstable, equilibrium (the instability would give rise to gravitational clumping, which Newton proposed as the mechanism for the formation of stars). This 'Newtonian universe' is one that is both homogeneous and isotropic, on sufficiently large scales. It is also *static*.

Of course, the real reason that the stars don't fall in on us is that they are in orbit about the galactic centre; they appear to be fixed, in inertial frames, because of their great distance from us (see Q.I.1). However, the problem with stars now returns as a problem with galaxies. Why do the galaxies not all collapse to a central point? In fact, all the galaxies in our local 'supercluster' *do* seem to be accelerating towards a point (the 'great attractor') about 40 Mpc away. However, if the cosmological principle is correct then there must be some unit for which we can again ask the question "why do they not all collapse to a point?" For convenience, we shall continue to speak of these units as 'stars'. Imagine an infinite cubic lattice with a star at each lattice point. Each line through any star has a mass distribution that is symmetric about the star, so the net gravitational force on it is zero and the system is in static equilibrium; at least, that is the gist of Newton's argument. But is it correct?

Choose star A as the centre of an isotropic and homogeneous universe. We may take the rest frame of A to be an inertial frame since the force on A in this frame due to all the other stars vanishes by symmetry. We now need to compute the force \mathbf{F} on some other star B at some distance d from A . Note that isotropy implies that \mathbf{F} is directed along the line AB , unless it vanishes. It is tempting to argue that \mathbf{F} must vanish because we could have chosen B as the centre and we have just concluded that a central star feels no force. But this argument is fallacious. It is true that the force on B must vanish *in its own rest frame*, by homogeneity. However, \mathbf{F} is the force on B *in the rest frame of A* . If \mathbf{F} is non-zero then B is accelerating towards A so the change of reference frame from one in which A is at rest to one in which B is at rest introduces an additional 'fictitious' force on B . By the equality of inertial and gravitational mass (as shown by Galileo's Leaning-Tower-of-Pisa experiment, and called the *equivalence principle* by Einstein) this fictitious force exactly cancels F , in agreement with our earlier conclusion that the force on B must vanish in its own rest frame, so there was no contradiction in supposing \mathbf{F} to be non-zero.

The force \mathbf{F} on star B in the rest frame of A can be expressed as the sum $\mathbf{F} = \mathbf{F}_{<} + \mathbf{F}_{>}$, where $\mathbf{F}_{<}$ is the force due to the stars at $r < d$ (including A itself) and $\mathbf{F}_{>}$ is the force due to the infinite number of other stars at $r > d$. The force $\mathbf{F}_{<}$ can easily be calculated by means of

Newton's Theorem: *For a spherically symmetric mass distribution the gravitational force on a particle at distance d from centre due to the mass within a radius $r < d$ is the*

same as if all the mass with $r < d$ were concentrated at the centre.

Thus $\mathbf{F}_{<}$ is an attractive force directed towards the central star A . If Newton's claim is correct then this should be cancelled by $\mathbf{F}_{>}$. To compute $\mathbf{F}_{>}$, we divide the stars with $r > d$ into shells of radius δr

Fig. I.4

It can be shown (exercise) that the net gravitational force of any such shell on a star at $r = d$ vanishes. Summing over all shells then yields $F_{>} = 0$, in contradiction to Newton!

The problem we have here is that the result of the computation we are trying to do is not really well-defined because we are effectively summing an infinite series that is only conditionally convergent. We could get any answer we wish by an appropriate summation method. Additional input is needed, and this is provided by GR. A theorem of Birkhoff about GR reads as follows when adapted to Newtonian gravity:

Birkhoff's Theorem: *For a spherically symmetric mass distribution the gravitational force on a particle at distance d from the centre is due entirely to the mass at distance $r < d$.*

This justifies the sum-over-shells result that $\mathbf{F}_{>}$ vanishes. We conclude that an isotropic and homogeneous universe cannot be static. Once we allow for a dynamic universe the problem of the stability of the universe must be re-analysed afresh, as we shall do in the following lectures.

1.3. Olbers' paradox

There is another, related, problem with the static Newtonian universe that was first raised by Halley (who proposed one of many erroneous solutions). It was rediscovered in the 19th century by various people, Olbers among them, and it was revived in the 1950's by Bondi, who called it **Olbers' Paradox**. Consider an infinite Newtonian universe, with

a random distribution of stars of uniform number density n (the average number of stars per unit volume). Each star radiates energy in the form of electromagnetic waves; this includes light so we shall call it light energy; the star's *luminosity* is the light energy it radiates in unit time. For simplicity, we shall assume that each star has the same luminosity, L , as our Sun. We want to compute the energy flux from these stars in the neighbourhood of our solar system, which we take as the centre of an isotropic and homogeneous distribution of stars. The flux (energy passing through unit area in unit time) from a single star of luminosity L is

$$\Phi = \frac{L}{4\pi r^2}, \quad (1.1)$$

because Φ can depend only on r (by isotropy) and its integral over a sphere of radius r must equal L . The flux Φ is also called the star's *brightness* because what we conventionally call brightness is a measure of the light energy per unit time impinging on the retina. Now consider a shell of width δr at distance r . The number of stars in this shell is $n \times (4\pi r^2)\delta r$, so the total flux from the shell is $nL\delta r$, *independent of r* . If we were to sum over the infinite number of shells in an infinite space we would arrive at the absurd result that the combined light of all the stars is infinitely bright. This is because a sum over all shells fails to take into account the fact that light from very distant stars will be obscured by closer stars. The average distance to the first star along any line of sight is finite and if we sum over shells up to this distance, but no further, we get a finite result for the total brightness. However, according to this calculation the night sky should be everywhere as bright as the Sun! (see Q.I.2). So why is the sky dark at night? This is Olbers' paradox.

1.4. Hubble's law

Convincing evidence that stars are similar to our own Sun had to await the development of spectroscopy in the 19th century. Each element has its own characteristic spectrum and the chemical composition of a stellar atmosphere can be determined by an examination of its absorption spectrum (dark lines that correspond to frequencies absorbed by elements in a star's outer atmosphere). The element Helium is so called because it was first identified from the Sun's absorption spectrum, and most other stars have a similar fraction of Helium in their atmosphere. Because stars are generally moving with respect to the Sun, their spectral lines are Doppler-shifted.

Consider a light source moving with velocity v , relative to the rest frame of some observation post, and at an angle θ to the line of sight:

Fig. I.5

Let λ_e be the wavelength of the light emitted by the source (as measured in its own rest-frame). Let λ_o be the wavelength of this light as detected at the observation post (in its rest-frame). Because of the Doppler shift we expect that $\lambda_o \neq \lambda_e$. From SR one learns that they are related by the formula

$$\frac{\lambda_o}{\lambda_e} = \frac{\sqrt{c^2 - v^2}}{c - v \cos \theta} = 1 + \frac{v \cos \theta}{c} + \mathcal{O}(v^2/c^2). \quad (1.2)$$

Observations of stars within our galaxy show that, typically, $v \sim 10^5$ m/s. The radial component of the velocity of other galaxies can be similarly deduced from the emission spectra of the stars they contain. A shift of the spectrum towards the red end of the visible spectrum corresponds to positive v , and hence a galaxy that is receding from us. A shift towards the blue end of the spectrum corresponds to a negative velocity of recession and hence to a galaxy that is approaching us. Cosmologists write

$$\frac{\lambda_o}{\lambda_e} = 1 + z \quad (1.3)$$

and call z the ‘redshift’. Nearby galaxies may be either receding from us or approaching us (as is the Andromeda galaxy). However, very distant galaxies always have a redshifted spectrum. In 1929 Hubble reported the results of his measurements of the redshifts of many distant galaxies. Interpreting the results as due to the Doppler shift he deduced that the velocity of recession of a distant galaxy is proportional to its distance r from us. Ignoring the ‘peculiar’ motions of galaxies, then leads to **Hubble’s law**

$$\mathbf{v} = H_0 \mathbf{r}. \quad (1.4)$$

The constant of proportionality H_0 is known as the Hubble constant. Its inverse H_0^{-1} has the dimensions of time and is called the *Hubble time*. Difficulties in determining distances produce a factor of 2 uncertainty in H_0 , so cosmologists usually write

$$H_0^{-1} = h^{-1} \times 10^{10} \text{ yrs} \quad (1.5)$$

where h is the ‘Hubble factor’ constrained by observations to lie within the range

$$0.5 < h < 1. \quad (1.6)$$

Hubble proposed his law on the basis of a straight line fit to a set of data points, but a straight line was not obviously warranted by his initial data. Most likely, he knew that it ‘ought’ to be a straight line from the following argument. Let X and Y be two galaxies, and let \mathbf{v}_X^Y be the velocity of Y as seen from X . Let Z be a third galaxy with velocity \mathbf{v}_Y^Z as seen from Y .

Fig. I.6

Assuming *non-relativistic* velocities, the velocity of Z as seen from X is

$$\mathbf{v}_X^Z = \mathbf{v}_X^Y + \mathbf{v}_Y^Z. \quad (1.7)$$

Now assume homogeneity. In a homogeneous universe the relative velocity of any two galaxies can depend only on their *relative* position vector. That is,

$$\mathbf{v}_A^B = \mathbf{v}(\mathbf{r}_B - \mathbf{r}_A) \quad (1.8)$$

for some vector field \mathbf{v} , where A and B are any two galaxies, with position vectors \mathbf{r}_A and \mathbf{r}_B , respectively. Using this in (1.7) we deduce that

$$\mathbf{v}(\mathbf{r}_Z - \mathbf{r}_X) = \mathbf{v}(\mathbf{r}_Y - \mathbf{r}_X) + \mathbf{v}(\mathbf{r}_Z - \mathbf{r}_Y). \quad (1.9)$$

Defining $\mathbf{r} = \mathbf{r}_Y - \mathbf{r}_X$ and $\mathbf{r}' = \mathbf{r}_Z - \mathbf{r}_Y$ we see that this is equivalent to

$$\mathbf{v}(\mathbf{r} + \mathbf{r}') = \mathbf{v}(\mathbf{r}) + \mathbf{v}(\mathbf{r}'). \quad (1.10)$$

The only vector field with this property is one for which the Cartesian coordinates v_i of \mathbf{v} are linear in the Cartesian coordinates r_i of \mathbf{r} :

$$v_i = \sum_{j=1}^3 H_{ij} r_j \quad (i = 1, 2, 3) \quad (1.11)$$

for constants H_{ij} , which may be viewed as the entries of a constant matrix H . Any such matrix can be written as a sum $H = S + A$ of a symmetric matrix S , with entries S_{ij} , and an antisymmetric matrix A , with entries

$$A_{ij} = \varepsilon_{ijk} \omega_k \quad (1.12)$$

This shows that the antisymmetric part of H may be removed, if it is present, by a change of reference frame to one in relative rotation with angular velocity $\boldsymbol{\omega}$. Let us call a reference frame for which $A = 0$ a *non-rotating* frame. In any given non-rotating frame we have $H = S$ for some symmetric matrix S . Then, in a rotated (but still non-rotating) frame we will have $H = R^{-1}SR$ for rotation matrix R . Isotropy requires that $R^{-1}SR = S$ for all rotation matrices R , and this implies that S is proportional to the identity matrix, so $H_{ij} = H_0 \delta_{ij}$ for some constant H_0 . Thus, homogeneity and isotropy (in non-rotating reference frames) implies Hubble's law, $\mathbf{v} = H_0 \mathbf{r}$.

1.5. Scale factor of the Universe

If we extrapolate Hubble's law to relativistic velocities we find that $v = c$ when $r = cH_0^{-1}$. Of course the extrapolation fails as v approaches c because we should then use the

relativistic law for the addition of velocities. However, we should then also have to take into account that the universe may have expanded appreciably during the time taken for the light to reach us, in which case the meaning of the ‘distance’ r appearing in Hubble’s law becomes ambiguous. We can circumvent these problems as follows. In an isotropic and homogeneous universe all motion can be attributed to a changing distance scale. Let $\mathbf{r}(t)$ be the displacement vector between two galaxies in such a universe. We may write

$$\mathbf{r}(t) = a(t)\mathbf{x}, \quad (1.13)$$

where \mathbf{x} is a time-independent displacement vector, and $a(t)$ is the **scale factor of the universe**. The distance $|\mathbf{x}|$ is called the *comoving* distance between the two galaxies; it is time-independent by definition. Differentiating (1.13) we find that

$$\mathbf{v}(t) = H(t)\mathbf{r}(t), \quad (1.14)$$

where

$$H(t) = \frac{\dot{a}(t)}{a(t)}. \quad (1.15)$$

Let t_0 be the time *now* (i.e., the time of our current cosmological era). We will always choose $a(t)$ such that

$$a(t_0) = 1, \quad (1.16)$$

so that the comoving distance is the distance *now*. Setting $t = t_0$ in (1.14) we recover Hubble’s law with $H_0 = H(t_0)$.

1.6. Dynamics of expansion

Let us consider the mass units to which the cosmological principle applies to be particles of a uniform fluid of energy density $\varepsilon(t)$ and pressure $P(t)$. We take ε to include the rest-mass energy of the particles, in addition to their kinetic energy (which will be non-zero when $P \neq 0$), and we write

$$\varepsilon(t) = \rho(t)c^2, \quad (1.17)$$

where $\rho(t)$ is a mass density. Now consider a small fluid element of fixed *comoving* volume V_0 . Then

$$V(t) = a^3(t)V_0 \quad (1.18)$$

is the actual volume of the fluid element at time t , from which it follows that

$$\dot{V}/V = 3H. \quad (1.19)$$

The energy of the fluid element is $E(t) = \rho(t)c^2V(t)$ and hence

$$\dot{\rho} = V^{-1} \left(\dot{E}/c^2 - \rho\dot{V} \right). \quad (1.20)$$

As the universe expands the fluid element also expands. As it does so its kinetic energy will decrease according to the formula $dE = -PdV$, so

$$\dot{E} = -P\dot{V}. \quad (1.21)$$

Using this in (1.20) we deduce that

$$\dot{\rho} = -(\rho + P/c^2) (\dot{V}/V) \quad (1.22)$$

and hence, from (1.19) that

$$\dot{\rho} = -3(\rho + P/c^2) H. \quad (1.23)$$

We shall call this the **Fluid Equation**.

Now consider a spherical fluid element of fixed co-moving radius R_0 , and hence actual radius $R(t) = a(t)R_0$, and place a test particle of unit mass at its boundary.

This particle will have a velocity of magnitude \dot{R} directed radially outwards; its kinetic energy is therefore

$$K.E. = \frac{1}{2} \dot{R}^2. \quad (1.24)$$

By Birkhoff's theorem, the gravitational force on the test particle is entirely due to mass contained *within* the fluid element; the fluid outside exerts no net force so any contribution from it to the gravitational potential energy must be an irrelevant constant. By Newton's theorem we may assume that the entire mass of the fluid element is concentrated at the centre. This mass is

$$M(t) = \frac{4\pi R^3(t)}{3} \rho(t). \quad (1.25)$$

The gravitational potential energy of the test particle is therefore

$$P.E. = -\frac{GM}{R} = -\frac{4\pi G}{3} \rho R^2. \quad (1.26)$$

Conservation of the total energy ($K.E. + P.E.$) of the test particle requires that

$$\dot{R}^2 - \frac{8\pi G}{3} \rho R^2 = \text{const.} \quad (1.27)$$

But $R(t) = a(t)R_0$, so this is equivalent to

$$\dot{a}^2 - \frac{8\pi G}{3} \rho a^2 = -kc^2, \quad (1.28)$$

for some constant k with dimensions of inverse length squared. This is the **Friedmann Equation**.

Both the Fluid equation and the Friedmann equation were derived by A. Friedmann in 1924 from Einstein's equations of GR. Given that Newton's theory is only an approximation to Einstein's theory, valid for weak gravitational fields, how is it that the equation we have derived from Newtonian theory is *exactly* the same as the one Friedmann derived from GR? The explanation for this is two-fold:

- In using the ‘mass’ M of (1.25) in Newton’s law of gravity we have actually incorporated one aspect of GR, deriving from its dependence on SR: the ‘mass’ in GR is not the rest-mass (which is what Newton himself would presumably have used had he done this calculation) but the energy divided by c^2 , where the energy includes not only the rest-mass energy but also the kinetic energy.
- GR corrections are small if the dimensionless ratio GM/c^2R is small where M and R are typical mass and length scales in the problem. In our case we can take M to be the (total) mass of the fluid element and R its radius. Using (1.25) we see that

$$\frac{GM}{c^2R} \sim \frac{GR_0^2}{c^2} a^2 \rho. \quad (1.29)$$

This is small for sufficiently small R_0 and goes to zero as $R_0 \rightarrow 0$. The GR corrections can therefore be made arbitrarily small by choice of R_0 . But our derivation of the Friedmann equation did not depend on the choice of R_0 , so the GR corrections must vanish. **Because of homogeneity, the behaviour of an arbitrarily small fluid element determines the behaviour of the entire fluid, and Newtonian theory is exact for an arbitrarily small fluid element.** So Newtonian theory suffices for a study of homogeneous cosmological models—a fact on which the viability of this course depends!

1.7. The Big Bang

Differentiate the Friedmann equation to get the **Acceleration Equation** (related to the Raychaudhuri equation in GR)

$$\ddot{a} = -\frac{4\pi G}{3}a(\rho + 3P/c^2). \quad (1.30)$$

[This assumes that $\dot{a} \neq 0$ but (1.30) is a consequence of GR, so the $\dot{a} = 0$ possibility is spurious]. It follows that the expansion is *decelerating* provided that

$$\rho + 3P/c^2 \geq 0. \quad (1.31)$$

This condition is satisfied by all *known* forms of matter, but could be violated by some unknown ‘exotic’ matter (it is not required by general principles: see Q.I.4 and Q.I.5). If we assume this condition and extrapolate back in time we see that the acceleration equation implies that the universe had a beginning (at $a = 0$) at some **finite** time in the past: the **Big Bang**. In fact, **the age of the universe must be less than the Hubble time**

Fig. I.8

By using $E = \rho a^3 V_0 c^2$ and $V = a^3 V_0$ in $\dot{E} = -P\dot{V}$ we deduce that

$$\frac{d(\rho a^3)}{da} = -\frac{3Pa^2}{c^2} \leq 0 \quad (\text{for } P \geq 0) \quad (1.32)$$

Typically, $P \geq 0$ is required for stability of matter (although negative P is not excluded by general physical principles). Assuming $P \geq 0$, it then follows that

$$d(\rho a^3) \geq 0 \text{ if } da < 0, \quad d(\rho a^3) \leq 0 \text{ if } da > 0. \quad (1.33)$$

The first of these inequalities shows that as a decreases ρa^3 must increase, so ρa^2 must increase at least as fast as $1/a$. If we write the Friedmann equation as

$$\dot{a}^2 = \frac{8\pi G}{3} \rho a^2 - kc^2, \quad (1.34)$$

then we see that the ρa^2 must dominate the constant kc^2 term as $a \rightarrow 0$; equivalently, as $t \rightarrow 0$. Thus

$$\dot{a}^2 \sim \frac{8\pi G}{3} \rho a^2 \rightarrow \infty \quad (\text{as } t \rightarrow 0). \quad (1.35)$$

What about the future? We know that $\dot{a} > 0$ *now*; it can change sign only if $\dot{a} = 0$ at some time in the future, but is this possible? From the second of the inequalities (1.33) it follows that that ρa^2 **must decrease at least as fast as $1/a$ as a increases**, and then from (1.34) that \dot{a} must also decrease as a increases. Whether it decreases to zero depends on the value of k . There are three cases to consider:

- $k < 0$. RHS of (1.34) > 0 always, so \dot{a} never vanishes; the universe must expand forever. Since $\rho a^2 \rightarrow 0$ as $a \rightarrow \infty$ we have $\dot{a} \sim \text{constant}$ and hence

$$a(t) \sim \text{const.} \times t \quad (\text{as } t \rightarrow \infty). \quad (1.36)$$

- $k = 0$. Similar to $k < 0$ but slower late time expansion. This case is analogous to the case of escape velocity for a projectile in a potential well.
- $k > 0$. Since ρa^2 is decreasing (as long as \dot{a} is positive) there must come a time when the RHS vanishes, at which time $\dot{a} = 0$. Since $\ddot{a} < 0$ we must have $\dot{a} < 0$ at later times: the universe is *contracting*; the quantity ρa^2 now increases, so \dot{a} remains negative thereafter, and the universe must collapse to $a = 0$ (the *big crunch*).

Fig. I.9

1.8. Flatness problem

Recall that k has dimensions of inverse length squared. This is the same as the dimension of the intrinsic curvature of a two-dimensional space (the inverse of the product of the two radii of curvature). In fact, in GR k is a measure of the curvature of certain two-dimensional slices of the three-dimensional space. If $k < 0$ the curvature is negative (like a saddle) and if $k > 0$ it is positive (like the surface of a sphere). If $k = 0$ the curvature vanishes and the universe is said to be *flat*.

The Friedmann equation can be rewritten as (recall that $H = \dot{a}/a$)

$$\frac{8\pi G}{3H^2} \rho = 1 + \frac{kc^2}{\dot{a}^2}. \quad (1.37)$$

Let us define the (time-dependent) *critical density* ρ_c by

$$\rho_c = \frac{3H^2}{8\pi G}, \quad (1.38)$$

and the *overdensity* Ω by

$$\Omega = \frac{\rho}{\rho_c}. \quad (1.39)$$

Then (1.37) becomes

$$\Omega(t) = 1 + \frac{kc^2}{\dot{a}^2}. \quad (1.40)$$

The quantity $\Omega_0 \equiv \Omega(t_0)$ is known as the *density parameter*. If $k = 0$ then $\Omega \equiv 1$, so $\Omega_0 = 1$.

The critical density *now* is

$$\rho_c(t_0) \sim 10^{-26} \text{ kg m}^{-3} \sim 10^{11} \frac{M_\odot}{(1 \text{ Mpc})^3} \quad (1.41)$$

where M_\odot is the solar mass. But $10^{11} M_\odot$ is the typical mass of a galaxy and 1 Mpc is the typical intergalactic separation, so $\rho(t_0) \approx \rho_c(t_0)$, or

$$\Omega_0 \approx 1, \quad (1.42)$$

This is a very rough estimate, which could be off by a factor of 5 or so, but it implies that the universe is now approximately flat. This is surprising, for a reason that will now be explained.

Because of the deceleration, \dot{a} is a decreasing function of time, so that c^2/\dot{a}^2 is an increasing function of time. Unless $k = 0$, it then follows from (1.40) that the approximation $\Omega \approx 1$ gets worse as time progresses; it is *unstable*. So a universe that is approximately flat

now must have been even flatter in the past. Suppose that we choose ‘initial’ conditions at some early time t_{in} ; then

$$\Omega_{in} - 1 = \frac{kc^2}{\dot{a}^2(t_{in})} = \left(\frac{\dot{a}(t_0)}{\dot{a}(t_{in})} \right)^2 (\Omega_0 - 1) . \quad (1.43)$$

Because $\dot{a} \rightarrow \infty$ as $t_{in} \rightarrow 0$, one has

$$(\Omega_0 - 1) \gg (\Omega_{in} - 1) . \quad (1.44)$$

It follows that to get $\Omega_0 \approx 1$ we must choose $\Omega_{in} \approx 1$ to a far greater precision, such that $\Omega_{in} = 1$ in the limit $t_{in} \rightarrow 0$. This is a very special initial condition and special initial conditions cry out for an explanation. The required special initial conditions are such that **the early universe was very nearly flat**, so this is known as the **flatness problem** of cosmology.

1.9. Equation of State and Cosmological Models

A homogeneous and isotropic cosmological model is determined by the three functions of time a , ρ and P , but so far we have two equations (Fluid and Friedmann) for these three unknowns. We need one more equation. This comes from a specification of the type of matter in the universe. We can provide this specification through an *Equation of State* of the form

$$P = P(\rho) . \quad (1.45)$$

That is, we specify the pressure as a function of the density. Clearly we should have $P(0) = 0$. For cosmological purposes it is usually sufficient to consider a linear equation of state:

$$P = \sigma \rho c^2 \quad (1.46)$$

for some constant σ . The speed of sound in such a material is $\sqrt{|P'|}$. As this should not exceed the speed of light (for causality) we have

$$|\sigma| \leq 1 . \quad (1.47)$$

As mentioned before, normal matter has $P \geq 0$, so that σ is both non-negative and less than unity for normal matter. Two important cases are

- **Dust:** $P = 0$ ($\sigma = 0$)
- **Radiation:** $P = \frac{1}{3}\rho c^2$ ($\sigma = 1/3$)

1.9.1. Einstein-de Sitter universe

For non-relativistic matter, K.E. \ll rest-mass energy, so $P = 0$ is a good approximation. We call this the *matter-dominated* case.

When $P = 0$ the Fluid equation becomes $\dot{\rho} = -3\rho H$, which implies that

$$\frac{4\pi\rho a^3}{3} = M = \text{constant} \quad (1.48)$$

and the Friedmann equation is

$$\dot{a}^2 = \frac{2MG}{a} - kc^2. \quad (1.49)$$

The choice $k = 0$ yields the *Einstein-de Sitter* (EdS) universe

$$a(t) = \left(\frac{t}{t_0}\right)^{\frac{2}{3}}, \quad (1.50)$$

from which we deduce that $H(t) = 2/3t$, and hence $t_0 = \frac{2}{3}H_0^{-1}$. The age of an EdS universe is two-thirds of the Hubble time. The EdS universe is a good approximation to the Universe as we see it now.

1.9.2. Tolman universe

For a radiation dominated universe we have $P = \frac{1}{3}\rho c^2$. The Fluid equation becomes $\dot{\rho} = -4\rho H$, which implies that ρa^4 is constant. Thus,

$$\rho(t) = \frac{\rho_0}{a^4(t)} \quad (1.51)$$

where ρ_0 is the mass density now. The Friedmann equation then becomes

$$\dot{a}^2 - \frac{8\pi G\rho_0}{3a^2} = -kc^2 \quad (1.52)$$

The choice $k = 0$ yields the *Tolman* universe

$$a(t) = \left(\frac{t}{t_0}\right)^{\frac{1}{2}}, \quad (1.53)$$

from which we deduce that $H = 1/2t$ and hence $t_0 = \frac{1}{2}H_0^{-1}$. As the universe is almost certainly older than half the Hubble time (radiocarbon dating gives the age of Earth as about 5×10^9 years, which is already half the Hubble time for $h = 1$) we can deduce that it has not been radiation dominated for most of its history. However, since $\rho_{mat} \sim 1/a^3$ and $\rho_{rad} \sim 1/a^4$ we have

$$\frac{\rho_{rad}}{\rho_{mat}} \sim \frac{1}{a} \rightarrow \infty \text{ as } t \rightarrow 0, \quad (1.54)$$

so if ρ has any radiation component now, this component must have dominated at some time in the early universe. We will return to this point later.

1.10. Redshift formula

Two points separated by a co-moving distance dx will be separated at time t by an actual distance of $a(t)dx$. The time taken for a light signal to go from one point to the other is therefore $dt = a(t)dx/c$, or

$$dx = \frac{cdt}{a(t)}. \quad (1.55)$$

For two points separated by a finite comoving distance x (which we may think of as a distant galaxy and a telescope on earth) we have

$$x = c \int_{t_e}^{t_o} \frac{dt}{a(t)}, \quad (1.56)$$

where t_e is the time of emission of the light signal from point 1 and t_o is the time that this signal is observed at point 2. The times of emission and observation can be chosen as the times at which a particular wavecrest of the light signal passes each of the two points Fig. I.10

But they might equally well be chosen as the times at which the following wavecrest, one wavelength away, passes these points. In this case

$$\begin{aligned} t_e &\rightarrow t_e + \lambda_e/c \\ t_o &\rightarrow t_o + \lambda_o/c, \end{aligned} \quad (1.57)$$

where λ_e and λ_o are the wavelength at emission from galaxy 1 and the wavelength observed at galaxy 2, respectively. So we also have

$$\begin{aligned} x &= c \int_{t_e + \lambda_e/c}^{t_o + \lambda_o/c} \frac{dt}{a(t)} \\ &\approx c \int_{t_e}^{t_o} \frac{dt}{a(t)} + \left[\frac{\lambda_o}{a(t_o)} - \frac{\lambda_e}{a(t_e)} \right] \end{aligned} \quad (1.58)$$

where the approximation is due to the fact that a one wavelength displacement is not infinitesimal—we could have chosen a displacement by a fraction of a wavelength, in principle, so the error could be made as small as we wish. It follows from a comparison of

(1.56) with (1.58) that

$$\frac{\lambda_o}{a(t_o)} = \frac{\lambda_e}{a(t_e)}, \quad (1.59)$$

or (recalling the definition of the redshift z)

$$\frac{a(t_o)}{a(t_e)} = \frac{\lambda_o}{\lambda_e} \equiv 1 + z. \quad (1.60)$$

Thus, the redshift of a distant galaxy is a measure of how much the universe has expanded during the time it takes light to reach us (for whom $t_o = t_0$). In reality this is true only of very distant galaxies for which the assumption of fixed comoving distance is a good approximation; the ‘peculiar’ velocities of galaxies cannot be ignored if they are too close to us.

For the EdS universe we have $a(t) = (t/t_0)^{2/3}$ so

$$1 + z = \left(\frac{t_o}{t_e} \right)^{\frac{2}{3}}. \quad (1.61)$$

Note that $z \rightarrow \infty$ as $t_e \rightarrow 0$. If the EdS universe were valid all the way back to $t = 0$ we could look arbitrarily far back into the past by observing light from objects at arbitrarily large z .

1.11. The cosmological horizon

The earliest time at which a light signal could, in principle, have been emitted is $t = 0$. Setting $t_e = 0$ and $t_0 = t$ in (1.56) we see that the maximum co-moving distance that a light signal could have travelled at a time t after the big bang is

$$x_{max}(t) = c \int_0^t \frac{dt'}{a(t')}. \quad (1.62)$$

This corresponds to an actual distance of

$$R(t) = ca(t) \int_0^t \frac{dt'}{a(t')}, \quad (1.63)$$

which is the *radius of the observable universe* at time t . This radius may be finite or infinite, depending on the function $a(t)$, but for cosmological models in which the universe is decelerating it is always finite. For example,

$$a(t) \propto t^\alpha \quad (0 < \alpha < 1) \quad \Rightarrow \quad R(t) = \frac{ct}{1 - \alpha}. \quad (1.64)$$

Since $0 < \alpha < 1$ for a decelerating universe we deduce that R is finite, although it grows linearly with t . For the EdS universe we have $\alpha = 2/3$. Setting $t = t_0$ we have

$R_0 \equiv R(t_0) = 3ct_0$. The light from any galaxy that is now further away from us than R_0 cannot yet have reached us, so the sphere of radius R_0 centred on us is said to be our *cosmological horizon*.

This resolves Olbers' paradox because as we look out in space we also look back in time, and even if one were to suppose that all the stars of an infinite universe came into existence at $t = 0$ there still would not have been time for the light from more than a finite number of stars (those within our cosmological horizon) to have reached us.

1.12. The horizon problem

Note that $R(t)$ increases with time, with $R(t) \rightarrow \infty$ as $t \rightarrow \infty$ for the EdS universe, so this universe is infinite even though only a finite part of it is now observable to us. But not only does $R(t)$ increase with time, it also increases faster than $a(t)$. This means that the ratio R/a *decreases as we go back in time*, in fact $R/a \rightarrow 0$ as $t \rightarrow 0$. This means that the *comoving* distance to the cosmological horizon decreases to zero as we go back in time. Recalling that galaxies are assumed to have *fixed* comoving positions (since the assumption of homogeneity and isotropy implies that their motion is entirely due to the time-dependence of the scale factor) one sees that the further back in time we go the fewer galaxies there are within the cosmological horizon of any one of them. At some early time each galaxy (or nascent galaxy if galaxies have not yet formed) must have been causally disconnected from any other one. Under these circumstances distant regions of the Universe, which we now see as they were at much earlier epochs, could not have 'colluded' to arrange for the universe to appear isotropic and homogeneous. So why *is* the Universe isotropic and homogeneous? This is the **horizon problem** of cosmology.

This problem is most acute in the context of the Cosmic Microwave Background Radiation (CMBR) that we will discuss later because this radiation arrives isotropically to one part in 10^5 from the most distant regions that can be probed, in principle, by instruments that detect electromagnetic signals. On the conventional Big Bang theory, the isotropy of the CMBR is an incredible coincidence.

1.13. The Cosmological Constant

Consider the 'exotic' equation of state $P = -\rho c^2$ ($\sigma = -1$). In this case the Fluid equation (1.23) implies that $\dot{\rho} = 0$, so we can write

$$\rho = \left(\frac{c^2}{8\pi G} \right) \Lambda \quad (1.65)$$

for *Cosmological Constant* Λ , with units of inverse length squared. The acceleration equation is now

$$\ddot{a} = \frac{c^2 \Lambda}{3} a \quad (1.66)$$

which shows that we have an *accelerating* universe (given that $\Lambda > 0$). The general solution is

$$a(t) = A_+ e^{\sqrt{\Lambda/3} ct} + A_- e^{-\sqrt{\Lambda/3} ct} \quad (1.67)$$

for constants A_+ and A_- to be determined by initial conditions. The Friedman equation now reduces to

$$A_+ A_- = \frac{3k}{4\Lambda} \quad (1.68)$$

For $k = 0$ we have the exponentially expanding *de Sitter* universe with

$$a(t) \propto e^{\sqrt{\Lambda/3} ct} \quad (1.69)$$

The cosmological constant was introduced by Einstein in 1917. He had discovered that his 1915 theory of GR implied a dynamic universe but, thinking that this could not be possible, he modified his equations to include a constant ‘cosmological’ component to the mass density. From the Friedmann equation we see that a static universe requires $k > 0$, and from the acceleration equation we see that it also requires the equation of state

$$P = -\frac{1}{3}\rho. \quad (1.70)$$

Einstein supposed that the pressure was entirely due to a constant ‘cosmological’ component of the mass density $\rho_\Lambda \propto \Lambda$ with $P = -\rho_\Lambda$ and he accounted for the visible matter by a component ρ_{mat} with zero pressure. The equation of state is then that of (1.70) if $\rho_{mat} = 2\rho_\Lambda$. This yields the ($k > 0$) *Einstein Static Universe*, although Einstein’s procedure was different because Friedmann’s simplifications still lay in the future. Later in 1917 de Sitter found his ‘empty space’ cosmological solution to Einstein’s modified equations, but because he found it in a different form he interpreted it as another static universe. Friedmann was the first to take seriously the implications of GR for a dynamic universe, but his general analysis published in 1924 was ignored and he died in 1925. In 1927 Lemaitre proposed a ‘big bang’ model with $\Lambda > 0$ (although the term ‘big bang’ was introduced, sarcastically, by Hoyle in the 1960s) but it was Hubble’s 1929 results and the subsequent re-evaluation of Friedmann’s work that convinced Einstein and everyone else that the Universe is indeed expanding; Einstein then came to consider his introduction of the cosmological constant as his ‘greatest blunder’.

However, once the genie is out of the bottle, it’s not so easy to get him back in again: modern theories of elementary particles typically predict the dimensionless ratio $c^3\Lambda/G\hbar$ to be of order unity, but astronomical observations imply that

$$\frac{c^3\Lambda}{G\hbar} \sim 10^{-120}. \quad (1.71)$$

Until recently, observations were consistent with $\Lambda = 0$ but it now seems that the cosmological constant is non-zero. Just why it is so small in natural units is a mystery, known as the **Cosmological Constant Problem**.

1.14. The inflationary universe *

Notice that the general solution (1.67) of the acceleration equation with equation of state $P = -\rho c^2$, corresponding to arbitrary initial conditions, approaches the $k = 0$ de Sitter universe within a time $\sim 1/c\sqrt{\Lambda}$. The initial choice of k is soon irrelevant because the exponential expansion rapidly ‘flattens’ the universe. If the dS universe were viable then there would be no ‘flatness problem’. Although the dS universe is not viable now, one can suppose that our universe might have gone through a dS phase very early in its history. If this phase were sufficiently long then any region within what is now our cosmological horizon could have been flattened by the expansion, thus accounting for the special initial conditions in a subsequent big bang phase. This is called the *inflationary universe hypothesis*; one supposes that the Big Bang was actually the transition to a hot Tolman universe (with $\Lambda = 0$) from an earlier de Sitter phase created by the rapid expansion due to an approximately constant ‘cosmological’ contribution to the energy density. After the transition, the universe expands and cools until it enters an EdS phase:

Fig. I.11

This inflationary universe hypothesis solves the horizon problem, as well as the flatness problem, because the dS universe has no cosmological horizon [see Q.1.5]. It also solves several other problems related to particle physics in the early universe. Most importantly, it makes predictions for density fluctuations of the CMBR that appear to be confirmed by recent observations. However, it does not solve the cosmological constant problem because it is just assumed that the final value of Λ is small.

2. Statistical Mechanics

2.1. Entropy and the laws of Thermodynamics

Most macroscopic systems (composed of a large number of particles of some kind) are adequately described by a few variables, e.g. energy E , volume V and number of particles N . For each choice of E, V, N there is a huge number Ω of possible ‘microstates’ (which we need not know about in any detail for the moment). We define the *entropy* of the system, S , by Boltzmann’s formula

$$S = k \log \Omega . \tag{2.1}$$

The constant k is known as Boltzmann’s constant; it has dimensions of energy/temperature. Temperature is really just a measure of kinetic energy so the natural dimensions of temperature are those of energy, in which case Boltzmann’s constant would be a dimensionless number that we might as well choose to be 1; however, the concept of temperature had already been introduced, with its own units, so Boltzmann had to introduce his constant to convert energy units to temperature units. In other words, we need to introduce k for historical reasons but the whole theory could be developed without it, so it has no fundamental significance.

For an *isolated* system E, V, N are usually fixed (for simplicity, we will take this as the definition of what we mean by ‘isolated’). Suppose that we have two such systems. The total number of microstates of the combined system is then

$$\Omega = \Omega_1 \Omega_2 \tag{2.2}$$

because for every choice of a microstate in system 1 we have a possible Ω_2 microstates of system 2. It then follows that

$$S = S_1 + S_2 , \tag{2.3}$$

i.e., entropy is additive.

Fig. II.1

Suppose that we now allow the two systems to exchange energy at fixed V, N (via heat exchange), volume at fixed N, E (via a moveable partition) and particle number at fixed E, V (by allowing the partition to be permeable). The possible partitions of the the total energy E , total volume V and total number of particles N , can be specified by, say, the energy E_1 , volume V_1 and number of particles N_1 of system 1 since the corresponding variables of system 2 are then fixed at $E_2 = E - E_1$, $V_2 = V - V_1$ and $N_2 = N - N_1$. Thus

$$S = S_1(E_1, V_1, N_1) + S_2(N - N_1, V - V_1, N - N_1) \quad (2.4)$$

and the total entropy may now be considered a function of the ‘partition variables’ (E_1, V_1, N_1) . Experience shows that the system will come to equilibrium with some definite values for these variables. How do we find these equilibrium values? We shall suppose that the equilibrium values are the most probable values, which makes sense if fluctuations are negligible. We then make use of the following

Hypothesis: *Each microstate of an isolated system is equally likely*

This could be interpreted as a partial definition of what we mean by a microstate but, ultimately, microstates can only be understood in terms of the quantum mechanics of the particles composing the system.

Given this hypothesis, the probability of any particular partition of our combined system is proportional to the number of microstates Ω consistent with those particular values of (E_1, V_1, N_1) . The most probable partition is therefore the one that maximizes Ω , and hence S with respect to variations in (E_1, V_1, N_1) . This requires

$$\frac{\partial S}{\partial E_1} = 0, \quad \frac{\partial S}{\partial V_1} = 0, \quad \frac{\partial S}{\partial N_1} = 0 \quad (2.5)$$

where $S(E_1, V_1, N_1)$ is the function given in (2.4). Since (for $x = x_1 + x_2$)

$$\frac{\partial f(x - x_1)}{\partial x_1} = -f' \equiv -\frac{\partial f(x_2)}{\partial x_2}, \quad (2.6)$$

these equations are equivalent to

$$\begin{aligned} \partial S_1(E_1, V_1, N_1)/\partial E_1 &= \partial S_2(E_2, V_2, N_2)/\partial E_2 \\ \partial S_1(E_1, V_1, N_1)/\partial V_1 &= \partial S_2(E_2, V_2, N_2)/\partial V_2 \\ \partial S_1(E_1, V_1, N_1)/\partial N_1 &= \partial S_2(E_2, V_2, N_2)/\partial N_2 \end{aligned} \quad (2.7)$$

Let us define the *temperature* T , the *pressure* P and the *chemical potential*, μ , of any subsystem with entropy $S(E, V, N)$ by

$$\frac{\partial S}{\partial E} = \frac{1}{T}, \quad \frac{\partial S}{\partial V} = \frac{P}{T}, \quad \frac{\partial S}{\partial N} = -\frac{\mu}{T}. \quad (2.8)$$

Then the conditions for equilibrium (2.7) can be written as¹

$$T_1 = T_2, \quad P_1 = P_2, \quad \mu_1 = \mu_2. \quad (2.9)$$

The first of these conditions is called the **zeroth law of thermodynamics**: all parts of an isolated system in thermal equilibrium are at the same temperature

An immediate consequence of (2.8) is that

$$dS = \frac{1}{T}dE + \frac{P}{T}dV - \frac{\mu}{T}dN, \quad (2.10)$$

which is equivalent to

$$dE = TdS - PdV + \mu dN. \quad (2.11)$$

This is known as the **first law of thermodynamics**. The temperature T appearing in this law is *absolute* temperature, such that $T = 0$ is the *absolute zero*; classically this corresponds to zero kinetic energy (quantum mechanically there is always a zero-point energy, even at absolute zero). To complete the definition of T we must also choose a temperature scale. The choice is arbitrary, but it is customary to measure T in degrees Kelvin (K); this is essentially the Celsius scale (centigrade) but with zero temperature moved to absolute zero.

The first law is a mathematical expression of the interchangeability of three forms of energy: heat, work, and chemical energy. For simplicity let's assume that $dN = 0$, which allows us to ignore chemical energy. The first law can then be written as

$$dE = dQ + dW, \quad (2.12)$$

where we have defined

$$dQ = TdS, \quad dW = -PdV \quad (2.13)$$

as the heat transferred to the system, and work done on it, respectively. Thus, (2.12) says that the total change in energy of the system equals the sum of the heat absorbed and the work done on it. This might seem obvious enough but the subtlety is that one cannot consistently assign to the system any definite amounts of quantities which one might wish to write as Q and W and call 'heat' and 'work' (as the old 'caloric' theory of heat tried to do), whereas the system does have a definite amount of energy, E . This is part of the significance of the first law; mathematically, it is the statement that whereas neither dQ nor dW are exact differentials their sum dE is an exact differential. But the first law further states that although dQ is not an exact differential, it does have an *integrating factor*, namely $1/T$, so that $dQ/T = dS$ is an exact differential. It follows that there also

¹These conditions need modification in the presence of gravity because one must then take into account the gravitational potential energy, but we pass over this point here.

exists a quantity S of which any system in thermal equilibrium has a definite amount. This deduction was originally made by Clausius, and he called S the ‘entropy’. [By a similar argument you can deduce that a system in mechanical equilibrium has a definite ‘amount’ of something called V , but this is hardly a surprise.]

Clausius, and independently Lord Kelvin, showed that the quantity that Clausius had called entropy could only increase in any change undergone by an isolated system; this is the **second law of thermodynamics**. From our perspective, i.e. Boltzmann’s, the second law is obvious because we found the conditions for equilibrium by maximizing the entropy.

2.2. Chemical potential

For a gas composed of several different species of particle, the total pressure is the sum of the ‘partial pressures’ exerted by each species, so if a gas is composed of particles of type A and B its total pressure is $P = P_A + P_B$, where P_A and P_B are the partial pressures. If this gas is separated by a moveable partition from another gas composed of particles of type C then the equilibrium position of the partition occurs when

$$P_A + P_B = P_C . \quad (2.14)$$

Suppose now that we have a gas composed of particles of all three types, with no partition but such that the chemical reaction

$$A + B \leftrightarrow C \quad (2.15)$$

can convert a particle of type A and a particle of type B into a particle of type C , and vice versa. Since $N_A + N_B + N_C \equiv N$ and $N_A - N_B$ are unchanged by this reaction, we may consider N_C as the only variable. Its value in equilibrium is determined by the condition analogous to (2.14)

$$\mu_A + \mu_B = \mu_C . \quad (2.16)$$

If this system is initially out of equilibrium, with $\mu_A + \mu_B > \mu_C$ then the reaction (2.15) will proceed in the forward direction, creating more of C at the expense of A and B . If it is initially out of equilibrium, with $\mu_A + \mu_B < \mu_C$ then the reaction will proceed in the other direction. Thus, chemical potentials determine the direction of chemical reactions.

For relativistic systems the energy E should include the rest-mass energy. For a system of N particles of mass m we therefore write

$$E = Nmc^2 + U \quad (2.17)$$

where U is the ‘internal energy’, in the terminology of thermodynamics. The first law of thermodynamics can now be written as

$$dU = TdS - PdV + \mu_{NR}dN \quad (2.18)$$

where

$$\mu_{NR} \equiv \mu - mc^2 \quad (2.19)$$

is the Non-Relativistic chemical potential; it is this quantity that is often called ‘the chemical potential’ in books on thermodynamics because this subject is usually developed in the context of non-relativistic systems.

2.3. Thermodynamic Ensembles

So far we have considered an isolated system with equilibrium states that are determined by the three *extensive* variables E, V, N , so called because doubling the size of the system doubles their values. The set of all microstates consistent with specified values of E, V, N is called the **microcanonical ensemble**. The entropy $S(E, V, N)$, considered as a function of the extensive variables determines the remaining *intensive* variables (T, P, μ) (which are independent of the size of the system). The intensive variables fluctuate about their equilibrium values but these fluctuations are negligible for macroscopic systems.

We are often interested in systems that are *not* isolated; e.g., at fixed temperature due to thermal contact with an infinite heat bath. We then specify (T, V, N) . The collection of all microstates compatible with a given (T, V, N) is called the **canonical ensemble** (this is how statistical mechanics was initially formulated). Define the *Helmholtz Free Energy* F by

$$F = E - TS. \quad (2.20)$$

The first law can now be rewritten as

$$dF = -SdT - PdV + \mu dN \quad (2.21)$$

Integrate to get $F(T, V, N)$ with (see Q.II.3)

$$S = -\frac{\partial F}{\partial T}, \quad P = -\frac{\partial F}{\partial V}, \quad \mu = \frac{\partial F}{\partial N}. \quad (2.22)$$

It can be shown that F is a minimum in equilibrium, so equilibrium is now found by minimizing F . This determines the remaining variables (S, P, μ) —these fluctuate about their equilibrium value, but the fluctuations are negligible for macroscopic systems.

In many relativistic systems the particle number is not conserved so we cannot fix N . In this case we can fix μ instead. We then specify (T, V, μ) . The collection of all microstates compatible with a given (T, V, μ) is called the **Grand canonical ensemble**. We shall take this approach to statistical mechanics²

²It would also be possible to specify only the intensive variables (T, P, μ) , and this is standard in chemistry because the state of reagents in equilibrium in an open beaker is one in which both temperature and pressure are determined by the environment, but this is not convenient for statistical mechanics because the microstates of a system are its energy eigenstates and the energy of a microstate typically depends on the volume of the system.

2.4. The Gibbs Distribution

We shall suppose the system of interest, which we will call ‘system 1’, to be in thermal, mechanical, and chemical contact with a much larger system, which we will call ‘system 2’ (‘chemical’ denotes the possibility of an exchange of particles). We will suppose that the combined system is isolated, so equilibrium corresponds to a maximum of the entropy for the combined system comprising subsystems 1 and 2. Since system 2 is much larger than system 1 its extensive variables are also much larger:

$$E_2 \gg E_1, \quad V_2 \gg V_1, \quad N_2 \gg N_1 \equiv n. \quad (2.23)$$

If we now fix V_1 then the microstates of system 1 at fixed temperature and chemical potential are those of a grand canonical ensemble. We take these microstates to be the energy eigenstates available in volume V_1 to n particles, for variable n . For a given n we could label each n -particle energy eigenstate by an integer r , but since n is variable we must have a set of such labels for each n , so let $r^{(n)}$ be the r th n -particle energy eigenstate. Let $E_r^{(n)}$ be the corresponding energy eigenvalues. We may assume that the energy eigenstates are non-degenerate for any given n because any degeneracy could be lifted by an arbitrarily small perturbation; this means that *subsystem 1 is in microstate $r^{(n)}$ if, and only if, it has n particles and energy $E_r^{(n)}$* .

We want to find the probability $p(r^{(n)})$ that subsystem 1 is in a *particular* microstate, with label $r^{(n)}$. This probability is proportional to the number of microstates available to the *total* system given the constraints that $E_1 = E_r^{(n)}$ and $N_1 = n$. **Since these constraints specify a *particular* microstate of system 1 the number of microstates of the total system compatible with them is the number of microstates of system 2;** that is

$$p(r^{(n)}) \propto \Omega_2(E_2, N_2) = \exp\left(\frac{1}{k} S_2(E_2, N_2)\right) \quad (2.24)$$

where $E_2 = E - E_r^{(n)}$ and $N_2 = N - n$. Thus

$$p(r^{(n)}) \propto \exp\left[\frac{1}{k} S_2(E - E_r^{(n)}, N - n)\right]. \quad (2.25)$$

Note that we suppress the dependence on V_2 because $V_2 = V - V_1$ and both V and V_1 are being held fixed.

Now, $E \gg E_1$ and $N \gg N_1$, by hypothesis, so we may Taylor expand S_2 :

$$\begin{aligned} S_2(E - E_r^{(n)}, N - n) &= S_2(E, N) - \left(\frac{\partial S_2}{\partial E_2}\right) E_r^{(n)} - \left(\frac{\partial S_2}{\partial N_2}\right) n + \dots \\ &= \text{constant} - \left(\frac{1}{T_2}\right) E_r^{(n)} + \left(\frac{\mu_2}{T_2}\right) n + \dots \end{aligned} \quad (2.26)$$

But $\mu_2 = \mu_1 = \mu$ and $T_2 = T_1 = T$ in equilibrium, so

$$p(r^{(n)}) \propto e^{\beta(\mu n - E_r^{(n)})} \quad \left(\beta \equiv \frac{1}{kT} \right). \quad (2.27)$$

The sum of these probabilities over all microstates of subsystem 1 must be unity, so

$$p(r^{(n)}) = \frac{e^{\beta(\mu n - E_r^{(n)})}}{\mathcal{Z}} \quad (2.28)$$

where

$$\mathcal{Z} = \sum_{\text{states}} e^{\beta(\mu n - E_r^{(n)})}. \quad (2.29)$$

The probability distribution (2.28) is known as the **Gibbs Distribution**; the normalization factor \mathcal{Z} is known as the *Grand partition function*. The quantity

$$\mathcal{G} = -kT \log \mathcal{Z} \quad (2.30)$$

is the Gibbs *Grand Potential*. It is analogous to the Helmholtz free energy for the canonical ensemble in that, for a grand canonical ensemble, \mathcal{G} is a minimum at equilibrium. One sees from this that $\log \mathcal{Z}$ is analogous to the entropy, so that \mathcal{Z} is a measure of the number of microstates available in a grand canonical ensemble.

2.5. Ideal Gases

An ideal gas is one for which the internal energy U is entirely kinetic, so that $U = U(T)$, independent of the pressure or volume. This will certainly be the case if the particles do not interact (although there must be some interaction, e.g. with the walls of the container, in order for the gas to come to equilibrium). It will also be a good approximation if the particles interact only via a short range force provided that this range is much less than the mean free path of a gas particle. Gravity is long range but weak, so we neglect it here, although gravitational gradients (such as those of the Earth's atmosphere or the atmosphere of a star) will ultimately require some changes.

We will assume that the particles of the gas do not interact. Suppose that we build up to a gas of n particles by putting one particle at a time into a box of volume V . Each particle must go into one of the energy levels allowed to a *single* particle in the box. For a sufficiently asymmetric box we will have a series of non-degenerate energy levels (E_0, E_1, E_2, \dots) with $E_{k+1} > E_k$. An energy eigenstate of a gas of n particles is then specified by the set of *occupation numbers* (n_0, n_1, n_2, \dots) of each level; that is, n_1 particles with energy E_1 , etc. The label $r^{(n)}$ can now be replaced by the set of occupation numbers $\{n\} = (n_0, n_1, n_2, \dots)$.

Fig. II.2

Given $\{n\}$ we have³

$$n = \sum_k n_k, \quad E_r^{(n)} = \sum_k n_k E_k \quad (2.31)$$

We may now use this result to write the grand partition function as

$$\begin{aligned} \mathcal{Z} &= \sum_{\{n\}} e^{\beta \sum_k (\mu - E_k) n_k} \\ &= \sum_{\{n\}} \prod_k [e^{\beta(\mu - E_k)}]^{n_k}. \end{aligned} \quad (2.32)$$

Lemma: For any denumerable set (finite or infinite) of numbers $\{B\}$ we have

$$\sum_{\{n\}} \prod_k (B_k)^{n_k} = \prod_k \left(\sum_n (B_k)^n \right) \quad (2.33)$$

The proof is as follows:

$$\begin{aligned} LHS &= (\dots \sum_{n_2} \sum_{n_1}) [(B_1)^{n_1} (B_2)^{n_2} \dots] \\ &= (\dots \sum_{n_3} \sum_{n_2} \sum_n) [\sum_n (B_1)^n] [(B_2)^{n_2} (B_3)^{n_3} \dots] \\ &= [\sum_n (B_1)^n] (\dots \sum_{n_3} \sum_{n_2}) [(B_2)^{n_2} (B_3)^{n_3} \dots]. \end{aligned} \quad (2.34)$$

Iteration now yields the RHS.

Applying this lemma with $B_k = \exp[\beta(\mu - E_k)]$ we deduce that

$$\mathcal{Z} = \prod_k \mathcal{Z}_k, \quad \mathcal{Z}_k \equiv \sum_n e^{\beta(\mu - E_k)n} \quad (2.35)$$

The Gibbs distribution similarly factorizes:

$$p(\{n\}) = \frac{e^{\beta(\sum_k (\mu - E_k) n_k)}}{\mathcal{Z}} = \prod_k p_k(n_k) \quad (2.36)$$

where

$$p_k(n) = \frac{e^{\beta(\mu - E_k)n}}{\mathcal{Z}_k}. \quad (2.37)$$

³These are *not* constraints on the numbers n_k because neither n nor $E_r^{(n)}$ are fixed. They *are* fixed in the microcanonical ensemble, and in that case one proceeds differently, but the end result is the same—see Q.II.1.

The average value of n_k is therefore

$$\begin{aligned}\bar{n}_k &= \sum_n n p_k(n) = \frac{\sum_n n e^{\beta(\mu - E_k)n}}{\mathcal{Z}_k} \\ &= \frac{1}{\beta \mathcal{Z}_k} \frac{\partial \mathcal{Z}_k}{\partial \mu},\end{aligned}\tag{2.38}$$

so that

$$\bar{n}_k = kT \frac{\partial \log \mathcal{Z}_k}{\partial \mu}.\tag{2.39}$$

2.6. Bosons and Fermions

We assumed that a state of the gas is specified by the set of occupation numbers $\{n\}$. This is equivalent to the assumption that the gas particles are identical, and hence *indistinguishable*. According to QM, identical particles are either *bosons* or *fermions*. For bosons there is no restriction on the occupation numbers, so the allowed values of n_k , for each k , are $(0, 1, 2, \dots, \infty)$. thus

$$\begin{aligned}\mathcal{Z}_k &= \sum_{n=0}^{\infty} [e^{\beta(\mu - E_k)}]^n \\ &= \frac{1}{1 - e^{\beta(\mu - E_k)}} \quad (\text{bosons})\end{aligned}\tag{2.40}$$

For fermions the Pauli Exclusion Principle states that the only allowed values of n_k , for each k , are $(0, 1)$, so

$$\begin{aligned}\mathcal{Z}_k &= \sum_{n=0}^1 [e^{\beta(\mu - E_k)}]^n \\ &= 1 + e^{\beta(\mu - E_k)} \quad (\text{fermions})\end{aligned}\tag{2.41}$$

Using these results in (2.39) we deduce that

$$\bar{n}_k = \frac{1}{e^{\beta(E_k - \mu)} - 1} \quad (\text{bosons})\tag{2.42}$$

$$\bar{n}_k = \frac{1}{e^{\beta(E_k - \mu)} + 1} \quad (\text{fermions})\tag{2.43}$$

These formulae give the the average number of particles in a *single* one-particle energy eigenstate. If the k th energy level has degeneracy g_k then there are g_k eigenstates with energy E_k , so the average number of particles with energy E_k is

$$\bar{n}(E_k) = \frac{g_k}{e^{\beta(E_k - \mu)} - 1} \quad (\text{bosons})\tag{2.44}$$

$$\bar{n}(E_k) = \frac{g_k}{e^{\beta(E_k - \mu)} + 1} \quad (\text{fermions})\tag{2.45}$$

The function $\bar{n}(E)$ for bosons is called the *Bose-Einstein* distribution, and the function $\bar{n}(E)$ for fermions is called the *Fermi-Dirac* distribution. When the ground state energy E_0 is such that

$$e^{\beta(E_0-\mu)} \gg 1 \quad (2.46)$$

then $e^{\beta(E_k-\mu)} \gg 1$ for any k , and hence

$$\bar{n}(E_k) \approx g_k e^{-\beta(E_k-\mu)}, \quad (2.47)$$

irrespective of whether the particles are bosons or fermions; this is called the *Maxwell-Boltzmann* distribution. Note that the validity of the MB distribution requires that $\bar{n}(E_k) \ll g_k$, so the average occupation number of any individual eigenstate is much less than unity—under these circumstances the fact that the particles are identical becomes irrelevant [see Q.II.1]

2.7. Density of states

Consider a particle in a cubic box of side L . Let ψ be a momentum eigenstate with eigenvalue p :

$$-i\hbar\nabla\psi = \mathbf{p}\psi. \quad (2.48)$$

Impose periodic boundary conditions. The eigenfunctions are then

$$\psi_{\mathbf{l}}(\mathbf{r}) = e^{\frac{2\pi i \mathbf{l} \cdot \mathbf{r}}{L}}, \quad \mathbf{l} = (l_1, l_2, l_3) \quad (2.49)$$

and the eigenvalues are

$$\mathbf{p} = \left(\frac{h}{L}\right) \mathbf{l}. \quad (2.50)$$

That is, states form a cubic lattice in momentum space with spacing h/L . The average number of lattice sites in a volume V_p of momentum space is therefore

$$\left(\frac{L}{h}\right)^3 V_p. \quad (2.51)$$

For particles with spin we must multiply this by a *spin-degeneracy* factor g_s , so the average number of momentum eigenstates in a volume V_p of momentum space is

$$\left(\frac{g_s V}{h^3}\right) V_p \quad (2.52)$$

since $L^3 = V$, the volume of the box. We derived this result by assuming a cubic box with periodic boundary conditions but it is true quite generally (excepting ‘pathological’ box shapes). Now consider a shell in momentum space of radius p and depth dp , for

which $V_p = 4\pi p^2 dp$. The average number of momentum eigenstates in this shell is $g(p)dp$, where⁴

$$g(p) = \frac{4\pi g_s V}{h^3} p^2. \quad (2.54)$$

This is the *density of states*, as a function of p . The average number of particles with momentum in the range $(p, p + dp)$ is therefore

$$\bar{n}(p)dp = \frac{g(p)dp}{e^{\beta(E(p)-\mu)} \mp 1} = \frac{4\pi g_s V}{h^3} \frac{p^2}{e^{\beta(E(p)-\mu)} \mp 1} dp, \quad (2.55)$$

where the upper sign is for bosons and the lower sign for fermions, and

$$E(p) = c\sqrt{p^2 + m^2 c^2}. \quad (2.56)$$

The total number of particles is

$$N = \int_0^\infty \bar{n}(p)dp, \quad (2.57)$$

and the total energy is

$$E = \int_0^\infty E(p)\bar{n}(p)dp. \quad (2.58)$$

These are really averages because they fluctuate, but the fluctuations are negligible for a macroscopic system.

Note that N and E were previously called N_1 and E_1 , with N and E being reserved for the fixed total number of particles and fixed energy of a larger isolated system, but we no longer need to refer to this larger system so we drop the suffix 1. Similarly V_1 is now V , although this was kept fixed and so does not fluctuate.

2.8. Ideal Gas Pressure

We have kept V fixed in arriving at the above formula, but there is an implicit dependence on V due to the fact that the momentum p of a given (momentum or energy) eigenstate depends on V ; in fact

$$p \propto V^{-\frac{1}{3}}, \quad (2.59)$$

so that

$$\frac{dp}{dV} = -\frac{p}{3V}. \quad (2.60)$$

⁴To obtain these distributions as functions of the energy E of a gas particle, we set $g(p)dp = \tilde{g}(E)dE = \tilde{g}(E)E'(p)dp$, so

$$\tilde{g}(E) = \frac{4\pi g_s V}{h^3} \frac{p^2}{E'(p)} \quad (2.53)$$

where p is the inverse function to $E(p)$. For our purposes it is easier to work with the density of states as a function of momentum.

Suppose that we change V *slowly*, so that equilibrium is maintained. In equilibrium all reactions that are capable of changing the occupation number of any one-particle eigenstate proceed equally in both directions, so these occupation numbers cannot change. In the approximation of a continuous distribution of momentum eigenvalues this means that $\bar{n}(p)dp$ is unchanged; we then deduce from (2.57) that N is unchanged, as it obviously must be if all occupation numbers remain unchanged. The energy, as given by (2.58) does change however since

$$\begin{aligned} dE &= \int_0^\infty dE(p)\bar{n}(p)dp = dV \int_0^\infty E'(p)(dp/dV)\bar{n}(p)dp \\ &= -\frac{dV}{3V} \int_0^\infty pE'(p)\bar{n}(p)dp. \end{aligned} \quad (2.61)$$

This change of E occurs at constant N and, since equilibrium is maintained, at constant S . We thus deduce that

$$\left(\frac{\partial E}{\partial V}\right)_{N,S} = -\frac{1}{3V} \int_0^\infty pE'(p)\bar{n}(p)dp. \quad (2.62)$$

But the first law of thermodynamics implies that

$$\left(\frac{\partial E}{\partial V}\right)_{N,S} = -P, \quad (2.63)$$

so we deduce the following formula for the pressure of an ideal gas:

$$P = \frac{1}{3V} \int_0^\infty pE'(p)\bar{n}(p)dp. \quad (2.64)$$

Recall that

$$E(p) = c\sqrt{p^2 + m^2c^2}. \quad (2.65)$$

We shall need to consider only the two limiting cases

- **Non-relativistic** (NR). For a gas of NR particles of mass m we have

$$E(p) = mc^2 + U(p), \quad U(p) = \frac{p^2}{2m}. \quad (2.66)$$

In this case $pE'(p) = 2U(p)$, so

$$P = \frac{2}{3V} \int_0^\infty U(p)\bar{n}(p)dp = \frac{2}{3}(U/V), \quad (2.67)$$

where U is the total ‘internal’ energy (energy minus rest-mass energy).

- **Ultra-relativistic (UR).** For a gas of UR particles we have

$$E(p) = pc \quad (2.68)$$

and hence $pE'(p) = E(p)$. It then follows from (2.64) that

$$P = \frac{1}{3V} \int_0^\infty E(p) \bar{n}(p) dp = \frac{1}{3} (E/V), \quad (2.69)$$

where E is the total energy

2.9. Classical NR Gas

For a classical gas we may use the MB distribution, so

$$N = \left(\frac{4\pi g_s V}{h^3} \right) e^{\beta\mu} \int_0^\infty dp p^2 e^{-\beta E(p)}. \quad (2.70)$$

If we further assume that the gas is non-relativistic, then

$$\begin{aligned} N &= \left(\frac{4\pi g_s V}{h^3} \right) e^{\beta(\mu - mc^2)} \int_0^\infty dp p^2 e^{-\beta \frac{p^2}{2m}} \\ &= \left(\frac{4\pi g_s V}{h^3} \right) e^{\beta(\mu - mc^2)} \left[\left(\frac{\sqrt{\pi}}{4} \right) \left(\frac{2m}{\beta} \right)^{\frac{3}{2}} \right]. \end{aligned} \quad (2.71)$$

The number density $n \equiv N/V$ is therefore given by the formula

$$n = g_s e^{\beta(\mu - mc^2)} n_Q \quad (2.72)$$

where

$$n_Q = \left(\frac{2\pi m}{\beta h^2} \right)^{\frac{3}{2}}. \quad (2.73)$$

The quantity n_Q is called the *quantum concentration*. The classical approximation requires that

$$e^{\beta(\mu - mc^2)} \ll 1 \quad (2.74)$$

and hence that $n \ll n_Q$. In other words, the classical approximation fails when the gas density approaches n_Q . The formula (2.72) can be rewritten as a formula for μ :

$$\mu = mc^2 - kT \log \left(\frac{g_s n_Q}{n} \right). \quad (2.75)$$

The internal energy of classical NR gas is given by a formula analogous to that of (2.71) for N but with an insertion of $U(p)$ inside the integral:

$$\begin{aligned}
U &= \left(\frac{4\pi g_s V}{h^3} \right) e^{\beta(\mu - mc^2)} \int_0^\infty dp \left(\frac{p^2}{2m} \right) p^2 e^{-\beta \frac{p^2}{2m}} \\
&= - \left(\frac{4\pi g_s V}{h^3} \right) e^{\beta(\mu - mc^2)} \frac{\partial}{\partial \beta} \left[\left(\frac{\sqrt{\pi}}{4} \right) \left(\frac{2m}{\beta} \right)^{\frac{3}{2}} \right] \\
&= \frac{3}{2\beta} N.
\end{aligned} \tag{2.76}$$

Equivalently,

$$U = \frac{3}{2} N k T. \tag{2.77}$$

If we combine this result with our earlier one that $PV = \frac{2}{3}U$ for a NR gas, we deduce the **Boyle-Charles law**

$$PV = NkT \tag{2.78}$$

Note that a classical NR (monatomic) gas of N particles has $3N$ ‘degrees of freedom’, because each of the N particles is free to move in three dimensions, so the internal energy is $\frac{1}{2}kT$ times the number of degree of freedom. This is an illustration of the general principle of

Equipartition of energy: *the total energy of a weakly interacting system in thermal equilibrium at temperature T is partitioned equally among all its degrees of freedom, with $\frac{1}{2}kT$ per degree of freedom.*

This principle assumes that each degree of freedom is accessible at any temperature, as is true for translational motion. The assumption is not generally true, however, because QM may require a minimum energy, as happens for rotational and vibrational modes of a molecule; such degrees of freedom only become accessible above a certain temperature. At room temperature vibrational modes are inaccessible but rotational modes are important for molecules, so the perfect gas result derived above applies in practice only to monatomic gases.

2.10. Fermi Gas and degeneracy pressure

For a gas of fermions, the average occupation number of a single energy eigenstate is

$$\bar{n}_k = \frac{1}{e^{\beta(E_k - \mu)} + 1}. \tag{2.79}$$

As $T \rightarrow 0$, corresponding to $\beta \rightarrow \infty$, \bar{n} tends to 1 or 0 depending on whether E_k is less than or greater than μ . If we define the *Fermi energy* by

$$\mu|_{T=0} = \epsilon_F \tag{2.80}$$

then at $T = 0$ we have

$$\bar{n}_k \rightarrow \begin{cases} 1 & E_k < \epsilon_F \\ 0 & E_k > \epsilon_F \end{cases} . \quad (2.81)$$

Fig. II.3

At $T > 0$ this distribution is smoothed out so that \bar{n}_k is no longer zero for $E_k > \epsilon_F$ but still falls rapidly to zero as long as $kT \ll \epsilon_F$. A Fermi gas with $kT \ll \epsilon_F$ is said to be *degenerate*.

Fig. II.4

For a classical gas $P = nkT$, so $P \rightarrow 0$ as $T \rightarrow 0$. But the classical approximation fails as $T \rightarrow 0$. For a Fermi gas this is because we must take into account the *degeneracy pressure*. Consider a Fermi gas at $T = 0$. All momentum eigenstates are filled up to the *Fermi momentum* p_F corresponding to the Fermi energy; that is

$$\epsilon_F^2 = p_F^2 c^2 + m^2 c^4, \quad (2.82)$$

which implies

$$\epsilon_F = \begin{cases} mc^2 + \frac{p_F^2}{2m} & \text{NR} \\ p_F c & \text{UR} \end{cases} \quad (2.83)$$

Since all eigenstates with $p > p_F$ are empty at $T = 0$ we have

$$N = \left(\frac{4\pi g_s V}{h^3} \right) \int_0^{p_F} p^2 dp = \frac{4\pi g_s V p_F^3}{3h^3} \quad (2.84)$$

and hence

$$n = \left(\frac{4\pi g_s}{3} \right) \left(\frac{p_F}{h} \right)^3 \sim \frac{1}{\lambda_F^3} \quad (2.85)$$

where

$$\lambda_F = \frac{h}{p_F} \quad (2.86)$$

is the de Broglie wavelength associated with the Fermi momentum. In other words *each particle of a degenerate Fermi gas occupies a volume $\sim \lambda_F^3$* . Let us note for future use that (2.85) is equivalent to

$$\frac{p_F}{h} = \left(\frac{3n}{4\pi g_s} \right)^{\frac{1}{3}}. \quad (2.87)$$

We will now consider the two limits, NR and UR, separately:

- In the NR limit we have

$$\begin{aligned}\frac{U}{V} &= \frac{4\pi g_s}{h^3} \int_0^{p_F} \left(\frac{p^2}{2m} \right) p^2 dp = \left[\frac{4\pi g_s h^2}{5m} \left(\frac{p_F}{h} \right)^5 \right] \\ &= \frac{3h^2}{5m} \left(\frac{3}{4\pi g_s} \right)^{\frac{2}{3}} n^{\frac{5}{3}},\end{aligned}\tag{2.88}$$

where we have used (2.87) in the last line. For a NR gas we have $P = \frac{2}{3}(U/V)$ so

$$P \sim \frac{h^2}{m} n^{\frac{5}{3}}.\tag{2.89}$$

- In the UR limit we have

$$\begin{aligned}\frac{E}{V} &= \frac{4\pi g_s}{h^3} \int_0^{p_F} (pc) p^2 dp = \pi g_s (hc) \left(\frac{p_F}{h} \right)^4 \\ &= \frac{3hc}{4} \left(\frac{3}{4\pi g_s} \right)^{\frac{4}{3}} n^{\frac{4}{3}}.\end{aligned}\tag{2.90}$$

But $P = \frac{1}{3}(E/V)$ for a UR gas, so

$$P \sim hc n^{\frac{4}{3}}.\tag{2.91}$$

We will need these formula when we discuss White Dwarf stars.

3. Stars and gravitational collapse

3.1. Thermal Pressure Support

A star is a self-gravitating ball of gas, mostly hydrogen, in mechanical equilibrium. If we assume spherical symmetry then the pressure $P(r)$ and mass density $\rho(r)$ are functions only of radial distance r from the centre. Let $m(r)$ be the mass within radius r and consider a shell at radius r of width dr :

Fig. III.1

The mass of the shell is $dm = 4\pi r^2 \rho(r) dr$, so $m(r)$ and $\rho(r)$ are related by

$$m'(r) = 4\pi r^2 \rho(r). \quad (3.1)$$

Given $\rho(r)$ we can integrate this to find $m(r)$, and hence m as a function of ρ . The boundary condition is $m(R) = M$, where R is the star's radius and M its total mass.

What is the gravitational force on an element of the shell of surface area dA ? We have an isotropic mass distribution so we can apply both of the theorems, Birkhoff's and Newton's, that we used previously in the context of cosmology. According to Birkhoff's theorem the force on the shell element due to matter beyond radius r cancels. According to Newton's theorem, the force due to the mass $m(r)$ within radius r is the same as if it were all concentrated at the centre. The mass of the shell element is $\rho(r) dr dA$, so Newton's law of gravity tells us that the shell element experiences a force of magnitude

$$F_{grav} = \frac{Gm(r)\rho(r)drdA}{r^2}, \quad (3.2)$$

directed radially inwards. In equilibrium this must be canceled by a radial outward force due to the pressure gradient; its magnitude is the pressure difference across the shell times the area dA of the shell element, so

$$F_{press} = [-P'(r)dr]dA, \quad (3.3)$$

which is positive if (as we expect) $P(r)$ is a decreasing function of r .

Fig. III.2

The condition for mechanical equilibrium is $F_{grav} + F_{press} = 0$, or

$$P' = -\frac{Gm\rho}{r^2}. \quad (3.4)$$

This is the (Newtonian) *pressure support* equation. Given an *equation of state* $P = P(\rho)$ or, equivalently, $\rho = \rho(P)$, the pressure support equation can be integrated to yield P as a function of r , subject to the obvious boundary condition that $P(R) = 0$.

Differentiating (3.4) and then using (3.1), we obtain the second order ODE

$$\left(\frac{r^2 P'}{\rho}\right)' = -4\pi G r^2 \rho. \quad (3.5)$$

Given $\rho(P)$ we can integrate this subject to the boundary conditions

$$P(0) = P_c, \quad P(R) = 0, \quad (3.6)$$

where P_c is the central pressure, which is related to the total mass M . The relation depends on the equation of state (which will *not* be linear, as it was in the cosmology context). Without knowing the equation of state one can still derive a lower bound on P_c in terms of M and R (see Q.II.4).

Near the centre of the star,

$$m(r) \approx \frac{4}{3}\pi r^3 \rho_C \quad (3.7)$$

where $\rho_C \equiv \rho(0)$ is the central mass density. Thus (from the pressure support equation)

$$P' \sim -\left(\frac{4\pi G \rho_C^2}{3}\right) r \quad (r \rightarrow 0). \quad (3.8)$$

and hence

$$P(r) = P_C - \left(\frac{2\pi G \rho_C^2}{3}\right) r^2 + \mathcal{O}(r^3). \quad (3.9)$$

Not surprisingly, the pressure is a maximum at the centre. It will decrease monotonically away from the centre until it reaches zero at $r = R$.

3.2. Virial Theorem

For a gravitationally bound system, mechanical equilibrium implies a relation between its (negative) gravitational potential energy E_{grav} and its kinetic energy E_{kin} . This is called a ‘virial theorem’.

For a star, the KE determines the average pressure

$$\langle P \rangle = \frac{1}{V} \int_{star} P dV, \quad (3.10)$$

so we expect the condition for equilibrium, the pressure-support equation, to relate E_{grav} and $\langle P \rangle$. Multiplying the pressure-support equation by $4\pi r^3$ we have

$$(4\pi r^3)P' = -\frac{Gm}{r}(4\pi r^2 \rho) = -\frac{Gmm'}{r}, \quad (3.11)$$

and hence

$$(4\pi r^3 P)' - 3(4\pi r^2 P) = -\frac{Gmm'}{r}. \quad (3.12)$$

Integrate from $r = 0$ to $r = R$. Since $P(R) = 0$ the first term does not contribute and we have

$$-3 \int_0^R P(r) [4\pi r^2 dr] = - \int_0^R \frac{Gm[m' dr]}{r}. \quad (3.13)$$

The RHS is the total gravitational potential energy E_{grav} , and $[4\pi r^2 dr]$ is the volume of a shell of depth dr at radius r , so

$$-3 \int_{star} P dV = E_{grav}, \quad (3.14)$$

and hence

$$\langle P \rangle V = -\frac{1}{3} E_{grav}. \quad (3.15)$$

This is the *Virial Theorem* (for stars). We now consider its implications for ideal gases in the NR and UR limits:

- For a NR gas we have $P = \frac{2}{3}(U/V)$, or $(U/V) = \frac{3}{2}P$. Integrating the latter relation over the star we have

$$E_{kin} = \frac{3}{2} \int_{star} P dV = \frac{3}{2} \langle P \rangle V, \quad (3.16)$$

and hence

$$E_{grav} = -2E_{kin} \quad (NR). \quad (3.17)$$

The total energy (excluding rest-mass energy) is

$$U \equiv E_{kin} + E_{grav} = -E_{kin} < 0 \quad (3.18)$$

so the star is indeed gravitationally bound, with a binding energy equal to E_{kin} .

- In the UR case we have $P = \frac{1}{3}(E/V)$ or $(E/V) = 3P$. Since the rest mass energy is now negligible, (E/V) is the kinetic energy density, so integrating both sides of the equation $(E/V) = P$ over the star we have

$$E_{kin} = 3 \langle P \rangle V, \quad (3.19)$$

and hence, from (3.15),

$$E_{grav} = -E_{kin}. \quad (3.20)$$

In this case

$$E \equiv E_{kin} + E_{grav} = 0 \quad (3.21)$$

so the binding energy is zero. What this means is that a gas of UR particles cannot form a stable gravitationally bound system. This result has several important consequences. One arises from the fact that for a sufficiently massive star the central pressure would be so great that the gas particles near the centre would be

ultra-relativistic. Since these particles could not hold the core together there is a theoretical upper bound on the mass of a star; detailed calculations show this to be $\sim 50M_{\odot}$ and this is confirmed by observations. A second application is to White Dwarfs, which we consider later.

3.3. Elementary particle interactions

There are three ‘families’ of elementary fermions (which interact with each other via the exchange of various types of elementary bosons). Each family contains two types of *quark* (each coming in three ‘colours’), and two types of *lepton*. We need only concern ourselves with the first family because this contains all the known stable particles. The leptons of the first family are the (electrically-charged) electron (e^{-}) and the (electrically-neutral) electron-neutrino (ν_e), which we shall simply call ‘the neutrino’; the neutrino is now believed to have a very small (but still unknown) mass, but we shall suppose it to be massless. The quarks of the first family bind to form the (‘colourless’) proton (p) and neutron (n), which are jointly referred to as *nucleons*. The proton carries an electric charge of opposite sign but equal magnitude to that of the electron, while the neutron is electrically neutral. All four of these particles have anti-particles: the positron (anti-electron) (e^{+}); anti-neutrino ($\bar{\nu}_e$); anti-proton (\bar{p}); anti-neutron (\bar{n}).

These particles are subject to gravity and three non-gravitational forces, the long-range electromagnetic force (transmitted by photons); the short (nuclear) range Weak force (transmitted by ‘W’ and ‘Z’ bosons); the short (nuclear) range Strong forces (transmitted by ‘gluons’). It is the Strong force that is responsible for binding the quarks into protons and neutrons, and for binding the latter in atomic nuclei; it has to be very strong to overcome the electrostatic repulsion of protons in atomic nuclei.

Although the proton, electron and neutrino are stable (with respect to all *known* interactions) a free neutron is unstable, albeit with a very long lifetime (for an elementary particle) of about 15 minutes. It decays via the Weak interaction

$$n \rightarrow p + e^{-} + \bar{\nu}_e . \quad (3.22)$$

This reaction is responsible for the radioactive decay of some unstable atomic nuclei; the electrons are ejected from the nucleus and are detected as ‘beta-particles’ (the name predates the identification of these particles as electrons) so the reaction (3.22) is known as *beta-decay*. The anti-neutrino escapes the nucleus undetected, apart from the energy it carries away with it, because it interacts only via the Weak interaction

$$\bar{\nu}_e + p \rightarrow n + e^{+} . \quad (3.23)$$

where e^{+} is a positron (anti-particle to the electron). The neutrino was first postulated by Pauli as a means of avoiding an apparent violation of energy conservation in β decay.

Note that the reaction (3.23) can be deduced from (3.22) by the rule that a particle on one side of a reaction can be replaced by its anti-particle on the other side. The same rule shows that electrons can be captured by protons via an *inverse beta decay*

$$e^- + p \rightarrow n + \nu_e, \quad (3.24)$$

but this reaction is endothermic (it requires an input of energy) so it does not imply an instability of matter under normal circumstances. Another application of the rule yields

$$p \rightarrow n + e^+ + \nu_e \quad (3.25)$$

but this is forbidden by energy conservation: the proton is stable (at least against all *known* interactions). Suppose we now add a proton to both sides of (3.25). This does not change the energy balance by itself, but a proton and a neutron can bind to form a deuteron d (nucleus of deuterium), and the reaction

$$p + p \rightarrow d + e^+ + \nu_e \quad (3.26)$$

is exothermic (produces energy). This reaction is important for nuclear fusion in stars. A nuclear fusion reactor might also rely on this reaction if we could build one, but the Coulomb barrier makes it difficult to get the protons close enough for the reaction to proceed.

3.4. Stellar evolution

Stars like the Sun are prevented from undergoing complete gravitational collapse by their thermal pressure. Since energy is being radiated into space, a constant temperature requires an energy source. One such source is the gravitational potential energy itself. If the sun were to suddenly shrink then its gravitational energy would be lowered so its kinetic energy, and hence its temperature, would have to rise. Alternatively, and more realistically, it could maintain a constant temperature in spite of losing energy into space by gradually shrinking. However, the sun would not last more than a million years if this were the source of its energy, and this is much less than the age of the Earth. Instead, the energy comes from a series of nuclear reactions in the core that fuse hydrogen to helium, releasing energy in the process. In the first step, protons fuse to deuterium, according to the reaction (3.26) releasing both energy and neutrinos; additional energy is released subsequently when the positrons annihilate with electrons. A high central pressure is needed to overcome the ‘Coulomb barrier’ before the nuclear reactions can begin, so this puts a lower limit on the size of a star. The neutrinos produced by fusion of hydrogen to helium escape from the star, virtually unhindered because the weak interaction is so weak. Some also pass through the Earth, again almost unhindered but this *solar neutrino flux*

can be, and has been, detected. [However, the results of these solar neutrino experiments agree with theory only if the neutrino has some small, as yet undetermined, mass.]

The next step towards fusion of helium from hydrogen is the formation of ${}^3\text{He}$ nuclei via the (very fast) electromagnetic interaction

$$p + d \rightarrow {}^3\text{He} + \gamma. \quad (3.27)$$

The ${}^3\text{He}$ is the (unstable) nucleus of the helium isotope with two protons and one neutron, and γ is energy in the form of photons. We then have conversion of ${}^3\text{He}$ nuclei into ${}^4\text{He}$ nuclei (alpha-particles) and protons via the Strong Interaction

$${}^3\text{He} + {}^3\text{He} \rightarrow {}^4\text{He} + p + p, \quad (3.28)$$

where ${}^4\text{He}$ is the nucleus of the stable isotope of helium with two protons and two neutrons, also known as an *alpha-particle*. The net result of these reactions is the conversion of two hydrogen atoms to one helium atom with the production of energetic photons, which heat the star, and neutrinos which escape from it.

Eventually, the hydrogen at the core is used up, having been converted into helium. The core then contracts, and the pressure increases, until another set of nuclear reactions ‘burn’ helium to heavier elements such as carbon. These reactions pump more heat/second into the star than before (necessarily, since a higher temperature is needed to support the more compact core). Somewhat paradoxically, this results in the star expanding and cooling. This is implied by the virial theorem: from (3.18) we see that an increase in the total energy E implies a decrease in the kinetic energy, so the average temperature must decrease, implying a cooler surface. From (3.17) we see that a decrease in the kinetic energy implies an increase in the gravitational potential energy, which is now *less negative* than it was before. Since the mass M is fixed this implies an increase in the radius R . This cooler larger star, burning Helium in its core, is called a **Red Giant**. About 5×10^9 years from now our Sun will become a Red Giant, engulfing all the inner planets including the Earth. When the Helium is gone the carbon core will cool and contract until it is supported by electron degeneracy pressure (to be discussed further below). As the total energy of the star must decrease, the virial theorem now implies that it will contract and heat up (by the reverse of the previous argument for Red Giants). The result is a **White Dwarf**.

For a star with a mass much greater than M_\odot , the central pressure will rise to the point at which carbon ‘burns’ to heavier elements, such as nitrogen and oxygen and, ultimately, iron. There are no nuclear reactions that can burn iron to yet heavier elements (all such reactions are endothermic) so the iron core must eventually collapse—since the core mass exceeds the Chandrasekar limit (to be explained below) it cannot be supported by electron degeneracy pressure. The result is a (type II) supernova, and the end product is a *neutron star* or a *black hole*.

3.5. White Dwarfs

Since all nuclear reactions in a stellar core must eventually run out of fuel, all stars must eventually cool to $T = 0$. What are the possible ‘final states’ of these dead stars? At $T = 0$ the only force that could counteract the force of gravity is the force exerted by degeneracy pressure. For there to be a significant degeneracy pressure the hydrogen atoms must be close enough for the electrons in one atom to interact with those in neighbouring atoms, so we can think of the star as an electron-proton plasma (gas of charged particles).

For either the electrons or the protons to become degenerate they must be compressed to a number density

$$n \sim 1/\lambda^3, \quad (\lambda = h/p), \quad (3.29)$$

where p is a typical momentum (of the order of the Fermi momentum), and λ is the corresponding de Broglie wavelength. The actual number density of electrons or protons is the same because by charge neutrality the total number of electrons must equal the total number of protons, but the value of p (and hence λ) is different for protons and electrons. This follows from equipartition of energy in thermal equilibrium, which requires that

$$\frac{p_p^2}{m_p} \approx \frac{p_e^2}{m_e}. \quad (3.30)$$

This implies that

$$\lambda_p = \left(\frac{m_e}{m_p}\right)^{\frac{1}{2}} \lambda_e \ll \lambda_e. \quad (3.31)$$

So, as the number density n increases, the electrons will become degenerate long before the protons do. As long as electron degeneracy is sufficient to support the star at $T = 0$ we can ignore the protons.

But can electron degeneracy pressure support a star against gravitational collapse? The total gravitational potential energy is, on dimensional grounds,

$$E_{grav} \sim -\frac{GM^2}{R} \quad (3.32)$$

where M is the total mass and R the radius. The total kinetic energy is

$$E_{kin} \sim \left(\frac{\langle U \rangle}{V}\right) R^3 \sim \langle P \rangle R^3. \quad (3.33)$$

Assuming NR degenerate electrons and using the formula (2.89) for the degeneracy pressure we have

$$E_{kin} \sim \frac{h^2}{m_e} \langle n \rangle^{\frac{5}{3}} R^3. \quad (3.34)$$

Now, $M \approx m_p n R^3$ since the electrons' contribution to the total mass is negligible, so

$$\langle n \rangle \sim \frac{M}{m_p R^3}, \quad (3.35)$$

and hence

$$E_{kin} \sim \frac{h^2}{m_e R^2} \left(\frac{M}{m_p} \right)^{\frac{5}{3}}. \quad (3.36)$$

The total energy is therefore

$$E(R) \equiv E_{grav} + E_{kin} = -\frac{\alpha}{R} + \frac{\beta}{R^2} \quad (3.37)$$

where

$$\alpha \sim GM^2, \quad \beta \sim \frac{h^2}{m_e} \left(\frac{M}{m_p} \right)^{\frac{5}{3}}. \quad (3.38)$$

Note that α and β are constants because the total mass M is fixed, whereas the radius R is to be determined by minimization of $E(R)$.

Fig. III.3

There is a minimum, and it occurs at

$$R \sim \frac{h^2 M^{-\frac{1}{3}}}{G m_e m_p^{\frac{5}{3}}}. \quad (3.39)$$

In the 1930s a few **White Dwarf** stars were known that were very bright but far too compact to be normal (main sequence) stars. Fowler argued that they were compact because they were supported by electron degeneracy pressure. As long as $kT \ll \epsilon_F$, thermal pressure is negligible, so White Dwarfs may be ‘white hot’.

We have still to check that the assumption of non-relativistic electrons is self-consistent. The NR approximation requires that $p_F c \ll m_e c^2$. Since $p_F \sim \hbar n^{1/3}$ this is equivalent to

$$\langle n \rangle \ll \left(\frac{m_e c}{\hbar} \right)^3. \quad (3.40)$$

However, the average electron density at the equilibrium point is

$$\langle n \rangle \sim \frac{M}{m_p R^3} \sim \left(\frac{G m_e}{\hbar^2} \right)^3 m_p^4 M^2, \quad (3.41)$$

so the conclusion that there exists an equilibrium point is consistent with the assumption of NR electrons provided that

$$M \ll \frac{1}{m_p^2} \left(\frac{\hbar c}{G} \right)^{\frac{3}{2}}. \quad (3.42)$$

If this is not satisfied then our formula for $E(R)$ is invalid because the electrons are relativistic. We saw earlier from the virial theorem that ultra-relativistic particles cannot support a system against gravitational collapse, so we should now suspect that there is some maximum mass for White Dwarf stars.

To check this, let us now suppose that the electrons have become ultra-relativistic. In this case we may use the formula (2.91) to get

$$E_{kin} \sim \hbar c \langle n \rangle^{\frac{4}{3}} R^3 \sim \frac{\hbar c}{R} \left(\frac{M}{m_p} \right)^{\frac{4}{3}}, \quad (3.43)$$

and hence

$$E(R) = \frac{\gamma - \alpha}{R}, \quad (3.44)$$

where

$$\gamma \sim \hbar c \left(\frac{M}{m_p} \right)^{\frac{4}{3}}. \quad (3.45)$$

If $\gamma > \alpha$ then R will increase, and $\langle n \rangle$ will decrease until the UR approximation fails.

Fig. III.4

Eventually, the electrons will become non-relativistic and the star will be supported by electron degeneracy pressure at the radius we found above. If $\gamma < \alpha$ then R will decrease, and $\langle n \rangle$ will increase until the protons become degenerate.

Fig. III.5

We will then need to investigate whether the star can be supported by proton degeneracy pressure but, in any case, the star will not end up as a White Dwarf. Thus, the condition $\gamma = \alpha$, yields a *maximum* mass M_C for a White Dwarf, which is

$$M_C \sim \frac{1}{m_p^2} \left(\frac{\hbar c}{G} \right)^{\frac{3}{2}} \approx 1.85 M_\odot. \quad (3.46)$$

A proper calculation shows that $M_C \approx 1.4 M_\odot$; this is called the **Chandrasekhar limit**. No White Dwarf stars have been observed with $M > M_C$.

We have implicitly assumed that the GR corrections to Newtonian theory are negligible and we should now check this. Recall from (3.39) that $R \sim M^{1/3}$ for a WD. Since $M < M_C$, this puts a lower bound on the radius:

$$R > \frac{h^2 M_C^{\frac{1}{3}}}{G m_e m_p^{\frac{5}{3}}} \sim \left(\frac{m_p}{m_e} \right) \left(\frac{GM}{c^2} \right). \quad (3.47)$$

Equivalently,

$$\frac{GM}{c^2 R} \sim \left(\frac{m_e}{m_p} \right) \ll 1. \quad (3.48)$$

The dimensionless ratio $GM/c^2 R$ is a measure of the strength of the gravitational field. Since it is small, our neglect of GR corrections was justified.

3.6. Neutron Stars and Black Holes

A star that has exhausted its nuclear fuel cannot end its days as a WD if the mass of its core exceeds the Chandrasekhar limit. Gravity will compress it until the electrons become

degenerate, but will then keep compressing it. As the electron Fermi energy rises there will come a point at which electrons have sufficient energy for the inverse beta-decay reaction (3.24), which then removes the electrons and with it the electron degeneracy pressure; the neutrinos escape from the star, and can be detected on earth by the same methods used to detect solar neutrinos. With no pressure to support it the star's core now undergoes a free-fall collapse to nuclear matter density. The shockwave produced by this fall blows off all the outer layers of the star in a gigantic explosion that produces what we observe as a (Type II) *supernova* (Type I supernovas occur when a WD accretes matter that causes its mass to exceed the Chandrasekhar limit).

We are then left with an extremely compact neutron core of nuclear matter density. Its only means of support is neutron degeneracy pressure. If we make the same assumptions as we made when considering whether electron degeneracy pressure could support a star then we will get the same results but with m_e replaced by $m_n \approx m_p$. The formula (3.46) for the critical mass is independent of m_e so we again find that

$$M_C \sim \frac{1}{m_p^2} \left(\frac{\hbar c}{G} \right)^{\frac{3}{2}}, \quad (3.49)$$

whereas the formula (3.39) for the radius of a WD becomes

$$R \sim \frac{h^2 M^{-\frac{1}{3}}}{G m_p^{\frac{8}{3}}} \quad (3.50)$$

for a neutron star. For $M \sim M_C$ this implies that

$$\frac{GM}{c^2 R} \sim 1. \quad (3.51)$$

In fact, the left hand side is necessarily less than 1/2 for a reason to be explained below, but this still implies a strong gravitational field that invalidates the Newtonian approximation to GR. In addition, the ideal gas approximation is no longer a good one, so (3.49) cannot be trusted, although there must be *some* maximum mass. A proper calculation, using GR and nuclear physics, yields $M_{max} \approx 3M_\odot$. If, after the supernova explosion, the neutron core has a subcritical mass it will become a **Neutron Star**, supported by neutron degeneracy pressure (observationally, neutron stars make their presence known as *pulsars*). If, on the other hand, the neutron core has a mass exceeding the critical mass then it cannot be supported by neutron degeneracy pressure and must continue to collapse until

$$R \rightarrow \frac{2GM}{c^2} = R_S \quad (3.52)$$

This is the **Schwarzschild radius**. Any spherically-symmetric object with a radius less than its Schwarzschild radius is invisible because light cannot escape from its surface. In

fact, $R = R_S$ is really a minimum radius, and an object that has collapsed to this radius is a **Black Hole** [the surface at $R = R_0$ is the black hole's *event horizon*, analogous to the cosmological event horizon of the de Sitter universe]. Black holes formed from the collapse of massive stars might have masses of the order of ten solar masses, but supermassive black holes that contain a large fraction of the total mass of a galaxy are now believed to be at the centre of many galaxies, including our own. However, Black Holes can only be properly understood within GR.

4. Thermal History of the Universe

4.1. Photon gas

- Photons interact very weakly (light beams pass through one another), so the ideal gas approximation is good; and photons are bosons, so we can use the Bose-Einstein distribution.
- Photons are massless, so $E = pc$.
- Photon number is not conserved, so the number N of photons will adjust itself, in an isolated system, so as to maximize the entropy; this means that $\partial S/\partial N$ must vanish in equilibrium, so (from its definition) $\mu = 0$.

These properties imply that

$$\bar{n}(p) = \frac{g(p)}{e^{\beta cp} - 1} \quad (4.1)$$

for a photon gas in thermal equilibrium. Because the spin degeneracy factor for photons is 2 (two possible polarization states) we have

$$g(p) = 2 \times \left(\frac{4\pi V}{h^3} \right) p^2 = \left(\frac{8\pi V}{h^3} \right) p^2, \quad (4.2)$$

and hence

$$N = \frac{8\pi V}{h^3} \int_0^\infty \frac{p^2 dp}{e^{\beta cp} - 1}, \quad E = \frac{8\pi V}{h^3} \int_0^\infty \frac{(pc)p^2 dp}{e^{\beta cp} - 1}. \quad (4.3)$$

Now let

$$x = \beta cp \quad (4.4)$$

and define $n_\gamma = (N/V)$ and $\varepsilon_\gamma = (E/V)$ to be the photon number density and energy density, respectively. Then

$$n_\gamma = \frac{8\pi}{(hc)^3} (kT)^3 \int_0^\infty \frac{x^2 dx}{e^x - 1}, \quad \varepsilon_\gamma = \frac{8\pi}{(hc)^3} (kT)^4 \int_0^\infty \frac{x^3 dx}{e^x - 1}. \quad (4.5)$$

The integrals can be expressed in terms of the *Riemann zeta function*

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} \quad (4.6)$$

according to the following lemma:

$$\int_0^{\infty} \frac{x^n dx}{e^x - 1} = n! \zeta(n+1). \quad (4.7)$$

Thus,

$$n_{\gamma} = \frac{16\pi\zeta(3)}{(hc)^3} (kT)^3, \quad \varepsilon_{\gamma} = \frac{48\pi\zeta(4)}{(hc)^3} (kT)^4. \quad (4.8)$$

Note that

$$\frac{\varepsilon_{\gamma}}{n_{\gamma}} \approx 3kT, \quad (4.9)$$

since

$$\zeta(3) \approx 1.2, \quad \zeta(4) = \frac{\pi^4}{90}. \quad (4.10)$$

The expression for the energy density can be written as

$$\varepsilon_{\gamma} = 4c^{-1} \sigma T^4, \quad (4.11)$$

where the constant

$$\sigma = \frac{\pi^2 k^4}{60 \hbar^3 c^2} \quad (4.12)$$

is called the *Stefan-Boltzmann* constant. This result for the energy density is directly related to the *Stefan-Boltzmann law* which states that the energy flux radiated per unit time through unit surface area of a perfect absorber (black body) is σT^4 . Defining the *Radiation Density* constant

$$\alpha = 4c^{-1} \sigma = \frac{\pi^2 k^4}{15 \hbar^3 c^3}, \quad (4.13)$$

we have

$$\varepsilon_{\gamma} = \alpha T^4. \quad (4.14)$$

4.2. Planck's radiation formula and Wien's law

Returning to (4.5) we can rewrite the energy density integral as

$$\varepsilon_{\gamma} = \int_0^{\infty} \epsilon_{\gamma}(\nu) d\nu \quad (4.15)$$

where ν is the photon frequency. Since $p = h\nu/c$ we have

$$\epsilon(\nu) = \left(\frac{8\pi h}{c^3} \right) \frac{\nu^3}{e^{\frac{h\nu}{kT}} - 1}. \quad (4.16)$$

This is the **Planck radiation formula** giving the energy density of electromagnetic radiation in a cavity at temperature T (historically, the origin of quantum theory). For $h\nu \ll kT$ Planck's formula reduces to the Rayleigh-Jeans (RJ) formula

$$\epsilon(\nu) = \left(\frac{8\pi\nu^2}{c^3} \right) kT \quad (h\nu \ll kT). \quad (4.17)$$

Note how Planck's constant no longer appears. The RJ formula was originally derived on the assumptions of classical electromagnetism; its extrapolation yields the evidently absurd result that $\epsilon(\nu) \rightarrow \infty$ as $\nu \rightarrow \infty$. This is the 'ultraviolet catastrophe' of classical physics that was resolved by Planck's quantum hypothesis. For $h\nu \gg kT$, the Planck formula implies that

$$\epsilon(\nu) \approx \left(\frac{8\pi h}{c^3} \right) \nu^3 e^{-\frac{h\nu}{kT}}. \quad (4.18)$$

This is the *exponential tail* of the Planck distribution; it is of great importance to cosmology. A further feature of Planck's radiation formula is that $\epsilon(\nu)$ has a single critical point, a maximum, at a value ν_{peak} of ν given by

$$h\nu_{peak} \approx 3kT. \quad (4.19)$$

Prior to Planck's derivation of his radiation density formula, Wien had shown from general thermodynamic arguments that

$$\epsilon(\nu) = \nu^3 g(\nu/T) \quad (4.20)$$

for some function g . The RJ formula takes this form but with a function g that has no maximum. Wien supposed that the correct formula would result in a function g with a maximum at some frequency ν_{peak} . It then follows that

$$\nu_{peak} \propto T. \quad (4.21)$$

This is known as **Wien's law**. We may rewrite (4.20) as

$$\epsilon(\nu) = \nu^3 f(x) \quad (4.22)$$

for some function f , where we have re-introduced the variable

$$x = \beta c p = \frac{h\nu}{kT}. \quad (4.23)$$

Planck's formula confirms Wien's arguments, and shows that

$$g(y) = \left(\frac{8\pi h}{c^3} \right) [e^{hy/k} - 1]^{-1} \quad (4.24)$$

In terms of the dimensionless variable x we can write

$$\epsilon(\nu) = \left(\frac{8\pi}{h^2 c^3} \right) (kT)^3 f(x) \quad (4.25)$$

where

$$f(x) = \frac{x^3}{e^x - 1} \quad (4.26)$$

This function has a maximum at $x \approx 3$, which yields the result (4.19)

Fig. IV.1

Wien's law can be written in terms of the wavelength $\lambda_{peak} \equiv c/\nu_{peak}$ as

$$\lambda_{peak} \approx \left(\frac{hc}{3k} \right) T^{-1} \quad (4.27)$$

This gives the wavelength of a typical photon in a thermal gas of photons at temperature T . If we know, or assume, that the radiation is thermal then a measurement of λ_{peak} determines the temperature; e.g.,

$$\lambda_{peak} \sim 10^{-3} m \Leftrightarrow T \sim 3^\circ K \quad (4.28)$$

4.3. Photon pressure from Kinetic Theory *

As we saw earlier, the pressure of any ideal UR gas equals one-third of its energy density. This result can also be derived from Maxwell's *kinetic theory of gases* (which was the forerunner of statistical mechanics).

Let $f(\mathbf{p})d^3p$ be the contribution to the photon number density n_γ from photons with momentum in the volume element d^3p in momentum space. Writing d^3p as $dpd\Omega$, where $d\Omega = \sin\theta d\theta d\phi$ is the (momentum space) *solid angle* element, we have

$$n_\gamma = \int_0^\infty dp \int d\Omega f(\mathbf{p}), \quad (4.29)$$

where the first integral is over the unit sphere. Similarly, the energy density is

$$\varepsilon_\gamma = \int_0^\infty dp \int d\Omega (pc) f(\mathbf{p}). \quad (4.30)$$

We will now assume that f actually depends only on p and not on the angles θ and ϕ (if this were not the case, there would be some preferred momentum and hence a preferred direction in space). In this case

$$\varepsilon_\gamma = 4\pi \int_0^\infty (pc) f(p) dp. \quad (4.31)$$

This has to be compared to the photon pressure P_γ .

According to the kinetic theory of gases, the pressure on any surface is due to elastic collisions of gas molecules with it. If a molecule of momentum p approaches the surface at an angle θ to the normal then the collision transfers momentum $2p \cos \theta$ to the surface. The number of photons striking *unit surface area in unit time* with momentum in the momentum space volume element at momentum \mathbf{p} is $f(\mathbf{p})$ times the volume $c \cos \theta$ of the ‘sloping cylinder’ of length c and unit area base.

Fig. IV.2

The gas pressure is the momentum transfer to this unit surface area in unit time, so

$$\begin{aligned}
P_\gamma &= \int_0^\infty dp \int_0^{\frac{\pi}{2}} \sin \theta d\theta \int_0^{2\pi} d\phi (2p \cos \theta)(c \cos \theta) f(p) \\
&= 2 \int_0^\infty (pc) f(p) dp \int_{-1}^1 \cos^2 \theta d(\cos \theta) \int_0^{2\pi} d\phi \\
&= \frac{4\pi}{3} \int_0^\infty (pc) f(p) dp = \frac{1}{3} \varepsilon_\gamma.
\end{aligned} \tag{4.32}$$

4.4. Photon gas entropy

Since $\mu = 0$ for photons, the first law of thermodynamics states that $dE = TdS - PdV$, or

$$\begin{aligned}
TdS &= dE + PdV \\
&= d(\varepsilon_\gamma V) + \frac{1}{3} \varepsilon_\gamma dV \\
&= \frac{4}{3} \varepsilon_\gamma dV + V d\varepsilon_\gamma
\end{aligned} \tag{4.33}$$

Now use the radiation density formula (4.14) to get

$$TdS = \alpha \left(\frac{4}{3} T^4 dV + 4T^3 V dT \right) \tag{4.34}$$

and hence

$$\begin{aligned}
dS &= \frac{4}{3} \alpha (T^3 dV + 3T^2 V dT) \\
&= \frac{4}{3} \alpha d(VT^3)
\end{aligned} \tag{4.35}$$

Integrating, we have

$$S = \frac{4}{3} \alpha VT^3 + S_0 \tag{4.36}$$

for some constant S_0 . According to the *third law of thermodynamics* the entropy must vanish at $T = 0$. This law does not have the same status as the other laws of thermodynamics but it is obeyed by ideal gases, so S_0 must vanish. It then follows that the photon entropy density $(S/V) \equiv s_\gamma$ is given by the formula

$$s_\gamma = \frac{4}{3} \alpha T^3. \tag{4.37}$$

4.5. Adiabatic expansion and the CMBR

When a thermally isolated system passes from one equilibrium state to another the entropy will generally increase. However, the increase in entropy may be very small if the transition is slow and smooth. If the typical timescale determined by the process of change is much longer than the relaxation time of the system (the time it takes to return to equilibrium after a small perturbation) then the system effectively passes from the initial equilibrium state to the final one through a continuous sequence of other equilibrium states. This is called a ‘quasi-static’ change of state. A quasi-static change of an isolated system is always *isentropic* because the entropy cannot change as long as the system remains in equilibrium. This follows from the fact that any reaction capable of changing the occupation numbers of the microstates proceeds equally in both directions when the system is in equilibrium, but these occupation numbers determine the entropy. As a check, suppose that a photon gas undergoes a change of its volume V . If the occupation numbers do not change then neither does the total number of photons $N_\gamma \propto VT^3$, so the temperature T must change so as to keep VT^3 fixed. But $S \propto VT^3$ so the total entropy will also not change.

Suppose that we have a universe dominated by radiation in the form of a photon gas in thermal equilibrium. This photon gas is not *mechanically* isolated because gravity can perform work on it, but it *is* thermally isolated. The expansion of the universe can be considered as being quasi-static, so the total entropy will be conserved. As we have seen, this implies that VT^3 is constant, and since $V \propto a^3$ we deduce that

$$T \propto \frac{1}{a}. \quad (4.38)$$

In other words, the photon gas cools as it expands. This is called *adiabatic cooling* because a change under conditions of thermal isolation is, by definition, **adiabatic**. For this reason, cosmologists usually say that the photon gas has undergone an **adiabatic expansion**. We shall conform to this usage here even though it is not strictly accurate: a *sudden* adiabatic expansion will leave the temperature of an ideal gas unchanged. **It will therefore be implicit in what we mean here by adiabatic expansion that the expansion is quasi-static and hence isentropic.** As the photon gas cools it loses energy since $E_\gamma \propto VT^4$ and $VT^4 \propto T$ at constant entropy. Using (4.38) we then see that $E_\gamma \propto 1/a$. The energy lost by the photon gas goes into an increase of the gravitational potential energy.

In 1965 Penzias and Wilson detected an isotropic cosmic microwave background radiation (CMBR). They were looking at wavelengths $\sim 7\text{cm}$ (in the ‘microwave’ range). This was immediately interpreted as a photon gas that had been cooled by the expansion of the universe. For this to be correct, the frequency spectrum should be both thermal and isotropic. The thermal nature of the CMBR was confirmed to high precision in 1992,

and its temperature was determined to be

$$T = 2.728 \pm 0.004 K. \quad (4.39)$$

This corresponds to a peak at $\lambda_{peak} \sim 0.1 cm$, so the original observations were made quite far from the peak. The 1992 observations also confirmed the isotropy of the CMBR to one part in 10^5 , but fluctuations were found with

$$\frac{\Delta T}{T} \sim 10^{-5}. \quad (4.40)$$

For convenience we will round off the CMBR temperature to 3 K. The photon number density at this temperature is

$$n_\gamma \sim 4 \times 10^8 m^{-3}. \quad (4.41)$$

4.6. The baryon to photon ratio

As mentioned earlier, observations suggest that the total energy density of the universe is close to the critical density

$$\rho_c(t_0)c^2 \sim 10 GeV m^{-3}. \quad (4.42)$$

The photons in the CMBR provide a negligible fraction of this, so the universe is now matter dominated. If all the NR matter were in the form of baryons (we can neglect the electrons) the baryon number density would have to be about $10 m^{-3}$ since the proton rest-mass energy is about 1 GeV. However, observations imply that no more than 1/100 of the total mass can be baryonic, the remainder being *dark matter* of unknown type. Thus, the actual baryon number density must be much lower. A lower bound is provided by the visible matter in galaxies, and the best current estimate is

$$n_b \sim \frac{1}{20} m^{-3}. \quad (4.43)$$

This gives a ratio

$$\frac{n_b}{n_\gamma} \equiv \eta \sim 10^{-10}. \quad (4.44)$$

This ratio remains constant during the expansion of the universe because (i) baryon number conservation implies $n_b \sim a^{-3}$ and (ii) $n_\gamma \sim T^3 \sim a^{-3}$ for adiabatic expansion.

The smallness of the ratio η seems to call for an explanation. However, the real puzzle is why it should not be zero. The photon is its own antiparticle but protons and neutrons each have antiparticles (anti-proton and anti-neutron) with baryon number -1 . Given sufficient energy, baryon-antibaryon pairs can be pair-produced by processes such as

$$\gamma + p \rightarrow p + p + \bar{p}, \quad (4.45)$$

so at sufficiently high temperature there will always be approximately as many baryons as antibaryons. According to the Big Bang theory of cosmology there must have been a time at which the temperature was this high. If there were then an *exact* equality between numbers of particles and anti-particles they would all annihilate into photons as the universe cooled, leaving no baryons at all. Such a universe would be characterized by having $\eta = 0$. The fact that η is not exactly zero means that there was a small asymmetry between particles and antiparticles in the very early universe. But why should this be? From his exile in Gorki, the Russian physicist Sakharov pointed out that this asymmetry could arise spontaneously during a period in which the universe is out of thermal equilibrium provided that baryon number is not exactly conserved. It is now believed that baryon number is indeed not exactly conserved. Theories that purport to explain how a net baryon number is then created while the universe is out of equilibrium go by the name of *baryogenesis*, but there is no compelling theory that yields the observed value of η .

Note that the energy density in baryons is approximately $n_B(m_p c^2) = \eta m_p c^2 n_\gamma$, whereas $\varepsilon_\gamma \sim kT$, so if all the NR matter were in the form of baryons, the total energy density of NR matter would equal that in the CMBR when $kT \sim \eta m_p c^2 \sim 10^{-1} eV$. However, because baryons are estimated to form only about 1/200 of the total mass, the cross over from radiation domination to matter domination actually occurs when $kT \sim 20 eV$, corresponding to $T \sim 10^5 K$.

4.7. Recombination

In thermal equilibrium the occupation number n_k of the k th energy eigenstate is a function of $(E_k - \mu)/kT$. Since occupation numbers are unaffected by adiabatic expansion the temperature must change so as to keep $(E_k - \mu)/kT$ fixed [See Q.III.4]. Since $p_k \propto 1/a$ we can determine how E_k depends on a , from the formula $E_k^2 = p_k^2 c^2 + m^2 c^4$. Consider the the NR and UR limits:

- **UR case:** $E = pc$, $\mu = 0$ (as we have seen, for photons, but typical of UR particles), so $(E_k - \mu) \propto 1/a$. Hence $T \propto 1/a$
- **NR case:** $E_k = mc^2 + U_k$. Using (2.75) for μ we have

$$e^{\beta(E_k - \mu)} = \left(\frac{n}{g_s n_Q} \right) e^{\beta U_k} \quad (4.46)$$

Neglecting the slowly varying prefactor of the exponential we see that U_k/kT is constant when $\beta(E_k - \mu)$ is constant. As $U_k = p_k^2/2m$ we have $U_k \propto 1/a^2$. Hence $T \propto 1/a^2$

For an expanding universe that contains both NR and UR particles in thermal equilibrium, some compromise between these two cases will apply. For example, during an epoch for which

$$kT \ll 2m_e c^2 \approx 10^6 eV \quad (1 \text{ MeV}), \quad (4.47)$$

the main NR constituents of the universe must be electrons and protons, either as free particles or bound together as hydrogen atoms (for the moment we ignore neutrons; we will return to consider them later). The only other stable particles, apart from the UR neutrinos, are the positron and anti-proton, but at these temperatures there is insufficient thermal energy for the creation of electron positron pairs, and hence also insufficient energy for the creation of proton anti-proton pairs, and any pre-existing positrons or anti-protons will have annihilated into photons via the reactions

$$e^- + e^+ \rightarrow \gamma, \quad p + \bar{p} \rightarrow \gamma. \quad (4.48)$$

Let B be the binding energy of hydrogen (the energy required to ionize it). When

$$kT \gg B \approx 13.6 eV \quad (4.49)$$

there are essentially no hydrogen atoms but only a plasma of electrons and protons, with $n_e = n_p$ by charge neutrality. These NR particles will be kept in thermal equilibrium with the photons by electromagnetic scattering, e.g.,

$$\gamma + e^- \rightarrow \gamma + e^-. \quad (4.50)$$

The energy will be distributed equally among all particles, photons, electrons and protons (equipartition) but there are *far more photons than electrons or protons*; The ratio is $\eta^{-1} \sim 10^{10}$ (since $n_p = n_B$). This means that the temperature will be determined by the photons, so $T \propto 1/a$.

As the Universe expands and cools to $kT < B$ hydrogen atoms will begin to form. This is called **recombination**, even though electrons and protons have never previously been combined into atoms, because to study the reaction in a laboratory one first creates the plasma by ionizing hydrogen, which then recombines into hydrogen atoms (see Q.III.5). Recombination will occur when there are too few photons with energy sufficient to ionize hydrogen. Since the typical photon energy at temperature T is kT , a naive estimate of the temperature at recombination would be $kT \sim B$. But this fails to take into account the huge imbalance in the number of photons. Although kT is the typical photon energy there are *atypical* photons with much larger energies in the ‘exponential tail’ of the Planck distribution. Suppose that kT is significantly less than B . The number of photons with energy $\geq B$ will then be approximately

$$n_\gamma \times e^{-\frac{B}{kT}}, \quad (4.51)$$

and recombination will occur when this number is approximately equal to $n_p = \eta n_\gamma$. In other words, recombination occurs when $\exp(-B/kT) = \eta$, or

$$T \approx \frac{B}{k \log(\eta^{-1})} \approx 3000K. \quad (4.52)$$

Note that the universe is already matter-dominated well before recombination, so the rate of expansion at the time of recombination and afterwards is given by the EdS model.

4.8. Decoupling

Photons scatter only off *charged* particles. They do not interact with neutral particles. To the extent that a hydrogen atom can be considered as a ‘particle’ it is invisible to photons. In practice this means that a photon will not interact with a hydrogen atom if it has an energy insufficient to kick the electron out of its ground state. The energy needed is about the same as the ionization energy. As we have seen, the fraction of the total number of photons with at least this energy is already small at recombination, and rapidly decreases as the Universe cools. Thus, **around the same time as recombination the photons decouple from the non-relativistic matter**, which is mostly hydrogen (with an admixture of helium that we will consider later). At this point the photon gas continues to cool according to $T \propto 1/a$ but it is now out of thermal contact with the hydrogen gas, which cools much faster, according to the NR formula $T \propto 1/a^2$.

How big was the universe at the time of decoupling? Since $T_\gamma \propto 1/a$ and $T = T_0 \sim 3K$ when $a = 1$, the scale factor of the universe when the CMBR had temperature T is given by

$$a(T) = \frac{T_0}{T}. \quad (4.53)$$

At decoupling, when $T \sim 3000K$ this yields $a = 10^{-3}$. How old was the universe at decoupling? Since $a = (t/t_0)^{2/3}$ and $t_0 \sim 10^{10} \text{ yrs}$, the age of the universe at decoupling was

$$t_D \sim (10^{-3})^{3/2} t_0 \sim 3 \times 10^5 \text{ yrs}. \quad (4.54)$$

Since decoupling, the photons of the CMBR have been travelling unhindered. Those we detect on earth have therefore arrived from a surface in the sky that lies at a distance at which the age of the universe was only 3×10^5 years, and at which it was 1/1000 times smaller and 1000 times hotter. This is called the **surface of last scattering**. This surface represents the furthest distance that can be observed by any optical instrument, because further distances correspond to times at which the universe was opaque. Galaxies formed at a much later epoch when the universe was quite transparent, and so we have no difficulty seeing them (especially from space-based telescopes).

Fig. IV.3

4.9. The early universe

Although we cannot make direct observations of the Universe before recombination and decoupling, we can extrapolate backwards in time using physics established in the laboratory. The furthest back we can reach without running into the problem of unknown physics is about 10^{-12} seconds after the big bang, corresponding to $kT \sim 10^{12} \text{eV} = 1$ TeV. At this time, and until $t \sim 10^{-6} \text{s}$, the universe would have consisted of one family of UR quarks and leptons, with a few particles from the other two families. At around $t \sim 10^{-6} \text{s}$ the quarks condense into hadrons, consisting of mesons and baryons, of which only the proton is stable (although the neutron is nearly stable and hence still relevant). At around $t \sim 10^{-5} \text{s}$ the temperature has fallen to 10^{12} K. The density is already less than nuclear matter density and the nucleons are now non-relativistic; the Universe will now consist mainly of the following particles:

- UR: $\gamma, e^\pm, \nu_e, \bar{\nu}_e$
- NR: p, n

The electrons and positrons will remain ultra-relativistic as long as $kT \gg 2m_e c^2$ ($T \gg 5 \times 10^9$ K).

For the UR electrons and positrons we have $E = pc$ and $g_s = 2$ (as for photons). In addition we have $\mu = 0$, again as for photons, but for a different reason. Note first that for electrons and positrons in thermal equilibrium with photons, the electromagnetic reaction

$$e^+ + e^- \leftrightarrow \gamma \quad (4.55)$$

implies that $\mu_{e^+} = -\mu_{e^-}$ because $\mu_\gamma = 0$. If μ_{e^-} is negative then μ_{e^+} must be positive. This will imply an imbalance in the number densities, with $n_{e^-} > n_{e^+}$. We know that there *is* an imbalance because at lower temperatures, when all the positrons have annihilated with electrons, there must be a residual electron number density of ηn_γ , but this imbalance is negligible when $kT \gg 2m_e c^2$ because electron-positron pairs can then be easily produced by energetic photons, and their number densities will be comparable to n_γ . Setting $\mu = 0$ we have

$$n_{e^-} = \left(\frac{8\pi(kT)^3}{h^3 c^3} \right) \int_0^\infty \frac{x^2 dx}{e^x + 1} = \frac{7}{8} n_\gamma. \quad (4.56)$$

The factor of 7/8 relative to photons is because the electron is a fermion. The same calculation applies to positrons.

Now consider the Weak interaction

$$e^- + e^- \leftrightarrow \nu_e + \bar{\nu}_e. \quad (4.57)$$

We have just argued that the chemical potentials of the electrons and positrons vanish, so $\mu_\nu + \mu_{\bar{\nu}} = 0$. In this case we can't be sure that any imbalance of neutrinos and anti-neutrinos will be negligible at the temperatures we are considering, but it is a reasonable assumption. It then follows that

$$\mu_\nu = \mu_{\bar{\nu}} = 0. \quad (4.58)$$

For neutrinos, $g_s = 1$, however, so

$$n_\nu = n_{\bar{\nu}} = \frac{7}{16} n_\gamma. \quad (4.59)$$

Reactions such as (4.57) keep the neutrinos in chemical and thermal equilibrium with everything else as long as the reaction rate is much faster than the rate of expansion of the universe. But **weak interaction rates are strongly energy dependent** and at some characteristic energy the rate drops rapidly below the expansion rate. This causes a 'freeze out' of neutrinos, which then continue to cool according to $T \propto 1/a$ but out of thermal and chemical equilibrium with other particle species. This occurs well above the temperature ($kT \sim m_e c^2$) at which the positrons start to disappear due to annihilation with electrons, a process that 'reheats' the photons (actually it causes them to cool slightly less rapidly). As this reheating has no effect on the neutrinos the neutrino background should now be at a lower temperature than the CMBR. This temperature is calculable (Q.III.6) and constitutes a prediction of the Hot Big Bang theory, albeit one that is unlikely to be tested soon.

4.10. Primordial nucleosynthesis

Now we turn to the NR nucleons, labelled by $N = (n, p)$. From the formula (2.75) we learn that

$$\mu_N = m_N c^2 - kT \log \left(\frac{2n_Q^{(N)}}{n_N} \right) \quad \left(n_Q^{(N)} = \frac{(2\pi m_N kT)^{\frac{3}{2}}}{h^3} \right). \quad (4.60)$$

It follows that

$$\mu_n - \mu_p = (m_n - m_p)c^2 - kT \log \left[\left(\frac{m_n}{m_p} \right)^{\frac{3}{2}} \left(\frac{n_p}{n_n} \right) \right]. \quad (4.61)$$

In principle, the Weak reaction

$$p + e^- \leftrightarrow n + \bar{\nu}_e \quad (4.62)$$

can maintain the neutron and protons in chemical equilibrium with $\mu_n = \mu_p$ (since $\mu_e^- = \mu_\nu = 0$). We will later need to re-examine this point, but if we proceed on the assumption of chemical equilibrium then we may use (4.61) with $\mu_n = \mu_p$ to deduce that

$$\left(\frac{n_n}{n_p} \right) = \left(\frac{m_n}{m_p} \right)^{\frac{3}{2}} \exp \left[-\frac{(m_n - m_p)c^2}{kT} \right]. \quad (4.63)$$

Now, $(m_n - m_p)c^2 \approx 1.3\text{MeV}$, which corresponds to a temperature of about 10^{10}K . So

$$n_n = \begin{cases} n_p & T \gg 10^{10} \text{ K} \\ 0 & T \ll 10^{10} \text{ K} \end{cases} \quad (4.64)$$

and the neutron number density will, apparently, drop rapidly to zero around $T = 10^{10}\text{K}$.

This conclusion neglects the opportunities provided by other reactions to convert neutrons and protons into stable alpha particles (helium nuclei) before the neutron population has fallen to zero. There is a chain of reactions that can accomplish this, and it starts with the creation of the (stable) deuteron by the reaction

$$p + n \rightarrow d + \gamma. \quad (4.65)$$

Deuterons that are created in this way will be rapidly processed into alpha particles by the reactions

$$d + d \rightarrow {}^3\text{He} + n, \quad {}^3\text{He} + d \rightarrow {}^4\text{He} + p. \quad (4.66)$$

The helium nuclei will eventually end up as helium atoms. This is called **primordial nucleosynthesis**. As the universe is constantly cooling, the conditions for further nucleosynthesis of yet heavier elements never prevail, although trace amounts of Lithium will be synthesised. The relative abundance of helium depends crucially on the time or, equivalently, the temperature at which nucleosynthesis gets under way. To see this, let n_n and n_p be the nucleon number densities at the time of nucleosynthesis. The final helium number density will be $(1/2)n_n$ (since it takes two neutrons to make a ${}^4\text{He}$ nucleus), but $m_{\text{He}} \approx 4m_n$, so the total mass density ending up in helium is

$$\rho_{\text{He}} \approx \left(\frac{1}{2}n_n\right) \times 4m_n = 2n_n m_n. \quad (4.67)$$

This can be compared to the total mass density

$$\rho = n_n m_n + n_p m_p \approx (n_n + n_p)m_n \quad (4.68)$$

The fraction of the total mass ending up in helium is therefore

$$Y_4 \equiv \frac{\rho_{\text{He}}}{\rho} = \frac{2(n_n/n_p)}{1 + (n_n/n_p)}. \quad (4.69)$$

This fraction depends on the ratio (n_n/n_p) at the time of nucleosynthesis. If nucleosynthesis were to occur ‘too early’, when $T \gg 10^{10}\text{K}$, then $(n_n/n_p) = 1$ and so $Y_4 = 1$: a helium universe with no hydrogen and hence no stars like our Sun. If nucleosynthesis were to occur ‘too late’, when $T \ll 10^{10}\text{K}$, then we would have $(n_n/n_p) = 0$ and hence $Y_4 = 0$: no helium, in disagreement with observations.

Although deuterons are created by the reaction (4.65) they can also be destroyed by the reverse *photo-disintegration* reaction

$$d + \gamma \rightarrow p + n. \quad (4.70)$$

This requires an energy input equal to the deuteron binding energy, which is about 2.2 MeV. For $kT \gg 2.2$ MeV any deuteron that forms will immediately photo-disintegrate, and deuteron collisions will never occur. Thus, nucleosynthesis can only get started when the temperature has dropped to the point at which there are few photons with energies exceeding 2.2 MeV. Setting $kT = 2.2$ MeV provides a naive estimate of this temperature, and since this is about twice the temperature at which the neutrons start to disappear (at $kT \sim 1.3$ MeV) one might expect nucleosynthesis to start at a temperature above 10^{10} K when $n_n/n_p \approx 1$, yielding a helium universe. However, the naive estimate is wrong because it doesn't take into account the large photon to baryon ratio. Because there are so many photons there can be sufficient photons with energies exceeding 2.2 MeV in the 'exponential tail' of the Planck distribution. We have already seen how to take this into account in our calculation of the temperature at recombination: the actual temperature is lower than the naive estimate by a factor of $(1/\log \eta^{-1}) \sim 1/20$. An improved estimate of the temperature at nucleosynthesis is therefore obtained by setting $kT = 0.1$ MeV. This corresponds to a temperature well below 10^{10} K, at which $(n_n/n_p) \approx 0$, and hence almost no helium, again in conflict with observations.

However, we have still to take into account the fact that Weak interaction rates are strongly energy dependent. When the rate of the Weak reaction (4.62) falls below the rate of expansion of the universe there will be a 'freeze out' of neutrons. The forward reaction of (4.62) requires energy $(m_n - m_p - m_e)c^2 \approx (1.3 - 0.5) \text{ MeV} = 0.8 \text{ MeV}$. This constitutes the 'characteristic energy' for this process; the reaction rate falls rapidly below the expansion rate when kT becomes less than 0.8 MeV, and the neutron-proton ratio is 'frozen in' at this temperature, at which time

$$\frac{n_n}{n_p} \approx e^{-\frac{1.3}{0.8}} \approx \frac{1}{5}. \quad (4.71)$$

However, neutrons can still undergo beta-decay in the few minutes (the 'first three minutes') that the Universe takes to cool from $kT \approx 0.8$ MeV to the temperature at which nucleosynthesis 'cooks' any surviving neutrons into helium nuclei. Calculations show that this reduces the neutron to proton ratio to

$$\frac{n_n}{n_p} \approx \frac{1}{7}, \quad (4.72)$$

which yields $Y_4 \approx 1/4$ in good agreement with observations. [The fraction of helium in stars decreases as one looks deeper in the sky and hence further back in time, levelling

off at about 1/4. The increase in helium abundance over time is due to stellar nucleosynthesis.] Detailed calculations of the primordial abundances of both Helium and Lithium are in good agreement with observations; this is considered the best evidence we have for the correctness of the Hot Big Bang model of cosmology.

5. Example sheets

5.1. Example Sheet 1

- **1.** In our region of the galaxy, stars are typically separated from their nearest neighbours by a distance of about one parsec (pc). Relative velocities of stars within the galaxy typically have magnitudes of around 10^5 m/s. Use these figures to show that the typical angular speed ω at which one star moves across the sky relative to another one is subject to the approximate upper bound

$$\omega \leq 10^{-4} \text{ rad/yr.}$$

$$[1 \text{ yr} \sim 3 \times 10^7 \text{ s; } 1\text{pc} \sim 3 \times 10^{16} \text{ m.}]$$

- **2.** In a simple cosmological model, an average cosmic mass density ρ is concentrated in an infinite number of stars, of mass M , radius R and luminosity L . The stars are evenly, but randomly, distributed throughout an infinite universe, so that any line of sight in the night sky must meet some star. Let d be the average distance to a star along a line of sight from the earth. Explain why

$$d \sim \frac{M}{\rho R^2}.$$

Given that the energy flux per unit area of a single star at a distance r is $\Phi = L/4\pi r^2$, show that the model predicts a total energy flux per unit area on earth of $\Phi \sim L/R^2$. Show further that this equals the energy flux per unit area that the earth would receive if each point on the sky were as bright as the nearest star, i.e. the sun. So why is the sky dark at night? This is ‘Olbers’ paradox’.

In our universe, the average cosmic number density of hydrogen is approximately 1 atom per cubic metre. Use this to show that

$$\frac{(d/c)}{H_0^{-1}} \sim 10^{13}.$$

How does this help to resolve the paradox? [$M_\odot \sim 2 \times 10^{30} \text{ kg}$, $R_\odot \sim 7 \times 10^8 \text{ m}$, $m_H \sim 1.7 \times 10^{-27} \text{ kg}$, $cH_0^{-1} \sim 10^{26} \text{ m}$.]

- **3.** The ‘deceleration parameter’ q_0 is defined as $q(t_0)$, where

$$q(t) = -\frac{a\ddot{a}}{\dot{a}^2}.$$

Show that $q = [(3\gamma/2) - 1]\Omega$ for a universe with equation of state $P = (\gamma - 1)\rho c^2$. Hence show that $q_0 = \frac{1}{2}\Omega_0$ for a pressure-free universe.

- **4.** Consider an empty universe, with $\rho = 0$. Find the general solution of the Raychaudhuri equation, and then show that a solution with non-constant scale factor $a(t)$ solves the Friedmann equation only if $k < 0$. You have just found the Milne universe. Show that the age of the Milne universe equals the Hubble time H_0^{-1} . Show further that the general universe with this property has $\rho \propto t^{-2}$ and an equation of state $P = -\frac{1}{3}\rho c^2$.

- **5.** Show that the equation of state $P = -\rho c^2$ implies a *constant* mass density ρ , which we may write as

$$\rho = \left(\frac{c^2}{8\pi G} \right) \Lambda$$

where Λ is a ‘cosmological’ constant, with units of inverse length squared. Show that the acceleration equation now has the *de Sitter universe* solution

$$a(t) = a_0 e^{Ht}, \quad H = c\sqrt{\Lambda/3}.$$

What is the value of the parameter k for this solution? Show that

$$\int_{-\infty}^t \frac{dt'}{a(t')} = \infty$$

and hence that the de Sitter universe has no cosmological horizon. [Show, however, that the integral

$$\int_0^\infty \frac{dt'}{a(t')}$$

is *finite* and hence deduce that there is a *maximum* comoving distance that a signal emitted at time $t = 0$ can travel from its source. It follows from this that there are events in a de Sitter universe that an observer can never see; they are said to be behind a (cosmological) *event horizon*. This is a different kind of horizon to the cosmological horizon of decelerating model universes discussed in the lectures. Show that any model universe with $a \propto t^\alpha$ has no event horizon as long as $\alpha < 1$; in other words, as long as it is decelerating.]

- **6.** A homogeneous and isotropic model universe has pressure $P(t)$ and energy density $E/V = \rho(t)c^2$ such that

$$P = (\gamma - 1)\rho c^2$$

where γ is a constant. Assuming that the universe is expanding adiabatically, such that $dE = -PdV$, show that $\rho = \rho_0 a^{-3\gamma}$ for constant ρ_0 , where $a(t)$ is the scale factor of the universe. Let

$$\eta(t) = \int^t \frac{dt'}{a(t')}$$

be a new time parameter, and define the new function $y(\eta)$ by

$$y = a^{(3\gamma-2)/2}.$$

Show that the Friedmann equation for $a(t)$ implies that $y(\eta)$ satisfies

$$y'' + kc^2 \left(\frac{3}{2}\gamma - 1 \right)^2 y = 0.$$

Hence show that for a radiation-dominated universe ($\gamma = 4/3$) with $k = 1$ the graph of $a(t)$ against t is a semi-circle. Find the total time duration of this universe, from big bang to big crunch, as a function of ρ_0 .

- **7.** The apparent angular size $\delta\theta$ of a galaxy of proper size ℓ located at a comoving distance x is

$$\delta\theta = \frac{\ell}{a(t_e)x}$$

where t_e is the time of emission from the galaxy of the light that we see now. Taking $a(t_0) = 1$, show that $x = 3ct_0[1 - (t_e/t_0)^{1/3}]$ for an Einstein-de Sitter universe, and hence that

$$\delta\theta = \frac{\ell}{2cH_0^{-1}} \frac{(1+z)}{[1 - (1+z)^{-\frac{1}{2}}]}.$$

Sketch the graph of $\delta\theta$ against z and show that there is a minimum at $z = 1.25$.

5.2. Example Sheet 2

- **1.** N equal mass particles of total energy E populate a set of degenerate energy eigenstates with energies E_i and degeneracies g_i ($i = 1, 2, 3, \dots, \infty$). The set $\{n\}$ of numbers n_i of particles with energy E_i is assigned a weight of the form

$$\Omega(\{n\}) = \prod_i W(n_i, g_i). \quad (*)$$

The most probable distribution $\{\bar{n}\}$ is obtained by maximising $\log \Omega$ subject to the constraints of fixed particle number N and fixed total energy E . Show that \bar{n}_i is found by solving the equation

$$\frac{\partial \log W(n_i, g_i)}{\partial n_i} = \alpha + \beta E_i$$

where α and β are constants such that $\sum_i \bar{n}_i = N$ and $\sum_i \bar{n}_i E_i = E$. Write out this equation for each of the following three choices of the function W :

$$(i) W(n, g) = \frac{(g + n - 1)!}{n!(g - 1)!}, \quad (ii) W(n, g) = \frac{g!}{n!(g - n)!}, \quad (iii) W(n, g) = \frac{g^n}{n!}.$$

Assuming $g \gg 1$, $n \gg 1$, and $g \geq n$, use Stirling's formula [$\log n! = n \log n - n + \mathcal{O}(\log n)$] to simplify your result. Hence show that if α and β are appropriately related to the chemical potential μ and temperature T then $\{\bar{n}\}$ is the equilibrium distribution found in the lectures for a gas of (i) Bose-Einstein, (ii) Fermi-Dirac, and (iii) Maxwell-Boltzmann type. [The gas particles are said to obey BE, FD or MB 'statistics', respectively.]

- **2.** Assuming that $\Omega(\{n\})$ of the previous question equals the number of microstates available to the N particles for a given occupation number distribution $\{n\}$, explain why $\Omega(\{n\})$ must take the form $(*)$ if the N particles are identical.

Show that Ω is equal to the number of available microstates in cases (i) and (ii) assuming Bose-Einstein statistics and Fermi-Dirac statistics, respectively. [**Hint:** Consider how many different ways there are of painting n identical balls in g colours assuming (i) no restriction on the number of times each colour is used or (ii) that no colour may be used more than once.]

Show that in case (iii) Ω is $1/N!$ times the number of microstates available to N *distinguishable* particles. [This fact is related to the 'Gibbs paradox' of classical statistical mechanics.]

- **3.** The Helmholtz free energy F is defined by $F = U - TS$. Use the first law of thermodynamics to show that $dF = -SdT - PdV + \mu dN$. Regarding F as a function $F(T, V, N)$, and using the equality of mixed partial derivatives, derive the 'Maxwell relations'

$$\left(\frac{\partial P}{\partial T}\right)_{V,N} = \left(\frac{\partial S}{\partial V}\right)_{T,N} \quad \left(\frac{\partial \mu}{\partial T}\right)_{N,V} = -\left(\frac{\partial S}{\partial N}\right)_{T,V} \quad \left(\frac{\partial \mu}{\partial V}\right)_{N,T} = -\left(\frac{\partial P}{\partial N}\right)_{V,T}.$$

- **4.** Let r be the radial distance from the centre of a spherically symmetric star of pressure $P(r)$, and let $m(r)$ be the mass within a sphere of radius r . Use the pressure-support equations to show that the function

$$F(r) = P + \frac{Gm^2}{8\pi r^4}$$

is a decreasing function of r . Let M be the mass of the star and R its radius. Derive the lower bound

$$P_c > \frac{GM^2}{8\pi R^4}$$

on the central pressure P_c .

- **5.** A star is assumed to be a spherically-symmetric ball of ideal gas held together by gravity. Assuming that the number density $n(r)$, pressure $P(r)$ and temperature $T(r)$ are functions only of radial distance r from the centre, use the ideal gas law (Boyle-Charles law) to show that their gradients n' , P' and T' are related by

$$\frac{n'}{n} = \frac{P'}{P} - \frac{T'}{T}.$$

- **6.** The formation of a neutron star involves the removal of electrons, and hence their degeneracy pressure, by inverse beta-decay, $p + e^- \rightarrow n + \nu_e$. Why are white dwarf stars stable against inverse beta-decay? [You will need to use the fact that $m_n - m_p \approx 2.6m_e$.]
- **7.** Starting from the Fermi-Dirac distribution, obtain the relation

$$n = \frac{4\pi g_s}{3c^3 h^3} (\mu^2 - m^2 c^4)^{\frac{3}{2}}$$

for the number density of fully degenerate relativistic fermions of mass m and spin degeneracy g_s at chemical potential μ .

At densities much higher than those available in white dwarfs, inverse beta decay allows a star composed of protons and electrons to ‘neutronize’, i.e. to turn into neutrons with the neutrinos escaping from the star. The equilibrium concentration of protons and neutrons in the star is determined by the equation.

$$\mu_p + \mu_e = \mu_n.$$

Why is it reasonable to suppose that the neutrinos have zero chemical potential? Assuming that the nucleons (protons and neutrons) are non-relativistic, and that the electrons are ultra-relativistic, show that the fraction $\alpha = n_p/n_N$, where $n_N = n_p + n_n$ is the nucleon number density, satisfies

$$\alpha^{\frac{2}{3}} - \left(\frac{m_p}{m_n}\right) (1 - \alpha)^{\frac{2}{3}} = \frac{2}{\lambda_p (3\pi^2 n_N)^{\frac{1}{3}}} \left[\frac{(m_n - m_p)}{m_p \lambda_p (3\pi^2 n_N)^{\frac{1}{3}}} - \alpha^{\frac{1}{3}} \right]$$

where $\lambda_p = \frac{h}{m_p c}$ is the proton’s Compton wavelength. [This equation determines α as a function of n_N . A typical nucleon number density is $n_N \sim 10^{44} m^{-3}$, which yields $\alpha \approx 1/200$. Thus, ‘neutron’ stars are indeed composed mostly of neutrons.]

- **8.** Why are GR effects significant for neutron stars but not for white dwarfs? [You will need to consider the dimensionless quantity $GM/c^2 R$.]

5.3. Example Sheet 3

- **1.** A galaxy of constant luminosity L has redshift z , as seen from Earth. Show that the rate at which its radiant energy passes through a sphere that intercepts Earth, and is centred on the galaxy, is $L/(1+z)^2$. [This is its ‘apparent luminosity’.]
- **2.** A thermal (Planckian) cosmic radiation background is assumed to be isotropic with temperature T in an inertial frame S . The same radiation is detected in another (laboratory) inertial frame S' moving with velocity \mathbf{v} with respect to S . The Lorentz transformation relating the energy-momentum 4-vector in the two frames is

$$E = \gamma (E' - \mathbf{v} \cdot \mathbf{p}') , \quad \mathbf{p} = \gamma (\mathbf{p}' - \mathbf{v} E' / c^2) ,$$

where $\gamma = \sqrt{1 - v^2/c^2}$. Use this to show that the background will still be thermal in S' but with an anisotropic temperature

$$T'(\theta') = \frac{T}{\gamma [1 - (v/c) \cos \theta']} = T \left[1 + \frac{v}{c} \cos \theta' \right] + \mathcal{O}(v^2/c^2) ,$$

where θ' is the angle between the velocity \mathbf{v} of the frame S' and the momentum \mathbf{p}' of the photon arriving at the detector. Given that T'_+ and T'_- are the maximum and minimum temperatures seen in the inertial frame S' , show that

$$(i) \ T = \sqrt{T'_+ T'_-} , \quad (ii) \ \frac{v}{c} = \frac{T'_+ - T'_-}{T'_+ + T'_-} .$$

What is the significance of these results to observations of the CMBR?

It is believed that there is a thermal cosmic electron-neutrino background that is isotropic in the same ‘isotropy frame’ S as the CMBR, but with a slightly lower temperature. The above analysis will still apply if the neutrino is massless, but the neutrino may have a small mass; given that it does, show that the energy density in the cosmic neutrino background will not be thermal, even at fixed angle, when measured in any inertial frame S' with non-zero velocity relative to S .

- **3.** Let the internal energy U of a gas be related to its pressure P and volume V by the formula $PV = (\gamma - 1)U$. Assuming either fixed particle number or vanishing chemical potential, use the first law of thermodynamics to show that

$$(\gamma - 1)TdS = \gamma PdV + VdP$$

Hence show that PV^γ is constant for an isentropic ($dS = 0$) change of state. Show also that if S is proportional to V , at fixed pressure, then $TS = \gamma U$. [The constant γ is called the ‘adiabatic index’ because a ‘quasi-static’, i.e. slow, adiabatic change of state is isentropic.]

How are these results applicable to the CMBR? Use them, and the Stephan-Boltzmann law for blackbody radiation, to show that $s \propto T^3$ where s is the entropy density of the CMBR.

- **4.** Let $a(t)$ be the scale factor of an expanding universe. Assuming that the expansion is too slow to cause transitions between energy eigenstates with different energy, show that a particle of momentum p_0 at time t_0 will have momentum $p = p_0/a(t)$ at time t .

Use this to show that a thermal distribution of photons with temperature T_0 at time $t = t_0$ will still be thermal at time t , but with a temperature $T(t) = T_0/a(t)$. Hence show, using the result for the entropy density of the previous question, that the total entropy of the CMBR is conserved during the expansion.

Show that a thermal distribution of particles of a *non-relativistic* ideal gas will also remain thermal but with a temperature $T = T_0/a^2(t)$. Assuming that $PV = (\gamma - 1)U$ with $\gamma \neq 1$, and using the results of the previous question, deduce that $\gamma = 5/3$.

- **5.** Neutral hydrogen atoms can be ionized by collisions with sufficiently energetic photons via the photo-ionization reaction $\gamma + H \rightarrow e^- + p^+$. For simplicity we assume that only the ground state of the hydrogen atom is relevant, so that the minimum energy that the photon must have to ionize the atom is the ground-state binding energy B . The reverse reaction is called ‘recombination’ and at equilibrium the forward and reverse reactions balance. Let n_H , n_e and n_p be the equilibrium number densities of hydrogen atoms, electrons and protons, respectively. In equilibrium the chemical potentials must balance. Since $\mu_\gamma = 0$ this requires

$$\mu_H = \mu_e + \mu_p.$$

Assuming charge neutrality, and that all particles other than the photons are non-relativistic, show that the equilibrium electron number density is given by the ‘Saha’ equation:

$$n_e^2 \approx n_H \left(\frac{2\pi m_e kT}{h^2} \right)^{\frac{3}{2}} e^{-B/kT}.$$

[N.B. A NR approx is adequate since $B \approx 13.6 \text{ eV} \ll m_e c^2$.]

- **6.** Because of electron-positron pair creation and annihilation $\gamma \leftrightarrow e^- + e^+$, photons in the cosmic radiation background will be in thermal equilibrium with electrons and positrons at some temperature T . For $kT \ll m_e c^2$ the number densities of electrons and positrons is negligible, but for $kT \gg m_e c^2$ their number densities are approximately those of an ideal Fermi gas of massless particles at zero chemical

potential and temperature T . Discounting differences in spin degeneracies, a ideal Fermi gas of this type has an energy density that is $7/8$ times that of a Bose gas of massless particles at the same temperature and chemical potential. Why must this number be less than unity? Show that the combined energy density of photons, electrons and positrons for $kT \gg m_e c^2$ is $g_{eff} a_{RD} T^4$ where a_{RD} is the radiation density constant and g_{eff} an effective spin degeneracy factor which you should compute.

As the universe expands it cools adiabatically from a temperature $T \gg m_e c^2/k$ to a temperature $T \ll m_e c^2/k$. By equating the entropy of radiation at the higher and lower temperatures show that the later temperature of the radiation is a higher by a factor of $(11/4)^{1/3}$ than it would have been had the annihilation of electrons and protons not occurred. How is this fact relevant to the cosmic electron-neutrino background?

6. Answers to Example sheets

6.1. Answers to Example Sheet 1

- **1.** $\omega = v/r$, $v \sim 10^5 m/s$, $r \geq 3 \times 10^{16} m$, so $\omega \leq \frac{1}{3} \times 10^{-11} rad \sim 10^{-4} rad/yr$.
- **2.** Number density of stars is ρ/M . Visible volume $\sim d^3$, so number of visible stars is $N \sim \frac{\rho}{M} \times d^3$. But each star subtends solid angle $\delta\Omega \sim R^2/d^2$, so sky is filled by $N \sim d^2/R^2$ stars. Equating two estimates for N yields $d \sim M/(\rho R^2)$.

Average energy flux per unit area received from one star is $\Phi \sim L/d^2$, but there are $\sim d^2/R^2$ stars, so total flux per unit area is $\Phi \sim L/R^2$.

Energy flux per unit area from the nearest star is $L/4\pi r_{min}^2$ where r_{min} is the distance to this star. This star subtends solid angle $\delta\Omega = \pi R^2/r_{min}^2$, while the whole night sky subtends solid angle 2π . A night sky everywhere as bright as this star therefore receives $2r_{min}^2/R^2$ times the flux that it would from this one star, so $\Phi = L/2\pi R^2$. This is the energy flux radiated through unit area of the sun's disc at its surface (which is why it is independent of r_{min}).

Most of the light making the sky bright was emitted at a time d/c ago. Since $\rho \sim 1.7 \times 10^{-27} kg/m^3$ we have $d \sim 10^{39} m$, so $(d/c) \sim 10^{13} \times H_0^{-1}$, which is vastly greater than the age of the universe. The universe is not old enough for light from these stars to have had time to reach us. [This solution was first arrived at by Edgar Allan Poe; Lord Kelvin fleshed out the mathematics:

How distant some of the nocturnal suns!

So distant, says the sage, 'twere not absurd

To doubt that beams set out at Nature's birth

Had yet arrived at this so foreign world.]

- **3.** From Raychaudhuri,

$$q = \frac{4\pi G}{3H^2} (\rho + 3P/c^2) = \frac{8\pi G}{3H^2} \left(\frac{3}{2}\gamma - 1 \right) \rho = \left(\frac{3}{2}\gamma - 1 \right) \Omega.$$

Set $\gamma = 1$ for dust.

- **4.** $\rho = 0$ implies $P = 0$, so excluding $\dot{a} = 0$, we can use Raychaudhuri to learn that $\ddot{a} = 0$. General solution with $a(0) = 0$ and $a > 0$ for $t > 0$ is $a = \alpha t$ for constant $\alpha > 0$. Friedmann implies that $\alpha^2 = -kc^2$, which is possible only if $k < 0$. We then have the $k < 0$ Milne universe $a(t) = \sqrt{-k} ct$. This implies that $H(t) = 1/t$, so the age of the Milne universe is $t_0 = H_0^{-1}$.

If $H(t) = 1/t$ then $a(t) = \alpha t$. It follows that $\ddot{a} = 0$, so Raychaudhuri implies that $3P + \rho c^2 = 0$. The fluid equation is then $\dot{\rho} = -2\rho H = -2\rho/t$, which implies $\rho = A/t^2$ for constant A . The Milne model is the $A = 0$ case.

- **5.** When $P = -\rho c^2$ the fluid equation reduces to $\dot{\rho} = 0$. The acceleration equation yields $\ddot{a} = H^2 a$, which has the de Sitter solution given with $\dot{a} = Ha$. The Friedman equation is then solved if $k = 0$. The earliest time is now $t = -\infty$ and a light signal emitted at this time will have travelled, at time t , a comoving distance of c times the integral given. As this is infinite, there is no cosmological horizon [In such a universe the solution to Olbers' paradox of Q.2 would not apply, but the sky would still be dark at night because of the extreme redshift of distant galaxies caused by the accelerating expansion].
- **6.** Since $E = \rho c^2 V$, and $P = (\gamma - 1)\rho c^2$ the relation $dE = -PdV$ is equivalent to $Vd\rho + \gamma\rho dV = 0$. But $V \propto a^3$, so $d \log(\rho a^{3\gamma}) = 0$, which yields $\rho = \rho_0 a^{-3\gamma}$.

Now, $d/dt = \dot{\eta} d/d\eta = a^{-1} d/d\eta$, so Friedmann is equivalent to

$$\frac{1}{a^2} \left(\frac{da}{d\eta} \right)^2 - \frac{8\pi G \rho_0}{3} a^{2-3\gamma} = -kc^2$$

Assuming $3\gamma \neq 2$, we have $a^{-1} da/d\eta = [2/(3\gamma - 2)] y^{-1} dy/d\eta$, so

$$\left(\frac{dy}{d\eta} \right)^2 + \frac{k}{4} (3\gamma - 2)^2 c^2 y^2 = (3\gamma - 2)^2 \frac{2\pi G \rho_0}{3}.$$

Differentiate to get

$$2y' \left[y'' + kc^2 \left(\frac{3}{2}\gamma - 1 \right)^2 y \right] = 0$$

Excluding $y' = 0$, which implies $\dot{a} = 0$, we find the desired result. We assumed that $3\gamma \neq 2$ but if $3\gamma = 2$ then $\ddot{a} = 0$, which was analysed in Q.I.4.

If $\gamma = 4/3$ then $y = a$. With $kc^2 = 1$ we then have $d^2a/d\eta^2 + a = 0$. Now $d/d\eta = ad/dt$ so $d[a(da/dt)]/dt = -1$. It follows that $a\dot{a} = -(t - \alpha)$ for constant α . Integrating subject to $a(0) = 0$ we find that $a^2 + (t - \alpha)^2 = \alpha^2$, i.e. a semi-circle with diameter 2α , which is the total duration of the universe. Differentiate to get $\dot{a} = (\alpha - t)/a$ and substitute this into Friedmann to find that $2\alpha = 4\sqrt{2\pi G\rho_0}$.

- **7.** The distance $a(t)dx$ travelled by light equals cdt so

$$x = c \int_{t_e}^{t_0} \frac{dt}{a(t)} = 3ct_0^{\frac{2}{3}} \left(t_0^{\frac{1}{3}} - t_e^{\frac{1}{3}} \right)$$

where we use the EDS scale factor $a(t) = (t/t_0)^{2/3}$ in the second step. Since $3t_0 = 2H_0^{-1}$ for EDS, this is the same as

$$x = 2cH_0^{-1} [1 - (t_e/t_0)^{1/3}].$$

Now, for $a(t_0) = 1$,

$$(t_e/t_0)^{1/3} = [(t_0/t_e)^{2/3}]^{-\frac{1}{2}} = [a(t_0)/a(t_e)]^{-\frac{1}{2}} = (1+z)^{-\frac{1}{2}}$$

Since $3t_0 = 2H_0^{-1}$ for EDS, this yields the result for x . For $a(t_0) = 1$ we have $a(t_e)x = (1+z)^{-1}x$ so $\delta\theta = l(1+z)/(2cH_0^{-1}x)$, and hence the result claimed.

6.2. Answers to Example Sheet 2

- **1.** Maximise $\log \Omega - \alpha(\sum_i n_i - N) - \beta(\sum_i n_i E_i - E)$ without constraint. Variation with respect to n_i yields equation given. Solution is $\bar{n}(\alpha, \beta)$. The constraints determine α and β in terms of N and E . Set $\alpha = \beta\mu$ and $\beta = 1/kT$.
 - (i) $\log W_i(n_i) = (n_i + g_i - 1) \log(g_i + n_i - 1) - n_i \log n_i + \text{const.}$
This yields $\bar{n}_i = [1 - (1/g_i)]\bar{n}_i(BE) \approx \bar{n}_i(BE)$ since $g_i \gg 1$.
 - (ii) $\log W_i(n_i) = -n_i \log n_i - (g_i - n_i) \log(g_i - n_i) + \text{const.}$ This yields $\bar{n}_i = \bar{n}_i(FD)$.
 - (iii) $\log W_i(n_i) = n_i \log g_i - n_i \log n_i + n_i$. This yields $\bar{n}_i = \bar{n}_i(MB)$.
- **2.** If the particles are identical then it makes no difference which n_i particles we choose to put in the i th energy level. We have only to compute the numbers W_i of arrangements of particles within each level, and since an arrangement in one level can be chosen independently of the arrangements in the other levels, the total number of microstates is the product $\prod_i W_i$.

We now need to compute the number of ways that n identical particles can be distributed among g states. This is equivalent to computing the number of ways that n identical balls may be painted in g colours. With no restriction on the number of times each colour may be used we may proceed as follows. Place all N balls in a line. Now place $g - 1$ partitions between them, thus dividing them into g sets, each of which is to be painted with a different one of the g colours. The line now consists of $n + g - 1$ ‘objects’, n balls and $g - 1$ partitions. There are $(n + g - 1)!$ permutations of these objects but neither the $n!$ permutations of the balls nor the $(g - 1)!$ permutations of the partitions leads to a distinct arrangement. The total number of ways of painting the balls is therefore $(n + g - 1)!/n!(g - 1)!$. This equals $W(n)$ for case (i). If no colour may be used more than once then we have only to choose n of the g colours (this is possible because it is assumed that $g \geq n$). This may be done in $g!/[n!(g - n)!]$ ways. This equals $W(n)$ for case (ii).

If the particles are not identical then we must first decide how to partition them among the various energy levels. There are $N!/\prod_i n_i$ partitions since there are $N!$ permutations of the particles and permutations of the particles in each level do not yield distinct partitions. For each partition we have now to count the number of ways to put n particles into g levels. The first of the n particles can be put into any one of the g levels, as can the second particle and so on, leading to g^n combinations. The total number of microstates available to the N particles is therefore the number of partitions times $\prod_i g_i^{n_i}$, or $N!\prod_i W_i$ where W_i is the weight function of case (iii). This equals $N!$ times the MB weight function $\Omega(\{n\})$ of the previous question.

- **3.** $dF = dE - d(TS) = (dE - TdS) - SdT = -PdV + \mu dN - SdT$. Thus $(\partial F/\partial T)_{V,N} = -S$, $(\partial F/\partial V)_{T,N} = -P$ and $(\partial F/\partial N)_{V,T} = \mu$. It follows that

$$\frac{\partial S}{\partial V} = -\frac{\partial}{\partial V} \left(\frac{\partial F}{\partial T} \right) = -\frac{\partial}{\partial T} \left(\frac{\partial F}{\partial V} \right) = \frac{\partial P}{\partial T}$$

and similarly for the other cases.

- **4.** Using pressure support equations for P' and m' we find that

$$F' = -Gm^2/2\pi r^5 < 0,$$

so $F(0) < F(R)$. Now $m \propto r^3$ as $r \rightarrow 0$ so $F(0) = P_c$. Since $P(R) = 0$ and $m(R) = M$, we also have $F(R) = GM^2/8\pi R^4$, and hence the bound on P_c .

- **5.** For a spherically symmetric star composed of an ideal gas the ideal gas law is $P(r) = n(r)kT(r)$, so

$$\log n = \log P - \log T + \text{constant}$$

Differentiate w.r.t. r to get result.

- **6.** A white dwarf is supported by degenerate electrons so that almost all electrons have energy $< \epsilon_F$. The electrons are necessarily non-relativistic because they cannot otherwise support the star, so $\epsilon_F < m_e c^2$. These two facts imply that almost all electrons have kinetic energy $< m_e c^2$. But the minimal energy needed for inverse beta decay is $(m_n - m_p)c^2$ (the neutrino energy could be much less than this so it can be neglected). Since $(m_n - m_p)c^2 \sim 3m_e c^2$ there are almost no electrons with the requisite kinetic energy.
- **7.** For fully degenerate ideal fermion gas, $n = 4\pi g_s/3(p_F/h)^3$. In addition, $\mu^2 = \epsilon_F^2 = p_F^2 c^2 + m^2 c^4$, so $p_F c = \sqrt{(\mu^2 - m^2 c^4)}$. Hence result. Note that $g_s = 2$ for electrons and nucleons, so

$$p_F = \left(\frac{3n}{8\pi} \right)^{\frac{1}{3}} hc$$

Neutrino has $m = 0$ (or at least nearly so) so $\mu \propto n^{1/3}$, but the effective density is zero because the neutrinos escape into an infinite volume. So $\mu = 0$.

Since electrons are ultra-relativistic we have $\mu_e = p_F c$ and using above formula for p_F this becomes $\mu_e = (3n_e/8\pi)^{1/3} hc$. Since nucleons are non-relativistic, we have $\mu_N = m_N c^2 + p_F^2/2m_N$, where N stands for n or p . Using formula for p_F again, this becomes $\mu_N = (3n_N/8\pi)^{2/3} (h^2/2m_N)$. In equilibrium $\mu_n - \mu_p = \mu_e$, or

$$(3n_p/8\pi)^{\frac{1}{3}} hc + (3n_p/8\pi)^{\frac{2}{3}} \left(\frac{h^2}{2m_p} \right) - (3n_n/8\pi)^{\frac{2}{3}} \left(\frac{h^2}{2m_n} \right) = (m_n - m_p)c^2$$

Multiply by $2m_p(8\pi/3)^{2/3}/h^2 n_N^{2/3}$ to get result.

- **8.** GR effects are important only in strong gravitational fields for which $GM/c^2 R \sim 1$. Take M to be critical mass and R the corresponding radius. For a white dwarf $M \sim M_{PL}^3/m_p^2$, where $M_{PL} = \sqrt{\hbar c/G}$ (the ‘Planck mass’). The corresponding radius is $R \sim GM_{PL}/m_e m_p$, so $GM/c^2 R \sim m_e/m_p \ll 1$ and GR effects are unimportant. For neutron stars m_e is replaced by m_p , because it is supported by nucleon degeneracy pressure, and this implies $GM/c^2 R \sim 1$.

6.3. Answers to Example Sheet 3

- **1.** Rate at which radiant energy passes through a sphere that just encloses the galaxy is L , by definition. This energy is in the form of photons. The photons are redshifted when they reach the Earth, and this reduces the energy of each photon by a factor of $(1+z)^{-1}$. In addition, the photons are now travelling through a space that has been expanded by a factor of $(1+z)$ so their rate of arrival is correspondingly reduced. The apparent energy flux is therefore $L/(1+z)^2$.

2. We have $\mathbf{v} \cdot \mathbf{p}' = vp' \cos \theta'$. Now, $p^2 c^2 = E^2 - m^2 c^4$ for a particle of rest-mass m . So $p' = E'/c$ for $m = 0$. From the given Lorentz transformation it then follows that $E = A(v, \theta')E'$ with $A(v, \theta') = \gamma[1 - (v/c) \cos \theta']$.

Now consider photons in the frame S . Each photon is in some energy eigenstate, which we may take to be non-degenerate for simplicity. The average number of photons in the i th state is $\bar{n}_i = 1/[e^{E_i/kT} - 1]$. In S' this state has energy E'_i but it is still the i th state and the average number of particles in it, which we may call \bar{n}'_i , remains the same. That is,

$$\bar{n}'_i = \bar{n}_i = \frac{1}{e^{E_i/kT} - 1} = \frac{1}{e^{E'_i A/kT} - 1}.$$

In the last step we use the formula $E = AE'$. Thus,

$$\bar{n}'_i = \frac{1}{e^{E'_i/kT'} - 1}$$

where $T' = T/A(v, \theta')$, which is the result quoted.

In the case of massive particles we could still write $E = AE'$ but A would now depend on E' as well as on v and θ . The function $\bar{n}'_i(E'_i)$ is then no longer a thermal distribution function.

The formulae involving T_{\pm} suggest a simple way to extract the temperature of the CMBR and the velocity of the lab frame relative to it. They follow from

$$T'_{\pm} = T \sqrt{\frac{c \pm v}{c \mp v}}.$$

- **3.** Differentiate $(\gamma - 1)U = PV$, and use first law to get result for dS . It directly follows that PV^{γ} is constant when $dS = 0$. It also follows that

$$V \left(\frac{\partial S}{\partial V} \right)_P = \frac{\gamma PV}{(\gamma - 1)T} = \frac{\gamma U}{T}$$

But if S is proportional to V at fixed P then the left hand side must equal S . Thus $ST = \gamma U$.

For $\gamma = 4/3$ these results are applicable to a gas of photons since in this case (i) $\mu = 0$ and (ii) $P = (1/3)E = (1/3)U$. In particular $ST = \gamma E = \gamma \epsilon V$, or $S/V = \gamma \epsilon/T$. The SB law says that $\epsilon \propto T^4$, while $S/V = s$, so $s \propto T^3$.

- **4.** In a box of volume V , a given energy eigenstate has a momentum that is a multiple of $\hbar/V^{1/3}$. Since $V \sim a^3$, $p(t) = p_0/a(t)$.

Photons are massless so the energy of an eigenstate with momentum p is $E = pc$. Thus, the energy at time t of the an eigenstate with energy E_0 at time t_0 is $E(t) = E_0/a(t)$.

For a distribution that is thermal at time t_0 with temperature T_0 , the average occupation number of (say) the i th eigenstate has the form $\bar{n}_i = f(x_0)$ where $x_0 = p_0 c/kT_0$. This occupation number is time-independent, by hypothesis, so its value at time t is again $\bar{n}_i = f(x_0)$ but we must now write this in terms of $E(t)$. Since $x_0 = a(t)E/kT_0 = E/kT$, where $T = T_0/a(t)$, we have $\bar{n}_i = f(x)$ where $x = E/kT$. This is thermal with temperature $T = T_0/a$. Since $V \sim a^3$ and $T \sim 1/a$ we conclude that VT^3 is constant. But $s \propto T^3$ implies that $S \propto VT^3$, hence S is constant.

A similar analysis applies to non-relativistic particles. We can replace E by U (by shifting μ by mc^2), but this is *quadratic* in p so we now get $T = T_0/a^2$. For an ideal gas $PV \propto T$, so $PV \sim 1/a^2$ and $PV^\gamma \sim a^{3\gamma-5}$. But PV^γ is constant for an isentropic change with the given equation of state (from Q.3.3), so either $P = 0$, corresponding to $\gamma = 1$, or $\gamma = 5/3$.

- **5.** For non-relativistic particles we may use the formula $\mu = mc^2 - kT \log(g_s n_Q/n)$, where g_s is the spin degeneracy, $n_Q = (2\pi mkT)^{3/2}/h^3$ is the ‘quantum concentration’, and n the number density. Balancing the chemical potentials then yields

$$(m_H - m_e - m_p)c^2 = kT \log \left[\frac{g_s(H)}{g_s(e)g_s(p)} \frac{n_e n_p}{n_H} \frac{n_Q^{(H)}}{n_Q^{(e)} n_Q^{(p)}} \right]$$

The electron and proton spin degeneracies are $g_s(e) = g_s(p) = 2$. Because we assume the hydrogen atoms to be in their ground state the spin degeneracy of a hydrogen atom is $g_s(H) = g_s(e)g_s(p) = 4$, so the first factor in the logarithm equals 1. The left hand side equals minus the binding energy B , so

$$e^{-B/kT} = \frac{n_e n_p}{n_H} \frac{n_Q^{(H)}}{n_Q^{(e)} n_Q^{(p)}}$$

Charge neutrality implies $n_e = n_p$ while $n_Q^{(H)} \approx n_Q^{(p)}$ since the hydrogen mass is almost entirely due to the proton. This yields the Saha equation.

- **6.** The photon spin degeneracy is 2. The electron and positron spin degeneracies are also 2 but because they are fermions we must multiply by $7/8$. Thus the effective spin degeneracy at the higher temperature is

$$g_{eff} = 2 + \frac{7}{8} \times 2 + \frac{7}{8} \times 2 = 11/2$$

This is to be compared with a spin degeneracy of 2 for the photons alone at the lower temperature. Since $S/V \propto E/T$ we have $S/V \propto g_{eff}T^3$ with a constant of proportionality that is independent of temperature or number of species. Thus conservation of entropy implies that $(11/2)T_+^3V_+ = 2T_-^3V_-$ where $+$ means ‘before annihilation’ and $-$ means ‘after annihilation’. Equivalently,

$$\frac{T_-}{T_+} = \left(\frac{11}{4}\right)^{\frac{1}{3}} \frac{V_+}{V_-}.$$

In the absence of the ‘reheating’ due to electron-positron pair annihilation the right hand side would just be V_+/V_- .

Since the neutrino background is not ‘reheated’ its temperature should be $(4/11)^{\frac{1}{3}}$ times that of the CMBR.