NOTES ON PLANETARY PHYSICS

(Background Material for Physics 214*)

A.H. Bridle

Department of Physics

Queen's University

December , 1979

INTRODUCTION

There are very few textbooks on planetary physics and astronomy at a level between the many introductory first-year texts and the few specialist monographs that are aimed at graduate students and professional astronomers. One of the major reasons for this is the tremendous pace of development of the subject in the present era of planetary exploration by unmanned spacecraft. As a result, a reasonably detailed undergraduate text dealing with planetary astronomy would become out-of-date in a year or two, which discourages authors and publishers alike from producing one.

These notes are intended to bridge part of the gap between elementary texts and the research-level material in topics needed for background to Physics 214* - "Planets and Life". They cover material that is treated in the first half of the course, where we review the systematics of planetary astronomy before going on to consider the planets of our Solar System individually in detail. They aim to provide an introduction to the physics behind the study of planetary environments, and to define nomenclature which students are likely to encounter in the research literature. The index at the end of the notes gives references to most of the definitions of planetary terminology which can be found throughout them.

The notes are not a complete text for Physics 214*; readings from the research literature and from Scientific American form a vital part of the background material for the second half of the course. (These readings will be assigned in class and on hand-outs throughout the course.) Most of them will be found in the Physics Library; many will be available in Xerox copies in Binders kept for Physics 214* in the Physics Librarian's office. You should become acquainted with the Library arrangements early in the course in order to plan your reading for the assigned topics which carry a major fraction of the course marks.

Some parts of these notes contain derivations of formulae which will be referred to, but not proven, in class. Those marked with daggers (†) are more advanced, and may be read for their conclusions, rather than for their detailed method, by students who find their mathematical level too demanding. (You will not be examined on the derivations in the sections marked †.)

The book "The Solar System", by J.A. Wood (Prentice-Hall Foundations of Earth Science series, published in 1979) is also recommended reading for the course.

TABLE OF CONTENTS

1. Planetary Radar and the Solar System Distance Scale

The modern distance scale is based on radar ranging of the planetary distances, using pulsed radio transmitters operating in the frequency range 100 MHz to 10 GHz (i.e., at wavelengths from about 3 m to 3 cm). We use the fact that radio waves travel through the almost perfect vacuum of interplanetary space at the known velocity $c = 2.998 \times 10^5$ km/sec and reflect at well-defined boundaries, such as those between gaseous atmospheres and rocky surfaces. The basic measurement is very straightforward:
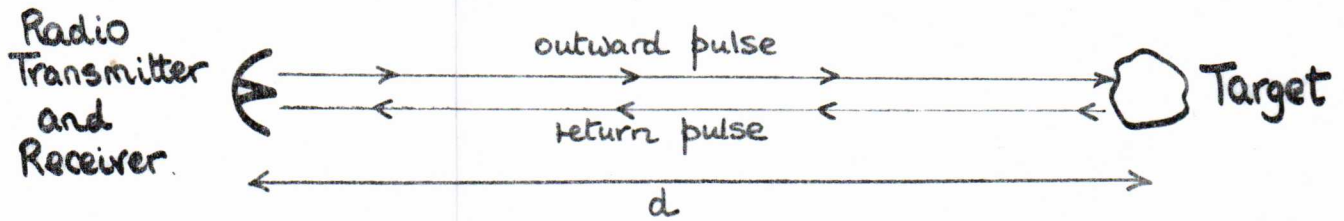
The total distance travelled by the pulse = 2 x the distance to the target

also $= c \cdot \Delta t$

where Δt is the time interval between transmission and rereception of the pulse. The unknown distance to the target is then

$$d = \frac{2.998 \times 10^5 \; \Delta t}{2} \quad km$$

if Δt is measured in **seconds**. In applying this technique to the determination of interplanetary distances, we run into two kinds of problems. First, the returned signals are very weak because of the enormous distances between planets. Second, not all objects in the Solar System are good reflective targets for our radar pulses.

1.1. The Basic Inverse-Square Law of Power

Suppose a radio transmitter emits a power P watts (joules/second) equally in all directions. We would call such a transmitter an ISOTROPIC radiator. Consider the energy E (joules) emitted by the transmitter in a very short time interval t; with these definitions $E = Pt$. At some time T after it was emitted, this energy will have travelled a distance $d = cT$ from the transmitter, where c is the velocity of radio waves (in vacuum the same velocity as that of light). So provided the velocity of radio waves is the same in all directions from the transmitter, the energy E is distributed equally all over the surface of a SPHERE of radius d at time T; provided that $T \gg t$, the fact that this is actually a finite

spherical shell of thickness ct will be of no consequence. Thus the <u>energy density</u> in joules per sq. metre on the surface of this sphere will be, if d is measured in metres;

$$\frac{E}{\text{Surface area of sphere}} = \frac{E}{4\pi d^2}$$

A small target of cross-sectional area A at right angles to the radius of the sphere will therefore intercept energy $EA/4\pi d^2$ in time t, i.e. the intercepted <u>power</u> will be

$$P_i = \frac{PA}{4\pi d^2} \qquad \text{watts} \qquad \text{(A in sq. metres if d in metres)}$$

1.2. 'Gain' of an antenna

In fact no real transmitter can be an isotropic radiator. Real radio antennas are constructed to maximise the power beamed in some directions while minimising the power beamed in others, i.e., they are highly directional. The directional characteristics of real radio antennas are complex, but a simplification will do for order-of-magnitude calculations. We can assume that a given antenna transmits only into a fraction f of the sphere around it; that it beams its power equally in all directions within the fraction f, and beams no power at all in other directions. Following the previous argument you now find that

$$P_i = \frac{PA}{f.4\pi d^2} \qquad \text{watts}$$

PROVIDED THE AREA A IS COMPLETELY IN THE 'BEAM' OF THE TRANSMITTER.

From its definition, f is less than 1. Provided the area A lies completely in the antenna's beam of transmission, the intercepted power increases as 1/f. For this reason the quantity 1/f is called the GAIN of a given transmitter. The gain is used by radio engineers as a 'figure of merit' for antenna directivity. For many types of antenna the gain is approximately 4π times the cross-sectional area of the antenna measured in wavelengths.

1.3. Reflection of Radio Waves at a Boundary

The propagation of radio waves through gases, liquids and solids can be described in exactly the same manner as the propagation of light; i.e., for most

materials the phenomena can be described by a REFRACTIVE INDEX n and an ABSORPTION COEFFICIENT a. The refractive index is related to the velocity of wave propagation in the material concerned: this velocity $v = c/n$ where c is the velocity of light in a vacuum. The absorption coefficient is such that the power in a given signal is reduced by a factor

$$e^{-ad}$$ after travelling any distance d in the material

The theory of electromagnetic radiation shows that radio waves, light, and other forms of electromagnetic wave are REFLECTED at boundaries where there are sudden changes in refractive index or absorption coefficient. Total reflection is rare; usually only a fraction of the radiation arriving at a boundary is reflected, and the rest is transmitted. The general equations describing reflection at a boundary are quite complicated, but a simple case contains the essential information for our present purpose.

If radio waves arrive at right angles to a boundary between two materials with refractive indices n_1 and n_2, then the fraction of the arriving power which will be reflected at the boundary is:

$$\text{Fraction reflected} = [(n_1-n_2)/(n_1+n_2)]^2$$

Note that the refractive indices must be the refractive indices of the materials FOR RADIO WAVES (not always, or even often, the same numerical values as their refractive indices for light).

Thus, to get a strong reflection, a sudden change in refractive index by a large amount is needed. The goemetry of the reflection of radio waves is the same as for light, i.e. angle of incidence equals angle of reflection, at a plane boundary.

Because most GASES have radio refractive indices which differ from unity by only a few parts per thousand, the transitions from one level to another in gaseous atmospheres are generally only very poor reflectors. The major exceptions are IONISED GASES, i.e., gases containing appreciable numbers of electrons and positive ions moving separately. The ionised layers of the Earth's upper atmosphere (the 'IONOSPHERE') can be good reflectors at radio wavelengths longer than a few metres (radio frequencies below $\sim$ 20 MHz); this fact is used for 'shortwave' communication around the curve of the Earth.

Many liquids, and most minerals and metals have large refractive indices and large absorption coefficients at radio wavelengths. The well-defined boundaries between gaseous atmospheres and ocean or land masses therefore tend

to be strong reflectors. Because of the large absorption coefficients, most
of the power <u>not</u> reflected at the first such boundary encountered by a radio
beam is absorbed shortly beyond it, so that little power reaches deeper boundaries
to produce further reflections. Thus, strong radar 'echoes' will come from
boundaries with solid or liquid surfaces, or from very dense cloud decks containing
many liquid particles or ice crystals.

1.4. The Radar Backscatter Factor, b

If a planet were a perfectly smooth reflecting sphere, only the point on
the planet closest to the transmitter would reflect a signal back to it. Real
planetary surfaces will be locally rough and have large-scale topography
however (mountains, craters, etc.) so that a sizeable fraction of a planet's
area may be able to reflect signals back to a given transmitter at any moment.
The complicated situation can be summarised in the 'backscatter factor' b. If
a total radar power P is intercepted by the planet, then b is defined so that
the power reflected back to the transmitter is AS IF the planet behaved as an
ISOTROPIC RADIATOR of power bP. The empirical value of b will vary with the
radar wavelength and with the presentation of the planet to the transmitter.
For non-reflective or very smooth planets, b would be near zero. Depending
on the topography of the planet, b can be greater than unity.

1.5. Solar System Radar Targets

The various objects in the Solar System are not all good radar targets.
The Sun does not have a well-defined surface and is itself a very powerful
source of natural radio waves. Attempts to get radar returns from the Sun there-
fore experience a very severe 'noise background' of the Sun's own natural emissions,
and do not give accurate results for the Earth-Sun distance.

The rocky planets Mercury, Venus and Mars, as well as Earth's Moon, give
strong radar reflections, whereas the outer planets, such as Jupiter and Saturn
give weak and confusing returns (implying that they do not have sharply-defined
surfaces). The very distant planets Uranus, Neptune and Pluto give only very
feeble radar returns due to the tremendous distances from Earth to these planets.

Fortunately, we do not need to range <u>all</u> of the planets to find out all
of their distances from the Sun; in fact, only one good measurement is needed from
radar studies, and all others can be derived from this measurement by geometry.

and orbit-timing. The key measurement is the distance to the planet Venus (which is the best radar target in the inner Solar System when size, distance and back-scatter factors are all taken into account) at a <u>particular configuration</u> of the Earth, the Sun and Venus.

To simplify the situation, suppose Venus travelled in an exactly circular orbit around the Sun (the actual orbit is an ellipse of eccentricity 0.007). From the diagram, note that <u>when Venus is at its greatest angular separation from the Sun</u>, as seen from Earth, the angle SVE must be 90°. (This configuration occurs when the line VE is <u>tangent</u> to Venus' circular orbit). We can recognise this configuration from Earth by carefully tracking Venus and the Sun, and so we can observe that <u>in this configuration the angle VES between Venus and the Sun</u> in Earth's sky <u>is 47°5</u>. (Knowing that SVE is 90° at this time, simple trigonometry relates the Sun-Venus distance SV and the Sun-Earth distance SE to the measurable Earth-Venus distance EV:

$$ES = EV \sec(47°5)$$
$$VS = EV \tan(47°5)$$

By this means, ES and VS can be found by measuring EV with radar at precisely the moment when Venus is at the correct configuration. The experiment could be repeated each time Venus comes to the right configuration, to improve the accuracy (and also, in fact, to allow for the small variations in VS, ES and in the angle VES due to the small departures of the planetary orbits from circles).

Once the Earth-Sun distance is known, we can find the distances to the outer (SUPERIOR) planets knowing their orbit times, without use of radar. To illustrate the method (neglecting geometrical complications due to noncircular and inclined orbits), consider the diagram on page 6.   If **we** used the same approach as before, we would need to measure the angle EPS at the time when angle SEP is 90°. <u>This requires travel to planet P!</u> So instead **we** use the fact that at some later time, Earth will be at E', on same line to the Sun as
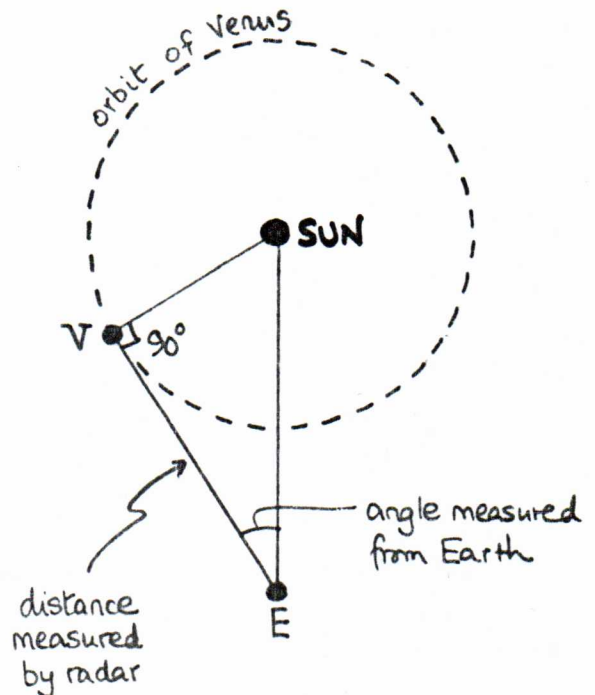
the planet, which is then at P'. This pair of configurations is <u>bound</u> to recur as Earth and other planet both go round Sun in their orbits.
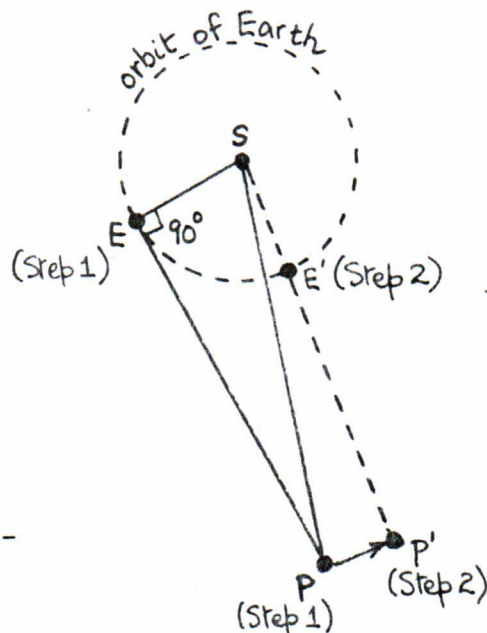
Step 1: Note when planet P is 90° from Sun as seen from Earth (i.e., note when angle SEP = 90°).

Step 2: Note when planet P is <u>next</u> 180° from Sun as seen from Earth (indicated configuration SE'P').

Record the time <u>difference</u> between these two configurations; call it t years. We can infer that the angle E'SE must = $360° \times (\frac{t}{1 \text{ year}})$. We can similarly find angle P'SP if we have measured the time it takes planet P to go once around Sun. Suppose we know that P goes once round the Sun every T years. Then P'SP must = $360° \times (\frac{t}{T})$ and angle PSE must = (Angle E'SE – Angle P'SP),

   i.e., angle PSE = $360° \times (t - t/T)$ ← all measured quantities .

We can then calculate the planet-Sun distance PS from:

$$PS = ES \sec (PSE)$$

where both ES and the angle PSE are now known from our measurements.

1.6. The Average Distances of Planets From Sun

Planet	Average Distance (km)	Average Distance (Earth = 1)	Orbit Time T (yrs)
Mercury	0.579×10^8	0.3871	0.241
Venus	1.082×10^8	0.7233	0.615
Earth	1.496×10^8	1.000	1.000
Mars	2.279×10^8	1.5237	1.881
Jupiter	7.783×10^8	5.20	11.86
Saturn	1.427×10^9	9.54	29.46
Uranus	2.870×10^9	19.2	84.02
Neptune	4.5×10^9	30.1	164.79
Pluto	5.9×10^9	39.5	247.7

Notes: 1) The average distance of the Earth from the Sun is a convenient scale for measurement of planetary distances, called the Astronomical Unit (A.U.).

2) The orbits are actually ellipses with the Sun at one focus of each ellipse (see later).

3) The ratios of orbit times are not the same as the ratios of orbit sizes, i.e., the outer planets travel their orbits at slower speeds. In fact $(\text{ORBIT TIME})^2 \propto (\text{DISTANCE FROM SUN})^3$

$$T^2 \propto r^3$$

This $T^2 \propto r^3$ relation provides the key to finding the MASSES of objects in Solar System.

2. The Period - Size Relation for Circular Orbits

The inverse-square law of gravitational attraction produces a simple relation between the period and size of an orbit performed around a fixed attracting mass. The relation also holds for the general elliptical orbit, as a relation between period and size of the major axis of the ellipse; the derivation for the general case is complicated by the elliptical geometry (see the notes in the Ring Binder if you are interested in the messy details!). In the circular-orbit case the derivation is very simple, however:
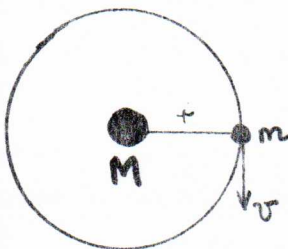
Consider a mass m in a circular orbit of radius d around a fixed mass M. The orbit is maintained by the gravitational attraction of M for m. Both masses are treated as mass POINTS.

Then, by the Newtonian formulae, the gravitational attractive force F of M for m is:

$$F = \frac{GMm}{r^2}$$

This must provide the required centripetal acceleration of m in the orbit

i.e., $a = v^2/r$

Using F = ma then gives $\underline{v^2 = GM/r}$ (note that m cancels)   Equation 2.1

We can introduce the orbital period of m as a parameter by noting that the circumference of the orbit $2\pi r$ is performed in the orbital period T at velocity v, so that

$\underline{2\pi r = vT}$ Equation 2.2

Eliminating v between Equations 2.1 and 2.2, and rearranging:

$$\boxed{T^2 = \frac{4\pi^2}{GM} r^3}$$ Equation 2.3

So for circular orbits under Newtonian gravitation around a fixed mass M, bodies orbiting at different distances r should have orbital periods T increasing so that T^2 is proportional to r^3. The constant of proportionality depends only on the value of the Newtonian gravitational constant G (known from the lab to be 6.67×10^{-11} $m^3 \cdot kg^{-1} \cdot sec^{-2}$) and on the mass M of the central object. If the central object can indeed be considered to be fixed, the masses of the orbiting objects cancel out.

This gives, in principle, a means of inferring the masses of gravitating objects by timing the orbits of bodies going around them at known distances. For example, timing planetary orbits will give a value for the mass of the Sun. Timing natural or artificial satellite orbits will give a value for the mass of a planet.

For elliptical orbits, r in the above expressions should be replaced by half the length of the major axis of the ellipse.

If M is not fixed but is free to move, r remains the distance between m and M, but the denominator in Equation 2.3 becomes G(M + m) instead of GM. In fact m << M for all but the Earth-Moon system, so this correction is relatively unimportant.

When several masses m orbit simultaneously, their interactions with one another should be considered. When m << M for all of them, the corrections are small.

3. <u>Timing the Planetary Orbits</u>

When we use the planetary motions to study the dynamics of the Solar System, e.g., to derive a value for the mass of the Sun or to obtain the distance to an outer planet, we need to consider the motions as they would appear to an observer who was at rest relative to the Sun. For example, the 'orbital period' of a planet should mean the interval of time during which the planet makes one complete circuit of the Sun as referenced to a 'fixed' set of space coordinates. In practice, the background of the distant stars is the appropriate reference system.

The observations we can actually make however must be made from the moving Earth. The time interval between a planet's appearance against a given stellar background on two successive occasions will NOT be the same as its orbital period, because of the motion of the Earth during that time. Allowing for the motion of the Earth when analysing the planetary motions is complicated; <u>there is</u>, however, <u>a simple relationship between the physically interesting</u> <u>orbital period of a planet and one easily-observable time interval, the so-called</u> <u>SYNODIC PERIOD of a planet</u>.

<u>The synodic period of a planet is defined as the interval between</u> <u>successive repetitions of the same Sun-Earth-Planet configuration</u>, e.g., the 'in-line' configuration (see diagram).
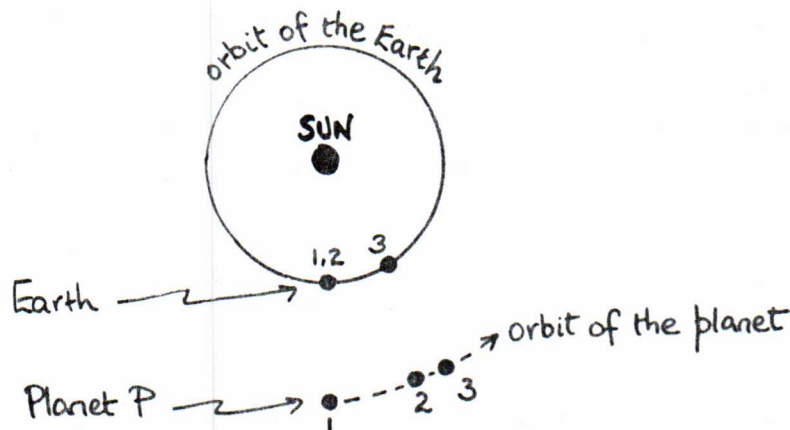
Beginning with the initial in-line configuration (#1), note that after one orbital period (year) of the Earth, the 'bent' configuration #2 results; the Earth has returned to the same position in its orbit, but planet P has moved further around its orbit. The in-line configuration is not repeated until slightly more than a year after #1, at #3. <u>The time interval between configuration</u>

#1 and configuration #3 is known as the SYNODIC PERIOD of planet P. Denote it by $t(P)$. Call the orbital periods $T(E)$ and $T(P)$ - $T(E)$ is one year.

We note that in time $t(P)$ planet P has travelled an angle $\theta°$ around the Sun while Earth has travelled an angle $360° + \theta°$. As Earth orbits at the constant rate $360°/T(E)$ and the planet at the constant rate $360°/T(P)$ (taking circular orbits as usual for simplicity) we know that

$$\theta° = \frac{360°}{T(P)} \times t(P) \qquad \text{and} \qquad 360° + \theta° = \frac{360°}{T(E)} \times t(P)$$

A little algebra eliminates θ to give the result

$$\boxed{\frac{1}{t(P)} = \frac{1}{T(E)} - \frac{1}{T(P)}}$$

Equation 3.1

Hence we can deduce the orbital period $T(P)$ from measurements of $t(P)$, which are easy to make, and knowledge of $T(E)$.

The formula holds for a planet whose orbit is OUTSIDE that of Earth. What is the correct formula for one whose orbit is INSIDE?

4. Sizes and Mean Densities of Solar System Bodies

There are two basic methods for finding the linear sizes of the planets or of the Sun:

a) from measuring the angular size of the object and its distance from Earth:
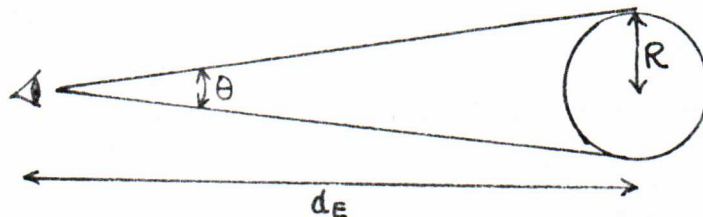
In this case we measure the angular diameter θ of the object when it is at a known distance d_E from Earth. Then $R = d_E \tan(\frac{\theta}{2})$,

b) from planetary radar:

Radar Transmitter
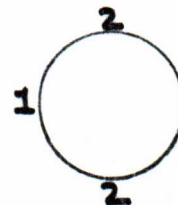
The pulses returned from the closest point on planet (1) travel a distance $(d_E - R)$ and back, while the pulses returned from the edges of the planet (2)

travel a distance $\sqrt{d_E^2 + R^2}$ and back. If we transmit sharp pulses, we therefore receive back a "blurred" echo:
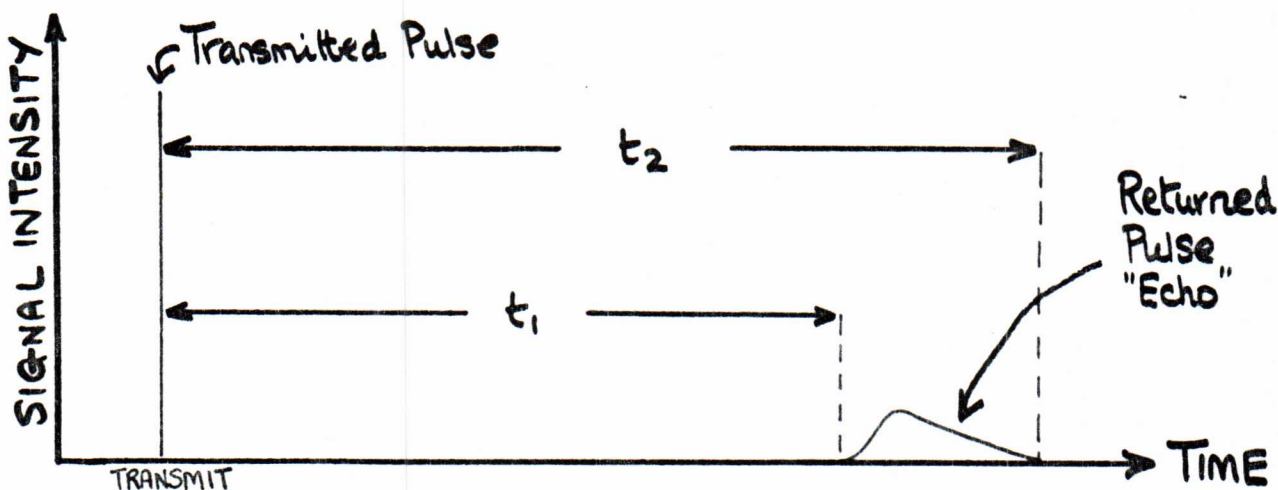
If $R \ll d_E$ (as will normally be true), then $R \sim \frac{1}{2} c (t_2 - t_1)$ and $d_E \sim \frac{1}{2} c t_2$. The radar method (measuring $t_2 - t_1$) is the best method of getting the sizes of SOLID planets (Mercury, Venus, Mars) as it is not affected by fuzzy images (atmospheres) or by small image sizes, which can cause inaccuracies in the angular-size method.

The table on p. 12 summarises the modern values of the masses, volumes and mean densities of the Solar System planets and, for comparison, the Sun. You may find small discrepancies between these figures and those given in some texts, particularly older texts. Where possible I have taken the latest or best-measured values from the scientific literature; there are recent measurements of the mean radii of the outer planets which represent revisions of about 10% from values adopted in the 1960's and early 1970's, and the quoted values should be considered to have uncertainties of several per cent. The planets are not exactly spherical (see Section 5 below). The "mean radius" quoted is the radius of the sphere whose volume equals that of the planet.

OBJECT	MASS (kg)	MEAN RADIUS (km)	MASS (EARTH = 1)	VOLUME	MEAN DENSITY $(kg.m^{-3})$
Sun	1.986×10^{30}	696,000	333,440	1.304×10^6	1406
Mercury	3.31×10^{23}	2,435	0.055	0.056	5470
Venus	4.87×10^{24}	6,050	0.815	0.856	5260
Earth	5.98×10^{24}	6,371	1.000	1.000	5517
Mars	6.42×10^{23}	3,390	0.107	0.151	3933
Jupiter	1.90×10^{27}	71,350	317.9	1400	1250
Saturn	5.69×10^{26}	57,800	95.2	750	700
Uranus	8.7×10^{25}	25,150	14.6	62	1310
Neptune	1.0×10^{26}	25,000	17	60	1530
Pluto	1.6×10^{22}	?	0.0026	?	?

Notes:

1. The size parameters for Pluto are extremely uncertain because of its great distance from the Sun. The mass for Pluto is estimated from the orbital parameters of the satellite discovered in 1978.

2. Observe the separation of mean densities into "Earth-like" (terrestrial) and "Jupiter-like" (Jovian) groups: Mercury, Venus and Earth contrasting with Jupiter, Saturn, Uranus and Neptune.

3. Observe that the high-density planets are also low-mass, and vice versa. Mars is normally classified as a "terrestrial" planet on these grounds.

4. Observe that the Jovian planets have mean densities comparable to that of the Sun.

5. Compare the mean densities with that of Earth-surface rocky materials, which is about 2800 $kg.m^{-3}$. All of the terrestrial planets are denser overall than Earth-surface (crustal) rock. All of the Jovian planets are less dense than Earth-surface rock, and Saturn would float in water!

These "bulk" properties of the planets already give us clues to the variety of compositions in the Solar System, and to the organisation of the Solar System. We will follow these clues in detail later on when we attempt to construct a history of the Solar System.

5. The Polar Flattening and Rotation Rates of the Planets

These prosaic quantities give important clues to the internal structure of the planets. We noted in Section 4 that planets are generally not quite spherical. To make this quantitative, define

| Optical (surface) OBLATENESS | $\varepsilon = \dfrac{a-b}{a}$ |
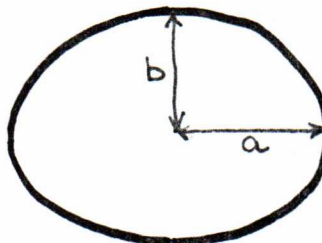
where a = longest radius of the planet and b = shortest radius. The observed oblatenesses vary from undetectable (Mercury, Venus) to 0.096 (Saturn). Measurements of oblateness provide clues to the INTERNAL STRUCTURES of the planets, when combined with data on their rates of rotation.

5.1 Methods of Measuring Rotation Rate

#1) Observe transits of visible features on the planetary disks, or monitor the total brightness variations of planets. If features repeat at regular intervals, or if the brightness "cycles" at a regular rate, the "repeat time" may be the rotation time. This approach has disadvantages: "features" may be cloud formations, etc. not rigidly attached to planetary surface.

#2) Use planetary radar and study the DOPPLER EFFECT in returned radar signals. Suppose that a source of radiation at rest relative to an observer emits radiation of wavelength λ. If the same source recedes from the observer at velocity v, the observer detects a longer wavelength $\lambda' = \lambda + \Delta\lambda$. If the source approaches the observer, he detects $\lambda' = \lambda - \Delta\lambda$. For source velocities v << c (the velocity of the radiation), this "Doppler Effect" is numerically

$$\frac{\Delta\lambda}{\lambda} = \frac{v}{c}$$

Equation 5.1

Positive v means recession, negative v means approach.

A point on a planet reflecting a radar signal back to Earth will generally be moving relative to the radar installation. The motion will have the following components:

i) Motion of the radar around the Sun

ii) Motion of the radar around the Earth's axis

iii) Motion of the planet around the Sun

iv) Motion of the reflecting point due to the rotation of the planet.

Of these, the first three can be considered <u>known</u> once we know the planetary orbit. Observations of the Doppler effect in the radar return will then tell us (iv). The "known" velocity components can easily be allowed for when interpreting differences in wavelength between the transmitted and returned signals. Assume this is done, then consider just the <u>effect of planetary rotation on the returned wavelength</u>:

Radar

Transmitter

Emits Wavelength λ
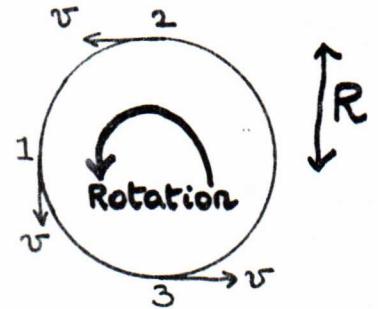
Suppose we bounce waves of wavelength λ off a planet whose radius is R, and which is rotating once in time interval T. Also suppose that the radar is located on the <u>equator</u> of the planet (if it is not, knowable but tedious geometry factors come in). Points on the equator of the planet travel at velocity $v = \dfrac{2\pi R}{T}$. Consulting the diagram above, consider reflection from:

<u>Point 1</u> (Earliest return). v is at 90° to observer, so there is no Doppler shift, $\lambda_{returned} = \lambda_{transmitted}$.

<u>Point 2</u> (Latest return). v is <u>towards</u> observer, so $\Delta\lambda/\lambda = -v/c = -2\pi R/cT$.

<u>Point 3</u> (Also part of latest return). Here $\Delta\lambda/\lambda = +v/c = +2\pi R/cT$
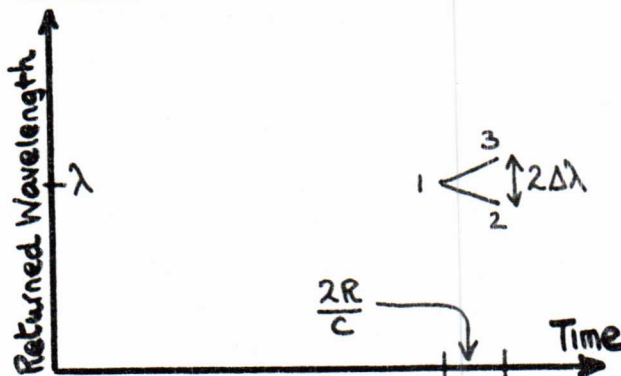
i.e., $\dfrac{\Delta\lambda}{\lambda} = \dfrac{2\pi R}{cT}$

and $\boxed{T = \dfrac{2\pi R \lambda}{c\Delta\lambda}}$ Equation 5.2

We can measure all quantities on the right and hence determine T by analysing the frequency spread of the returned pulse.

#3) Observe the <u>spectrum of reflected sunlight</u> from planet. This is also a DOPPLER EFFECT METHOD, in which natural sunlight replaces the radar signals of Method #2. The light from the Sun contains absorption (dark) lines at characteristic wavelengths corresponding to the atomic composition of the outer layers of the Sun. Reflection of sunlight from moving planetary features gives a

pattern of Doppler shifts $\Delta\lambda$ as before, recognisable as a <u>shifted</u> pattern of the characteristic absorption lines.

Slit of Spectroscope → **2** ← **image of planet**

1

rotation of planet

3
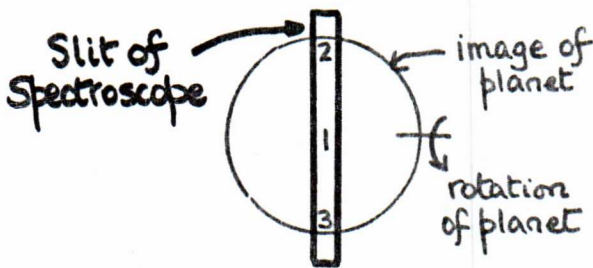
1. reflection from material travelling across line of sight → no Doppler -

2. **reflection from approaching material** → "blue shift" (λ decreased)

3. reflection from receding material → "red shift" (λ increased)

↓
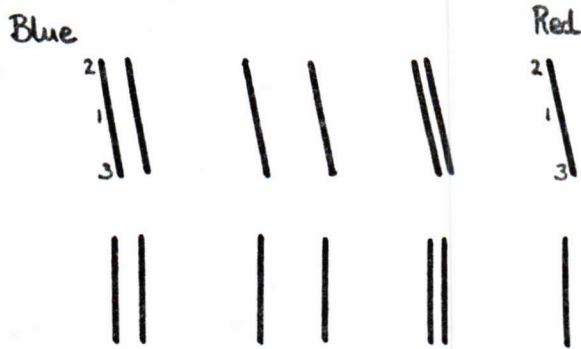
appearance of spectral lines in spectrum of rotating planet when slit is aligned perpendicular to rotation axis of planet

Blue **Red**

appearance of <u>same</u> lines when slit is <u>parallel</u> to rotation axis (or when there is <u>no</u> rotation).

Individual images of slit formed by light of ONE component wavelength in the sunlight spectrum.

For the inner planets, Radar method #2 is preferred as these planets are good targets nearby. For the outer planets, we are forced to use methods #1 and #3 by their large distances and the fact that they are poor radar targets. Large uncertainties still exist for some outer-planet rotations, e.g., Hayes and Belton, "The Rotational Periods of Uranus and Neptune", <u>Icarus</u>, vol. 32, pp. 383-401 (1977).

5.2 <u>Theory of the Oblateness of a Rotating, Gravitating Fluid Planet</u>

Once we have to consider the gravitational potential of bodies whose shapes are not spherical, the mathematical analysis becomes very messy. It is possible however to <u>illustrate</u> the basic concepts by which the oblateness and rotation of planets can be related, without getting into too much algebra. The analysis here is a "quick and dirty" treatment which attempts to get at the important parameters by <u>approximate</u> arguments based on the spherical shape. Those

interested in a proper treatment of the nonspherical shape might consult "Ellipsoidal Figures of Equilibrium" by S. Chandrasekhar (call no. QB410.C47); be warned that a strong stomach for mathematical analysis would be required to get much out of it!

Step #1. <u>Surface Forces on a Rigid Rotating Sphere</u>

First consider the effect of rotation on the force per unit mass at the surface of a large gravitating spherical planet of radius R and uniform density ρ, rotating once in time T.

The total mass M of the planet $= \frac{4}{3} \pi \rho R^3$

The angular velocity ω of the planet $= \frac{2\pi}{T}$

The <u>gravitational</u> force on mass m anywhere on the planetary surface will be
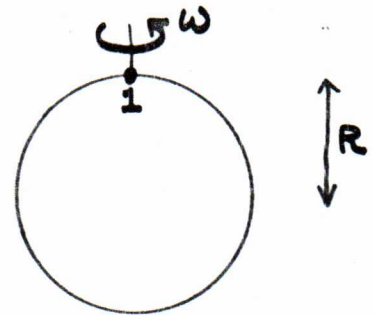
$$F_{grav} = \frac{GMm}{R^2} = \frac{G}{R^2} \cdot \frac{4}{3} \pi \rho R^3 m$$

A unit mass placed at the pole of the planet will be pressed against the surface by this force. Thus the weight of <u>unit</u> mass on the planetary surface at point #1 in the diagram will be

$$\frac{4}{3} \pi \rho G R \qquad\qquad\qquad \text{Equation 5.3}$$

At the equator of the planet however a force $F = m\omega^2 R$ is needed simply to provide the centripetal acceleration to keep mass m from flying off into space tangent to the planetary surface. Thus, although the gravitational force F_{grav} is the same at the equator as it is at the pole, part of it is simply keeping the mass m moving on the necessary circular path of radius R. Only the force <u>difference</u>

$$F_{grav} - m\omega^2 R = m \left[\frac{4}{3} \pi \rho G R - \omega^2 R\right] = \left[\frac{4}{3} \pi \rho G - \omega^2\right] mR$$

presses mass m onto the surface at the equator. Thus the weight of <u>unit</u> mass at the equator is reduced to

$$\left[\frac{4}{3} \pi \rho G - \omega^2\right] R \qquad\qquad\qquad \text{Equation 5.4}$$

by the planetary rotation. (It is sometimes said that this is due to the action of a "centrifugal force"; but this is an artifact. There is only one force present so far as any external observer is concerned, namely the gravitational force. At the planetary equator it is required to do two jobs; at the pole only one.)

The reduction in weight of unit mass at the equator means that a rotating planet is squeezed less across its equator than it is across its poles. (Consider what happens at intermediate latitudes).

A non-rigid planet will deform under this unequal self-squeezing.

Step #2. Total Force on a Column of Planetary Material

Next consider the total inwards force on a column of material from the surface to the centre of the planet. Let the column be a cylinder of cross-sectional area A. Let the planet be of radius R and uniform density ρ as before. Consider a small depth dr of the column at radius r in the planet, as in the diagram.
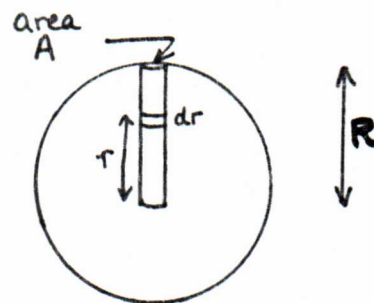
The mass in this small depth is $m = A\rho dr$

It is a property of the inverse-square law of the gravitational force that only the part of the planetary mass INSIDE radius r will exert a net attraction on the mass m. The attraction to all the material outside radius r cancels out (this is a consequence of the assumed spherical shape). Thus the inwards force on the small depth dr of the column is:

$$F = \frac{GM(r)m}{r^2} \quad \text{where } M(r) = \text{mass inside } r = \frac{4}{3}\pi\rho r^3$$

(It is also a property of the inverse-square law that $M(\mathbf{r})$ behaves gravitationally as if it were concentrated at the centre of the planet, so that the above expression is indeed correct).

The total force on the column is obtained by integrating the above expression from the centre of the planet (r = 0) to the surface r = R):

$$F_{total} = \int_0^R \frac{G}{r^2} \cdot \frac{4}{3}\pi\rho r^3 \cdot A\rho dr = \frac{4}{3}\pi\rho^2 GA \cdot \left[\int_0^R dr\right] = \frac{4}{3}\pi\rho^2 GA \frac{R^2}{2}$$

Thus the total gravitational <u>pressure</u> (force per unit area) at the base of the column can be written

$$P_g = F/A = [\frac{4}{3} \pi \rho G (\frac{R}{2})] \times \rho R \qquad \text{for a polar column} \qquad \underline{\text{Equation 5.5}}$$

i.e., P = [<u>average</u> force per unit mass] x [total mass above unit area in column]

By analogy, at the equator, using our results from Step #1 (Eq. 5.4)

$$P_g = [\frac{4}{3} \pi \rho G - \omega^2] [\frac{R}{2}] \times \rho R \qquad \text{for an equatorial column} \qquad \underline{\text{Equation 5.6}}$$

✝ Step #3. <u>A Deformable, but Incompressible Planet</u>

We now have the tools we need to make an <u>approximate</u> analysis of a more realistic situation in which the planetary material can <u>deform</u> under the unequal self-squeezing represented by these unequal pressures.

Suppose that the planet is modelled as an INCOMPRESSIBLE FLUID, i.e., <u>its shape is not fixed</u> by forces preserving solid rigidity, <u>but its density remains the same everywhere</u>. We can expect the greater forces across the poles to pull the planet in along the polar diameter; if it is incompressible it will bulge out all round the equator in compensation. In fact, it will come to shape equilibrium when the fluid comes to pressure balance, or HYDROSTATIC EQUILIBRIUM.

The equilibrium can be analysed by the following thought-experiment. Suppose that rigid-walled tubes of equal cross-sectional area A are sunk into the (now-fluid) planet and are filled with fluid as in the next diagram:

Let the polar radius of the rotationally-deformed planet be b, and the equatorial radius be a. When the fluid columns are both full to the surface, THEY MUST EXERT THE SAME PRESSURE AT THE CENTRE OF THE PLANET, otherwise one would force the other to a different level. Thus <u>we have to calculate how a and b will adjust to make the central pressures the same</u> at the base of each fluid column.
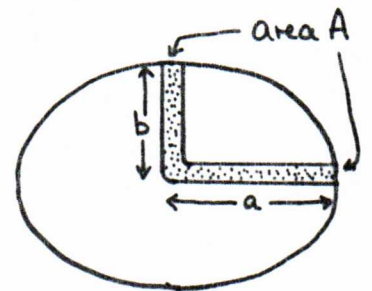
If the deformation of the planet from a sphere is <u>small</u>, i.e., if a and b are very nearly equal, we can take over the expressions derived in Step #2 for the pressures in columns through a <u>spherical</u> planet:

> Pressure = [average force per unit mass along the column concerned] x [total mass above unit area in column]

The second (total mass) term is easy to write down. For the polar column it is ρb and for the equatorial it is ρa. The average forces are harder.

To get close to the results of a proper calculation of the attraction of a non-spherical mass, we should slightly modify the expressions for the spherical case. For example, consider the average force per unit mass along the polar column. If we simply take over the result from Eq. 5.5, we would say the average force per unit mass along the polar column would be

$$\frac{4}{3} \pi \rho G \left(\frac{b}{2}\right)$$

But this would treat the planet as if it were a sphere of radius b overall. What we would be neglecting would be the attraction on the polar column of the equatorial "bulge" of the planet, i.e. the attraction of the matter lying <u>outside</u> the sphere of radius b. As this matter is <u>not</u> spherically distributed, its net attraction on the column does <u>not</u> cancel out but leaves a small extra contribution that has not been counted in by the above formula. The above expression for the average force per unit mass along the polar column is therefore SLIGHTLY TOO SMALL.

Similarly, for the equatorial column we could write down the force per unit mass from Eq. 5.6 as

$$\left[\frac{4}{3} \pi \rho G - \omega^2\right] \left(\frac{a}{2}\right)$$

but because this treats the planet as a sphere of radius a, it will be TOO LARGE.

It will be truer to the actual situation to replace b/2 and a/2 in <u>both</u> these expressions by R/2 where R is the MEAN RADIUS of the planet. As b < R < a, this replacement will make <u>rough</u> allowance for the gravitational attraction of the actual non-spherical shape of the planet.

Then pressure balance requires $\frac{4}{3} \pi \rho G \cdot \left(\frac{R}{2}\right) \cdot \rho b = \left[\frac{4}{3} \pi \rho G - \omega^2\right] \cdot \left(\frac{R}{2}\right) \cdot \rho a$

i.e. $\frac{b}{a} = 1 - \left(3\omega^2/4\pi G\rho\right)$

In terms of the oblateness $\varepsilon = \frac{a-b}{a}$, $\quad \varepsilon = \frac{3\omega^2}{4\pi G\rho}$

i.e., $\boxed{\varepsilon = \frac{\omega^2 R^3}{GM}}$

Equation 5.7

An EXACT calculation gives $\boxed{\epsilon = 1.25 \left(\dfrac{\omega^2 R^3}{GM}\right)}$ for an incompressible fluid planet (MACLAURIN SPHEROID), in place of Eq. 5.7.

Step #4. <u>A Deformable and Compressible Planet</u>

Real materials do not have constant density but are <u>compressed</u> when under pressure. The compressibility depends on the composition, temperature and pressure, and the exact behaviour will be difficult to compute. Qualitatively though, we would expect self-squeezing of a planet to <u>increase</u> the density inside it even if its <u>material</u> were the same throughout. Another "model planet" (called the "Roche spheroid") considers that most of the mass is concentrated at the planet's centre (in a very high-density core), but that most of the <u>volume</u> is in a low-density "halo". A proper calculation then shows that:

$$\boxed{\frac{GM\epsilon}{\omega^2 R^3} = 0.5}$$ for <u>strong</u> central compression   Equation 5.8

Now compare these <u>theoretical</u> extremes with the observed properties of the planets:

OBSERVED OBLATENESS, ROTATION, COMPRESSION

Planet	Oblateness ϵ	Rotation Time T	$\dfrac{GM\epsilon}{\omega^2 R^3}$	Deduced $\dfrac{\rho_{central}}{\rho_{average}}$
Mercury	small	59^d	?	?
Venus	small	$243^d.0$	?	?
Earth	0.00336	$23^h 56^m 04^s.1$	0.97	~ 2.5
Mars	0.009	$24^h 37^m 22^s.6$	1.14	~ 2
Jupiter	0.062	$9^h 50^m.5$	0.77	~ 3
Saturn	0.096	$10^h 14^m$	0.69	~ 6
Uranus	(0.01) uncertain	$(24^h \pm 3^h)*$	(0.69)	(~ 6)
Neptune	0.021	$(15^h.8 \pm 4^h)$	0.76	~ 3

*a value of $10^h.8$ used to be given.

NOTE: For all measured planets, $GM\epsilon/\omega^2 R^3$ <u>is</u> between 0.5 and 1.25, so there is evidence for <u>some</u> internal compression (central ρ_c > average $\bar\rho$). The implied inner densities for Jupiter and Saturn are about equal to the mean densities of the terrestrial planets. <u>Strong</u> central density increases are therefore implied for the ~~other~~ outer planets, raising the <u>possibility</u> that they might contain "terrestrial" <u>cores</u>.

6. Seismological Evidence for the Internal Compression of the Earth

The detailed nature of the density increase within the Earth has been studied by observing the way in which SEISMIC (earthquake) waves pass through the body of the Earth. The waves of interest here are not the highly-destructive SURFACE waves, which are fairly localised, but the waves which pass deep into the Earth's interior before re-emerging at the surface far from the epi-centre of the earthquake. These waves involve two kinds of distortion of the material through which they pass (see the accompanying diagram). The first class of wave is the compressional wave, in which the motions of adjacent pieces of material take place along the direction of propagation of the wave. The other class is the shear wave, in which the motions are trans-verse to the propagation of the wave. In the Earth, the compression waves travel faster and therefore arrive earlier at distant points. For this reason they are called primary, or P waves, and the slower-travelling shear waves are called secondary, or S waves. The most general waveform for a seismic wave is a combination of the two, as shown at the right of the diagram.

P waves   *S* waves   *S* and *P* waves
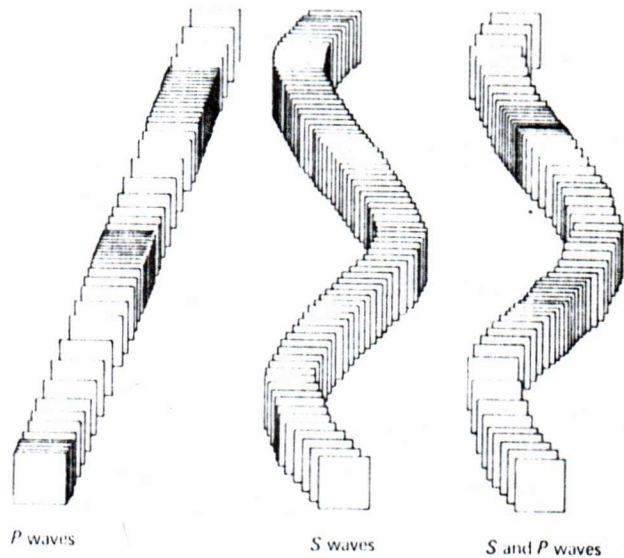
The velocities of the P and S waves depend on the density and on the elastic properties of the material through which they travel. From the theory of elasticity, it can be shown that the velocities of the P-waves and S-waves are

$$V_P = \sqrt{(k + \tfrac{4}{3}\mu)/\rho}$$ and $$V_S = \sqrt{(\mu/\rho)}$$ Equation 6.1

where ρ is the density of the material, k is its BULK MODULUS and μ its MODULUS OF RIGIDITY. Notice that propagation of shear waves requires significant adhesion between adjacent elements of the material, so that S-waves will not pass through liquids (which have $\mu = 0$). All of the quantities ρ, k and μ vary

with depth within the Earth, so the velocities of P and S waves vary with depth
and the body waves are REFRACTED as they path through the Earth.   Typical paths
of propagation from a single earthquake are shown in the next diagram.   By
observing the arrival of body waves at
many seismic stations from many different
earthquakes, it is possible to build up a
picture of how V_P and V_S vary with depth
in the Earth; this picture can then be
combined with known properties of materials
to put together an approximate model of the
variation of density and composition in the
Earth's interior.

        The clearest feature of
the velocity-depth data is the
existence of an abrupt transi-
tion in the seismic properties
of the Earth at a depth of
about 2900 km. Below this depth,
S-waves do not propagate at all,
and at this depth there is a sharp
discontinuity in the velocity of
the P-waves. This change corres-
ponds to an abrupt loss of rigi-
dity in the Earth's interior,
i.e. the transition to a LIQUID
CORE. This boundary produces
such strong refraction of the
P-waves that there is a 'shadow
zone' associated with any given

earthquake within which there is very little P-wave reception (see the previous
diagram). The main regimes of propagation can be divided into 1) the crustal
regime, in which V_P and V_S initially decrease with depth, 2) the mantle
regime, in which both velocities increase with depth, 3) the core regime, in
which $V_S = 0$, and 4) the inner core, in which V_P becomes more or less constant.
The boundary between the crust and the mantle is called the Mohorovicic discontin-
uity, after its discoverer (who identified it on a seismogram of an earthquake
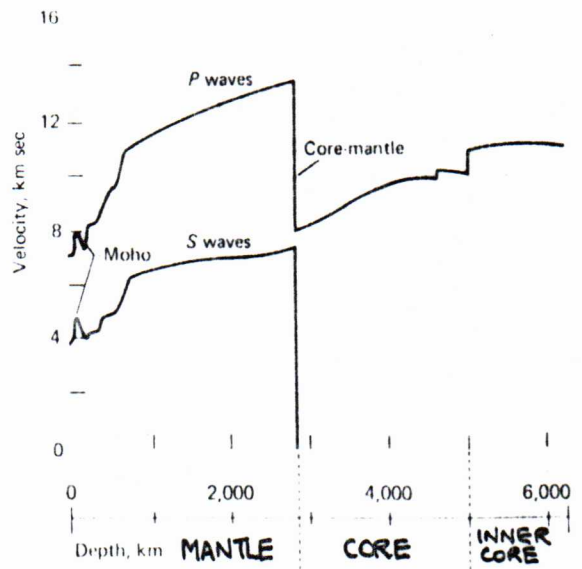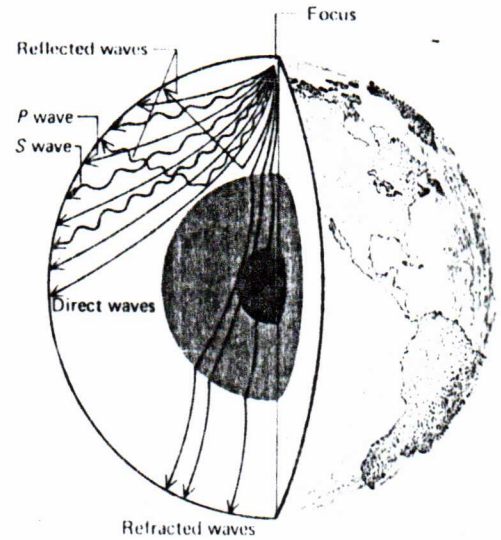
in Croatia in 1909). It occurs at a depth of about 5 km under the oceans and at a depth of about 35 km under the continents. The transition to the 'inner core' is less well-defined, and occurs at a depth of about 5100 km; there may in fact be several subsystems within the core.

Matching the observed velocity-depth curves to the properties of known materials is an extremely complicated task, the details of which go far beyond the scope of our present discussion. Basically we have to make educated guesses as to the nature of the materials in the Earth's interior, and to extrapolate their variation of density and elastic constants with pressure beyond the regime of laboratory measurements. As the inner parts of the Earth support the weight of the outer parts, we can compute that if each level is in hydrostatic equilibrium (i.e. pressure exactly balances the weight from above), then a certain variation of density with depth is required. The observed seismic parameters provide vital constraints on this calculation, and allow us to make a broadly satisfactory description of the observed data with a model in which the crust has a mean density of order 3800 kg/m^3, the mantle a mean density of order 4600 kg/m^3 and the core a mean density of order 10600 kg/m^3. With these parameters, the mantle contains about 70% of the mass of the Earth, and the core virtually all of the rest of the mass. The models give a range of permissible solutions for the detailed variation of density with depth, as indicated in the diagram to the right. The nature of the materials can also be guessed at from the seismic data. A fair fit to the observed mantle properties can be made with magnesium and iron-rich (MAFIC) silicate rocks such as OLIVINE $(Mg,Fe)_2SiO_4$, PYROXENE $(Ca,Mg,Fe)_2Si_2O_6$ and GARNET $(Ca,Mg)_3Al_2Si_3O_{12}$. The properties of the core are best matched by those of liquid iron, with some nickel, sulfur and silicon also present. It is important to realise how far from laboratory conditions we have to extrapolate to make such models however - at the centre of the core, where the density is probably near 13000 kg/m^3, we are probably dealing with a temperature
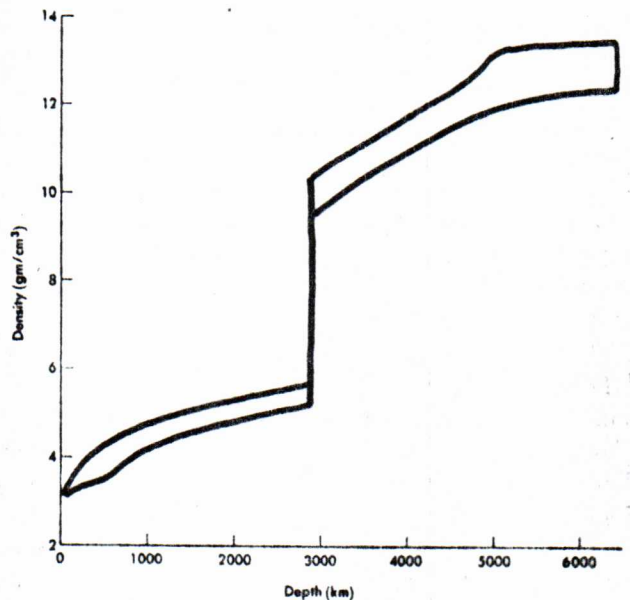
of $\sim$ 5000 K and a pressure of about 3 million atmospheres.

6.1 Evidence for High-Density Interiors of Other Planets

We have now landed spacecraft on the surfaces of Venus, Mars and the Moon, and have obtained <u>direct</u> evidence that the interiors of these bodies are more dense than their surfaces. For Venus, the Russian "Venera" probes gave a mean density of <u>surface</u> materials in the range 2700-2900 kg/m^3 (similar to the mean density of Earth crustal rock); this should be compared with the <u>overall</u> mean density of 5260 kg/m^3 given on p. 12. For Mars, the "Viking" landers found <u>surface</u> materials with densities in the range 2300-3200 kg/m^3, while the <u>overall</u> mean density is 3933 kg/m^3. Although we do not have seismic profiles of either planet, it is clear that their interiors must be denser than the surface rocks.

For the Moon, we have both this crude data and results from the seismic array left behind by the "Apollo" missions. We will treat the lunar interior in detail later on, but we can note here that the discrepancy between the mean density of the surface samples (3000 kg/m^3) and the overall mean density of the Moon (3340 kg/m^3) is much smaller than the discrepancy for any of the more massive inner-Solar System bodies, implying that <u>the Moon is much less centrally compressed than Mars, Venus or the Earth</u>.

While it is almost certainly an oversimplification, it is interesting to consider 'first-guess' models of the other 'rocky' planets in which we attempt to simulate their observed properties by assuming them to be composed of an Earth-like mantle and core, but in different <u>proportions</u>. The crusts probably contribute sufficiently little mass to be neglected altogether in a first approximation.

The models assume that each planet is composed of a <u>mantle of mafic silicates</u> and a <u>core of nickel-iron</u>. The experimental constraints are 1) the observed mean densities of the planets, 2) their total masses and 3) the known relations between pressure, density and temperature for silicates and for iron. The theoretical basis of the model calculations is 1) the assumption of hydrostatic equilibrium at all levels in each body and 2) the extrapolation of the pressure-density-temperature relations from laboratory measurements to high densities and high temperatures using the techniques of theoretical solid-state physics.

Such admittedly simplistic (but still not easy) calculations come up with an interesting picture of the variation of properties in the inner Solar System, as shown in the Table below. For each body we give both the observed mean density and the mean density which would be observed if the gravitational self-compression of each body could be removed (i.e. the mean density the planetary mixture would have if separated into small fragments which did not compress one another by their mutual gravity). This "decompressed" mean density (calculated using the model planetary structure) is a better indicator of composition than is the raw mean density, as it discounts the effects of the different total masses of the bodies (Earth squeezes its interior harder than Mercury does because Earth is more massive than Mercury). The final column gives the actual fraction of the planetary mass which would be in the iron-nickel core according to the model. See Wood, p. 66-75, for discussion of uncertainties in such models.

SIMPLE CORE-MANTLE MODELS OF TERRESTRIAL PLANETS

PLANET	OBSERVED MEAN DENSITY (kg/m^3)	"DECOMPRESSED" MEAN DENSITY (kg/m^3)	% OF PLANET MASS IN IRON-NICKEL CORE
Mercury	5470	5310	65%
Venus	5260	3950	27%
Earth	5520	4030	32%
Mars	3933	3650	19%
Moon	3340	3300	6%

The Table gives the first evidence for COMPOSITIONAL SEGREGATION in the inner Solar System. The innermost object, Mercury, is apparently richest in "iron"-core material. Earth and Venus are similar, and Mars is apparently the least rich in "iron" of the four PLANETARY bodies. Earth's satellite appears strangely iron-poor overall (if this type of modelling is in fact meaningful). We will pick up these clues, and some of the questions raised by them, later when we consider the processes which might cause segregation of properties in the Solar System.

7. Radiation Quantities

The following are definitions of terms used in making quantitative statements about radiation and radiating systems.

B
APPARENT BRIGHTNESS of an object = rate at which radiant energy of all wavelengths is received from the object, per square metre of detector. Symbol: B. Units: watts per square metre.

B_λ
FLUX DENSITY of an object = apparent brightness at a given wavelength, per unit wavelength interval, i.e. $B_\lambda d\lambda$ is apparent brightness of object including all wavelengths from λ to $\lambda + d\lambda$. Symbol: B_λ or in some texts S_λ. Units: watts per square metre per Hertz.

L
INTRINSIC LUMINOSITY of an object = rate at which energy of all wavelengths is radiated by object in all directions. Symbol: L. Units: watts. Sometimes normalised to one steradian of solid angle around object, in which case units are watts per steradian and the luminosity is reduced by a factor 4π.

L_λ
SPECTRAL LUMINOSITY of an object = L per unit wavelength interval. Symbol: L_λ. Units: watts per Hertz, or watts per steradian per Hertz.

E
SURFACE EMISSIVITY (SURFACE BRIGHTNESS) of an object = **rate** at which radiant energy of all wavelengths leaves one square metre of the object's surface in all directions, i.e. L per square metre of surface. Symbol: E. Units: watts per square metre, or watts per steradian per square metre.

E_λ
SPECTRAL EMISSIVITY of an object's surface = E per unit wavelength interval at given wavelength. Symbol: E_λ. Units: watts per **square** metre per Hertz, or watts per steradian per square metre per Hertz.

Geometrical Relations

$B = L/4\pi d^2$ – apparent brightness of object of luminosity L at a distance d from receiver.

$L = E \cdot 4\pi R^2$ – luminosity of sphere of radius R whose surface has uniform surface emissivity E.

7.1. The Perfect "Black-Body" Radiator

This a theoretical material which can interact completely with all wavelengths of radiation, freely absorbing and re-emitting all wavelengths.

Dilute materials - made up of effectively isolated atoms/molecules - have their absorption/emission processes governed by the internal quantum energy levels of these atoms/molecules. They exhibit the discrete-wavelength absorption and emission characteristic of their particular chemical species.

Dense materials - made up of strongly interacting atoms/molecules - have their individual atomic and molecular quantum energy levels distorted by interactions. Because the individual atoms/molecules in such materials have different energy levels, the ensemble can absorb or emit any wavelength. All sufficiently dense matter has a continuous spectrum characteristic only of the environment conditions (in particular the temperature T).

All sufficiently dense material behaves somewhat like the theoretical "perfect radiator"; the density required depends in detail on the material. Perfect Radiation theory predicts that

$$E_\lambda = \frac{2hc^2}{\lambda^5} \cdot \frac{1}{e^{hc/kT\lambda} - 1}$$

Equation 7.1

where h (Planck's constant) = 6.626×10^{-34} joules/Hz

k (Boltzmann's constant) = 1.381×10^{-23} joules/K

c (velocity of light) = 2.998×10^8 m/s

This relation is known as Planck's Law. The total emissivity over all wavelengths is:

$$E = \int_0^\infty E_\lambda d\lambda = [\frac{2\pi^5 k^4}{15c^2 h^3}]T^4 = \sigma T^4$$

Equation 7.2

This relation is known as the Stefan-Boltzmann law, and σ (Stefan's constant) = 5.669×10^{-8} watts/m^2/K^4.
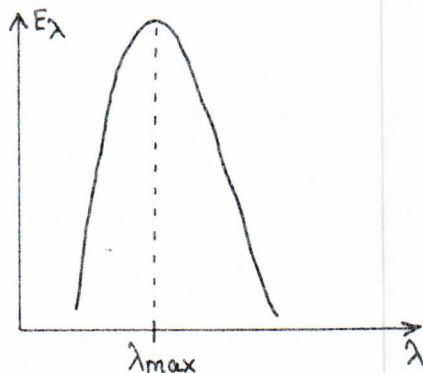
Wien's Law is an expression for the wavelength λ of maximum E_λ

$$\lambda_{max} = \frac{0.2898}{T} \text{ cm}$$

$$= \frac{2898}{T} \mu \text{ (microns)}$$

(1 micron = 10^{-6} metre)

8. Planetary Radiation

The study of the variation of intensity of planetary radiation with wavelength gives us clues to planetary temperatures and to the composition and condition of planetary atmospheres.

8.1. Broad Distribution of Energy with Wavelength
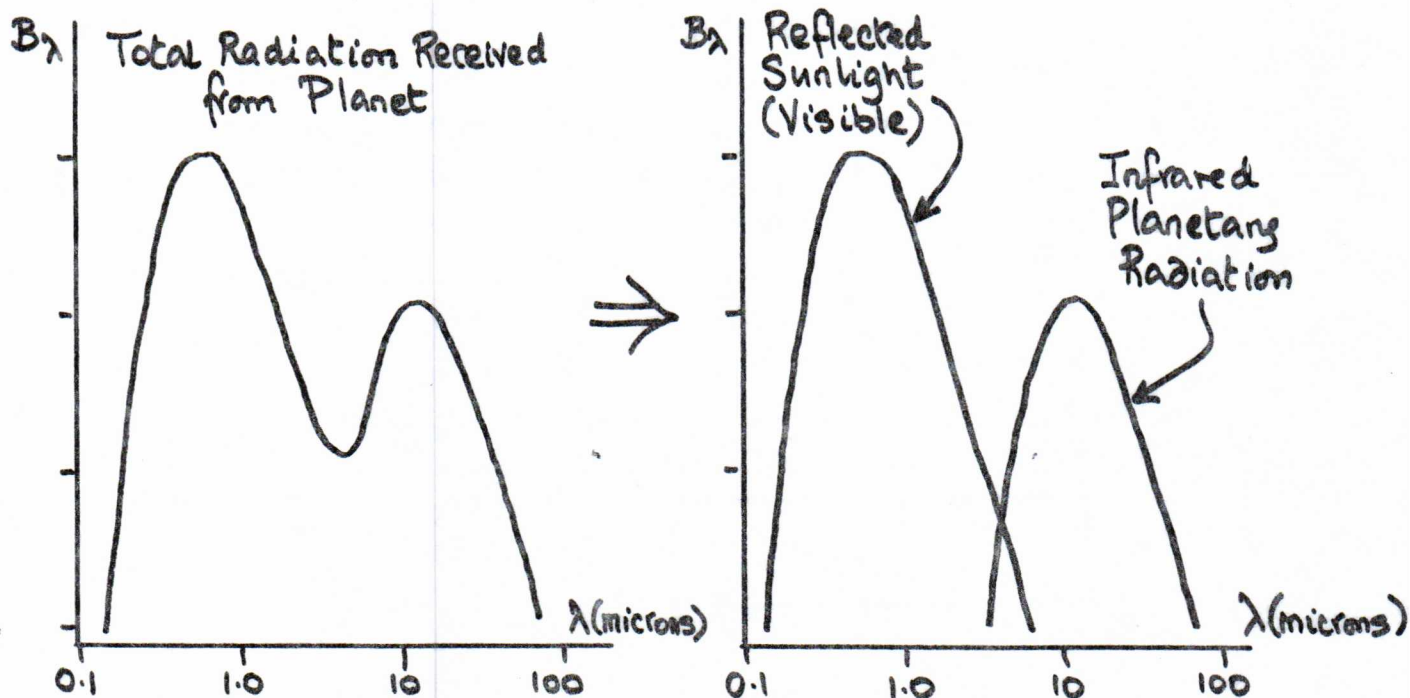
The example shown in the diagram is drawn from data for **Mars** and shows the typical features of a planetary spectrum. The spectrum contains two Planck-Law distributions, peaking at different wavelengths. The short-wavelength peak is always at ~ 0.5 microns wavelength. This is visible radiation, and the shape of the peak matches **the sunlight** spectrum at an effective temperature $T = 5780$ K. This peak is at the same wavelength for all planets. The long-wavelength peak is at $\stackrel{>}{\sim} 5$ microns wavelength. It is mostly infrared radiation (heat) at effective temperatures $\stackrel{<}{\sim} 600$ K. The infrared peak occurs at different wavelengths depending on which planet is being studied (outer planets generally give cooler temperatures).

We can make a simple interpretation of this two-peaked distribution. First, visible sunlight arrives at the planet's surface or at its upper cloud deck (if this is opaque). Then, some sunlight is reflected immediately and produces the short-wavelength (visible) peak in the spectrum. Some is absorbed

and heats the <u>planetary environment</u>, which then reradiates at <u>its</u> effective temperature to produce the long-wavelength (infrared) peak.

8.2. The Albedos of the Planets

The rate of reception of solar energy by a planet will be

$$S = \frac{L}{4\pi d^2} \cdot \pi R^2$$

where L is the solar luminosity, d is the sun-planet distance and R is the planet's radius. Note that this is governed by the <u>cross-sectional</u> area πR^2 of the planet, as this determines the planet's blockage of sunlight (see diagram).
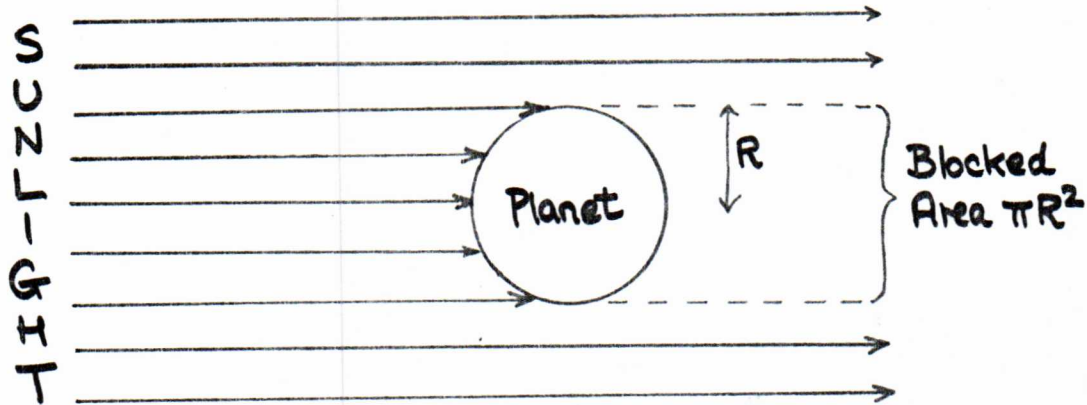
The fraction of this solar energy <u>reflected without absorption</u> is known as the ALBEDO of the planet. Because of the geometry of the reflections, reflected energy goes into all directions away from the planet. The detailed distribution of energy over different directions depends on the roughness and on the composition of the reflecting surface (which may be solid or an atmospheric cloud layer). The situation is analogous to radar backscatter (Section 1.4). In fact, the albedo varies both with direction (angle to Sun) and with wavelength λ. For our purpose, it will be sufficient to deal with the net albedo, <u>summed over all directions</u>, and appropriately <u>averaged over all the wavelengths</u> in the sunlight spectrum. We will call this net albedo A. To determine the net albedo of a planet, we first observe the total apparent brightness B_{VIS} of the planet's visible radiation at Earth. This is the area under the short-λ peak in the observed plot of B_λ versus λ for the planet. In practice this will be found to vary with the phase of planet, so observe (or extrapolate to) the "full"

phase, in which the sunlit disk of the planet is fully presented to Earth. Second, knowing the distance d_E to the planet when it is at full phase, we can compute $L_{VISIBLE} = B_{VIS} \cdot 4\pi d_E^2$, the effective visible luminosity of the planet. When this is done, we get the results in the following table:

PLANET	OBSERVED L_{VIS} (watts)	RATE OF SOLAR ENERGY RECEPTION S (watts)	ALBEDO A $= L_{VIS}/S$	NATURE OF REFLECTOR
Mercury	9.5×10^{13}	1.7×10^{15}	0.056	rocky surface
Venus	2.2×10^{17}	3.0×10^{17}	0.72	clouds
Earth	varies	1.7×10^{17}	average 0.36	rocks, ocean, clouds
Mars	3.4×10^{15}	2.1×10^{16}	0.16	rocky surface
Jupiter	5.6×10^{17}	8.0×10^{17}	0.70	clouds
Saturn	1.2×10^{17}	1.6×10^{17}	0.75	clouds
Uranus	5.1×10^{15}	7.3×10^{15}	0.70	clouds?
Neptune	2.4×10^{15}	2.9×10^{15}	0.82	?

Note the high albedos of planets surrounded by cloud decks, and the much lower net albedos of the planets which have visible rocky surfaces. Venus has the second-highest L_{VIS} and approaches closer to Earth than any other planet. This means it can be very bright in Earth's sky (and incidentally accounts for a significant number of UFO sightings). Mercury has the lowest L_{VIS} and is always near the Sun in our sky, so it is quite difficult to observe from Earth except when near maximum elongation from the Sun.

8.3. Black-Body Model for Temperature Distribution on a Planetary Surface

In the absence of significant heat transfer from one part of a planetary surface to another, each piece of planetary surface would come to its own equilibrium temperature T_e under the arriving solar radiation.

This situation will approximate the conditions on planetary bodies without much atmosphere, with only modest thermal conductivity (rocky materials), and with moderate rates of rotation (so that a given point on the surface does not significantly "remember" a recent time of more intense illumination by the Sun). In practice, these assumptions will be met by Mercury, the Moon, Mars and most asteroids (rotations slower than once every six hours or so). Planets with massive atmospheres, rapid rotation or small metallic bodies may depart strongly from this picture, due to heat transfer by the atmosphere, to thermal "memory" or to heat conduction respectively.
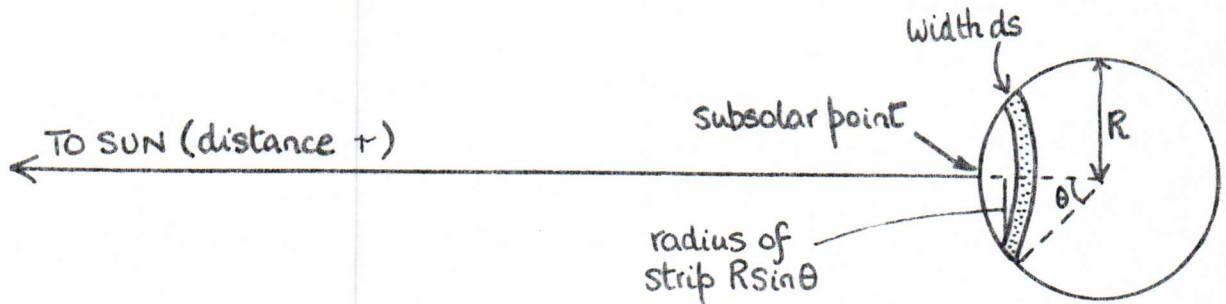
Consider a strip of planetary surface at an angle θ to the SUBSOLAR point where the Sun's illumination is vertical. Let the width of the strip be ds. If the planet's radius is R the strip subtends an angle $d\theta$ at the centre of the planet, where $ds = R\,d\theta$. The radius of the strip is $R\sin\theta$, so its total area is

$$2\pi(\text{radius})(\text{width}) = 2\pi\,R\sin\theta\cdot ds = 2\pi R^2\sin\theta\;d\theta$$

If the effective temperature of the strip is $T_e(\theta)$ then the rate of infrared radiation from the strip is (using Equation 7.2)

$$\sigma T_e^4(\theta)\cdot 2\pi R^2\sin\theta\cdot d\theta \quad \text{watts} \qquad (\sigma = \text{Stefan's constant})$$

But the rate of absorption of sunlight onto the strip is $(1-A)$ times the rate of arrival of sunlight on the strip, if A is the albedo. The rate of arrival of sunlight is equal to the power density of sunlight at the distance of the planet (r) times the underlined perpendicular area of the strip. Because the strip is tilted an angle θ away from the Sun, the perpendicular area is the actual area times $\cos\theta$. So the rate of absorption of sunlight onto the strip is

$$(1-A) \cdot \frac{L}{4\pi r^2} \cdot 2\pi R^2 \sin\theta \cdot d\theta \cdot \cos\theta$$

where L is the solar luminosity and r is the sun-pluanet distance, as in Section 8.2. At the equilibrium temperature, the rate of radiation from each strip would balance the rate of absorption of sunlight. Equating the two expressions given above, cancelling terms and rearranging, we get

$$T_e(\theta) = \left\{ \frac{(1-A)\ L\ \cos\theta}{4\pi\sigma r^2} \right\}^{1/4}$$

Equation 8.2

The temperature is a maximum at the subsolar point ($\theta = 0$, $\cos\theta = 1$) and falls off as the fourth root of $\cos\theta$ elsewhere on the surface of the planet. Infrared observations of planets can determine T_e at different θ; the observed variation of T_e across the planet can be compared with the expected law. The subsolar temperature is often quoted as a 'standardisable' parameter.

8.4. A "Smeared-Temperature" Approximation

In some cases however we cannot validly assume that each point on the planet comes separately to equilibrium, but must consider the effective temperature T_e to be "smeared" over a finite neighbourhood of the planetary surface. Obviously the detailed behaviour of T_e on the surface must vary with the process producing the smearing. For our purposes, it is worth considering a very rough approximation which will help us to interpret planetary temperatures derived from infrared data.

This rough approximation supposes a fraction f of the surface to be at the same effective temperature T_e, and the other (1-f) of the planetary surface to be at such a low effective temperature that its rate of radiation is negligible. (Recall that the total emissivity varies as T_e^4, so a two-fold decrease in temperature corresponds to a sixteenfold decrease in emissivity). Following the argument which led us to Eqn. 8.2 we would now say that the total rate of absorption of sunlight energy by the planet is (1-A)S watts, and in thermal equilibrium this must be balanced by the total rate of radiation with emissivity σT_e^4 from the hot surface of area $4\pi R^2 f$. Making this equation for the whole "hot fraction" f at once gives:

$$(1-A) \cdot \frac{L}{4\pi r^2} \cdot \pi R^2 = \sigma T_e^4 \cdot 4\pi R^2 \cdot f$$

Note that the size of the planet cancels through, and we can write the prediction for the <u>uniform</u> T_e of the "hot fraction" f as

$$T_e = \left\{ \frac{(1-A)L}{16\pi\sigma f r^2} \right\}^{1/4}$$

Equation 8.3

This is very similar in form to Equation 8.2, where we considered each part of the surface coming separately to its own equilibrium. Extreme assumptions regarding f might be to put $f \sim \frac{1}{2}$ (perfect temperature smearing over the "day" side of the planet only, perfect cooling on the "right" side) or $f \sim 1$ (perfect temperature redistribution all around the planet, e.g. by convection in a dense atmosphere).

With Eqs. 8.2 and 8.3 as background, now consider the actual experimental results for various planets.

8.5. Observed Planetary Temperatures

PLANET	ALBEDO A	"EXPECTED" T_e			OBSERVED T_e (AVERAGE)		
		Subsolar Point	$f = 1/2$	$f = 1$	Subsolar Point	Day Side	Night Side
Mercury	0.056	624	524	440	624	620	100
Venus	0.72	337	283	238	240	240	240
Earth	0.36	352	296	249	295	290	280
Mars	0.16	305	257	216	250	220	190
Jupiter	0.70	128	107	90	135	125	125 !
Saturn	0.75	90	76	64	105 !	90	never seen
Uranus	0.70	67	55	47	65	58	never seen
Neptune	0.82	47	39	33	56		never seen

First note the values in the columns for the subsolar points. For Mercury, Uranus and Neptune there is fair agreement between the theoretical

expectation and the temperatures derived from infrared measurements. For
Venus, Earth and Mars, the observed subsolar temperatures from the infrared
measurements are lower than "expected" from Equation 8.2. This is reasonable,
as any heat exchange in the planetary environment should act to cool the
subsolar point, which would be the hottest point on the planet according
to Eqn. 8.2.

For Venus, there is good agreement with Eqn. 8.3 if we put $f = 1$,
and indeed the separate measurements of the infrared T_e on the day and
night sides confirm that very effective temperature smearing has taken place.

For Earth, things do not fall into place so easily. The infrared
measurements show little temperature variation between the night and day
sides (280 K versus 290 K); we notice a 10°C temperature drop but the effect
on the emissivity is relatively minor, and this would lead us to expect
fair agreement with the "$f = 1$" prediction for Earth also, corresponding
to good heat exchange around the planet. But in fact $f = 1$ predicts $T_e = 249$ K.
So, given that Earth is near the $f = 1$ situation, we have to explain that
its effective temperature is higher than expected.

For Mars, there is a bigger temperature drop between the day and
night sides, from 220 K to 190 K. We should expect $f \lesssim 1$ to work, and it
is a fair guess ($T_e \sim 220$ K). There may be a detailed discrepancy with
Eqn. 8.3, but it is not as severe as that for Earth.

For Jupiter and Saturn, something is clearly "wrong". The subsolar
temperatures are higher than Eqn. 8.2 predicts. This cannot be due to
"smearing", which would try to cool the subsolar points. This anomaly
was known before Pioneer 10 flew by Jupiter and took measurements on the
night side of the planet; as 135 K is not greatly in excess of 128 K, it
was hoped that the difference might prove to be due to measurement inaccuracy
or to some minor failure of the theoretical model. Pioneer 10 showed however
that the average "nightside" infrared temperature of Jupiter is about the
same (~ 125 K) as the average "dayside" temperature, so we ought to be looking
for a reasonable fit to both temperatures using Eqn. 8.3 with $f = 1$. As this
predicts $T_e = 90$ K, we can only conclude that Jupiter's infrared temperature
is much higher than expected if it is in black-body equilibrium. Note that

the factor (125/90) in T_e is a factor $(125/90)^4$ in infrared luminosity -
Jupiter therefore appears to radiate over 3.5 times as much energy as it
absorbs from the Sun. This is a major problem which we must return to when
we examine Jupiter more closely later.

The data for Saturn indicate that we may be going to encounter an
analogous problem there; but as yet there are no nightside measurements
of Saturn - Pioneer 11 (launched in 1973) reached Saturn in September 1979
and reduction of the data it sent back may tell us whether this planet also
has a "hot" night side and hence an "infrared excess" like Jupiter's.

8.6 The Detailed Radiation Balance of Earth

If we study the temperature balance of the planet we know best, we gain
insight into the complexity of heat balancing in a planetary environment when
there are important deviations from the ideal "black-body" behaviour. Basi-
cally, the complication which arises is that the surface behaves like a black
body as we assumed above, but the atmosphere can be at the same time "dilute"
(transparent) for some wavelengths and "dense" (absorbing) at others. Under
these circumstances the infrared "effective temperature" measures a hybrid
of atmospheric and surface temperature. The surface can be significantly
above an equilibrium black-body temperature due to "heat trapping", a phenomenon
sometimes known (erroneously) as the "greenhouse effect".

The diagram shows the fraction of incident radiation absorbed by the Earth's
atmosphere as a function of wavelength.
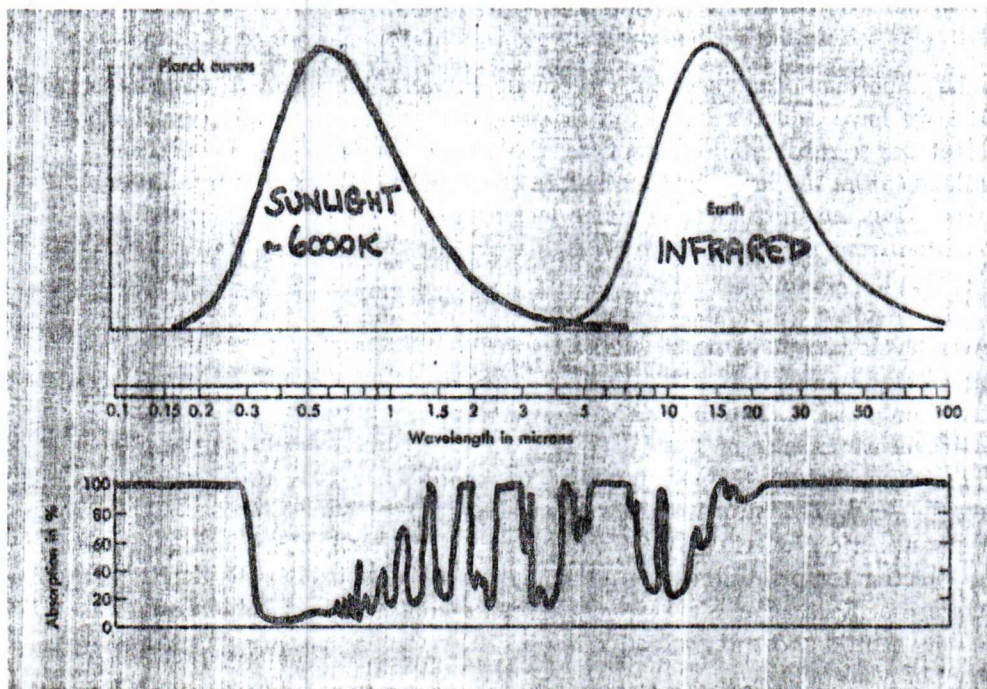
The Earth's atmosphere is fairly transparent through the band of visible wavelengths which contains most of the energy in sunlight, but is <u>strongly</u> absorbing through much of the infrared range of the reradiation from the surface. This is due to the presence of carbon dioxide (CO_2) and water vapour, whose molecules can be excited into vibration and rotation by infrared photons; although these are only trace constituents of the Earth's atmospheric gases, there are enough molecules in any air column for the total absorption to be vitually complete in some wavelength bands, especially longwards of ~ 12 microns and around 7 microns. None of the molecular or atomic species in the Earth's atmosphere absorbs so strongly in the <u>visible</u> wavelength range - the average absorption in the atmosphere at visible wavelengths is only $\sim 21\%$.

VISIBLE SUNLIGHT
100%

REFLECTED 28%

SCATTERED BY CLOUD 47%

REFLECTED 8%

CLOUDS

ABSORBED IN ATMOSPHERE 21%

SCATTERED TO SURFACE 19%

DIRECT TO GROUND 32%

Surface (Land and Water)

INFRARED RADIATION
64% RADIATED TO SPACE AS EARTH'S LONG-WAVELENGTH PEAK

RADIATED FROM SURFACE 112%

RADIATED FROM ATMOSPHERE TO SURFACE 99%

CONVECTION EVAPORATION 30%

Warm Surface
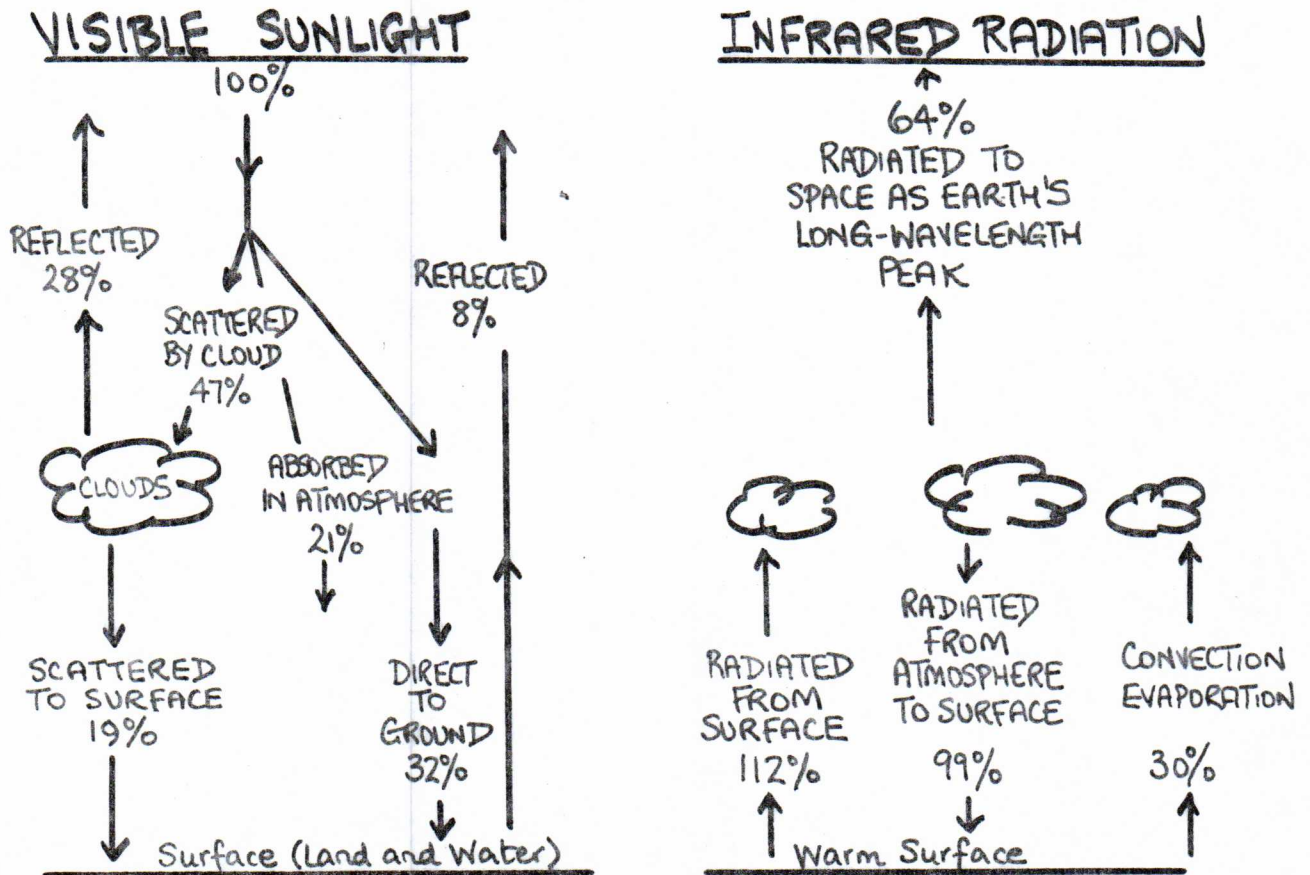
The diagram above shows the average flow of the energy in the sunlight radiation peak and in the infrared peak in the Earth environment, expressed as percentages of the <u>arriving</u> solar radiation at the top of the atmosphere.

Of the 100% arriving at Earth, 28% is reflected back into space from the atmosphere, and 8% from the surface. Thus a total of 36% contributes the average <u>albedo</u> of 0.36 given in the table on p. 30. An observer outside the

Earth would also note an energy flow into space in the <u>infrared</u>, amounting to 64% of the incident solar energy. The <u>whole Earth environment</u> is thus in thermal balance, with 36% of the incident energy being reflected in the visible, and 64% being reradiated in the infrared. To this extent things are as expected in our simple theoretical models. What does <u>not</u> proceed according to our models is the energy exchange <u>within</u> the Earth environment.

Of the 64% of the Solar visible radiation that is absorbed by the Earth, 21% is absorbed into the atmosphere, and 43% is absorbed at the surface. Both of these subsystems (atmosphere and surface) show <u>net</u> energy losses that exactly balance these absorptions (averaged over times longer than a seasonal cycle), but because the atmosphere can absorb much of the infrared radiation from the surface, <u>the surface's radiation does not escape freely into space</u> as our models assumed in Sections 8.3 and 8.4. Instead, it is largely <u>reabsorbed</u> in the atmosphere, which then radiates <u>both out into space and back to the surface</u>.

The reradiation from the atmosphere back to the surface sets up "planetary heat trap", warming the surface to a higher temperature than it would have under the solar visible that reaches it directly. At any instant, there is a greater energy reservoir trapped between atmosphere and surface than would be present if the atmosphere were transparent to the infrared. As a result, the surface does not come to <u>direct</u> equilibrium with the solar influx.

The right-hand figure on p. 36 shows the result quantitatively. The flux of infrared radiation emitted by the Earth's surface amounts now to 112% of the total arriving solar visible and is over 2 1/2 times the flux of solar visible radiation actually absorbed at the ground. In addition, convection and evaporation of surface water remove energy from the surface at a rate equal to 30% of the total arriving solar radiation. This elevated energy output from the surface can be maintained because the downwards flux of infrared radiation from the atmosphere is 99% of the arriving solar flux. The net rate of infrared energy loss from the surface is thus (112 − 99 + 30) = 43% of the incident solar visible, which balances the rate of visible energy gain by absorption. Similarly, the net rate of infrared energy loss from the atmosphere is (99 + 64 − 112 − 30) = 21% of the incident solar visible, which balances <u>its</u> rate of visible energy gain.

The elevation of surface temperature under an infrared-absorbing atmosphere is paralleled by a phenomenon which occurs in greenhouses which are transparent to visible wavelengths but absorbent for infrared. The heat trapping in domestic greenhouses also occurs through impeded convection however, so the

analogy is incomplete. Our effect would best be described as the "atmospheric heat trap effect", but the "greenhouse effect" terminology is well entrenched in scientific literature.

We can conclude that planetary surface temperatures can be determined by the absorption and transmission characteristics of their atmospheres as well as by the input of solar radiation. Also, depending on the atmospheric composition and the wavelength of interest, the "effective temperature" determined by infrared measurements may be an atmospheric temperature, or a surface temperature, or a mixture of both. The "effective temperatures" are therefore only a very approximate indicator of planetary conditions. As we saw at the end of Sec. 8.5 however, they suggest some interesting differences between the heat-transfer properties of the inner planets, and we will explore the meaning of these clues later on.

9. Planetary Atmospheres

In Section 8.5 we saw that the effective temperatures T_e of some planets indicate that there is heat transfer away from the subsolar point, and in Section 8.6 we saw how minor constituents of Earth's atmosphere set up a planetary "heat trap" which raises the surface temperature above that expected from simple radiative equilibrium. Evidently atmospheres may play important roles in controlling planetary environments and in this section we examine some features of planetary atmospheres in more detail.

9.1. The Barometric Law and Scale Height

The pressure and density of planetary atmospheres varies with height above the planetary surface. In an ideal gas, pressure P and density ρ are related to temperature T by

$$P = nkT \qquad \text{Equation 9.1}$$

where n is the number of particles (atoms or molecules) per unit volume, and k is Boltzmann's constant = 1.381×10^{-23} Joules/K. If a gas is comprised entirely of molecules of molecular weight μu kilograms (where u is the atomic mass unit 1.6606×10^{-27} kg) then

$$n = \frac{\rho}{\mu u}$$

where ρ is the density in kg/m^3, and P is in newtons per m^2. Then we can write

$$P = \frac{\rho}{\mu} \cdot \left(\frac{k}{u}\right) \cdot T = \frac{\rho \mathcal{R} T}{\mu} \qquad\qquad \text{Equation 9.2}$$

where $\mathcal{R} = k/u$ is the "universal gas constant" = 8314 Joules/k/kg-mol.

    Planetary atmospheres settle into HYDROSTATIC EQUILIBRIUM, wherein the gas pressure at any level supports the weight per unit area of the gas above that level. This weight per unit area can be thought of as a "gravitational pressure" P_g present in the gas as a result of the gravitational pull of the planet. In hydrostatic equilibrium, each finite volume of gas is neither falling towards the surface under gravity nor diffusing outwards under its own gas pressure (individual atoms/molecules are, of course moving randomly at a variety of velocities).

    Consider a horizontal layer of gas at height h above the planetary surface. Let its thickness be dh, its density $\rho(h)$ and temperature $T(h)$. The volume of this layer under area A is Adh, so the mass under area A is

$$dm = \rho(h)Adh$$

Going from height h+dh to height h in the atmosphere therefore adds

$$dW = dm \cdot g(h)$$

to the weight above a fixed horizontal area A, where g(h) is the gravitational acceleration at height h. The variation of gravitational pressure (weight/unit area) with height must therefore satisfy

$$\frac{dP_g}{dh} = -\frac{dm \cdot g(h)}{Adh} = -\rho(h)g(h)$$

In hydrostatic equilibrium $dP_g/dh = dP/dh$ where P is the gas pressure, so we have

$$\frac{dP}{dh} = -\rho(h)g(h) \qquad\qquad \text{Equation 9.3}$$

and $\quad P(h) = \rho(h)\,\mathcal{R}\,T(h)/\bar{\mu}(h) \qquad\qquad \text{Equation 9.4}$
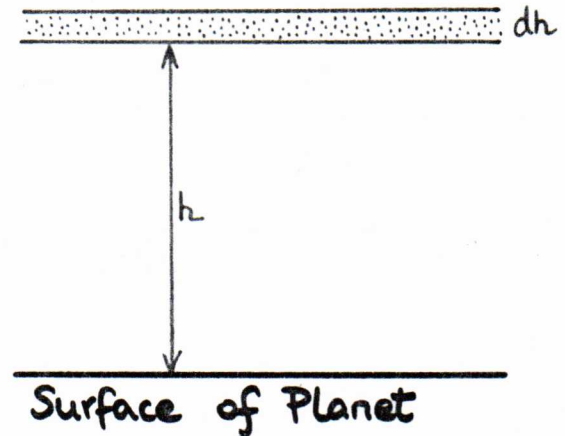
where in Equation 9.4 we use the <u>mean</u> molecular weight $\bar{\mu}(h)$ of the gas mixture in the atmosphere at height h. This mean molecular weight is

$$\bar{\mu}(h) = \sum_i N_i(h)\mu_i$$

where $N_i(h)$ is the relative concentration (or "mixing ratio") of the i-th molecular/atomic species at height h, normalised so that $\sum_i N_i = 1$. Dividing Eqn. 9.3 by Eqn. 9.4 and rearranging, we have

$$\frac{dP}{P} = -\frac{g\bar{\mu}}{\mathscr{R}T}\ dh$$

where it is understood that P, g, $\bar{\mu}$ and T are varying with h in general. This can be integrated to give

$$P(h_2) = P(h_1)\ \exp\left\{-\int_{h_1}^{h_2} \frac{g(h)\bar{\mu}(h)}{\mathscr{R}\ T(h)}\ dh\right\} \qquad \text{Equation 9.5}$$

We know that (to sufficient accuracy)

$$g(h) = \frac{GM}{(R+h)^2}$$

where M is the mass of the planet and R is its mean radius, so the integral in Equation 9.5 will normally be

$$-\frac{GM}{\mathscr{R}}\int_{h_1}^{h_2} \frac{\bar{\mu}(h)}{(R+h)^2 T(h)}\ dh$$

This demonstrates that in a real atmosphere, we must know the variation of atmospheric composition and temperature with height (i.e., all of the $N_i(h)$ for the species in the atmosphere, and the temperature profile $T(h)$), before we can predict the run of pressure and density in the atmosphere. This is a formidable task. <u>In many cases however it turns out that the quantity</u>

$$H(h) = \left[\frac{GM\bar{\mu}(h)}{\mathscr{R}(R+h)^2 T(h)}\right]^{-1} \qquad \text{Equation 9.6}$$

<u>varies only very slowly with h</u> when h << R (i.e., the bulk of the atmosphere is much thinner than the radius of the planet). In that case Equation 9.5 takes the much simpler form

$$P(h_2) = P(h_1) \exp \left\{ -\int_{h_1}^{h_2} dh/H(h) \right\} \sim P(h_1)\, e^{-(h_2-h_1)/H}$$

where H is the average value of H(h) over the height range h_1 to h_2. If we take the surface of the planet (h = 0) as a reference level, then at other heights h we have the <u>exponential</u> law

$$\boxed{P(h) = P(0)e^{-h/H}}$$

This gives a simple meaning to the quantity H--it is the change in height corresponding to a change in atmospheric pressure by a factor of e. Note that <u>if $T(h)/\overline{\mu}(h)$ is nearly constant</u>, Equation 9.7 corresponds to

$$\rho(h) = \rho(0)e^{-h/H} \qquad \text{(from Equations 9.7 and 9.4)}$$

so H would also be the change in height giving a change in <u>density</u> by a factor of e. H is usually referred to as the "SCALE HEIGHT" of a planetary atmosphere; remember that from its definition in Equation 9.6 it will generally vary with h, if only slowly.
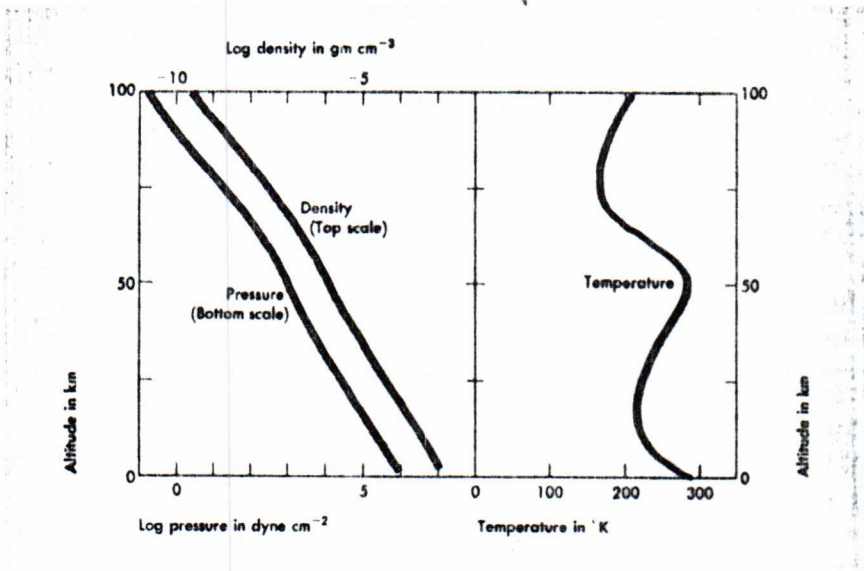
The figure shows actual data for the variation of P,ρ and T with height in the Earth's atmosphere. The height scale (vertical) is linear, but pressure and density are plotted LOGARITHMICALLY in the left-hand diagram, so the exponential law in Equation 9.7 would give a straight line of negative slope. Evidently the exponential law is a <u>fair</u> fit over a range of 100,000 to 1 in

pressure or density despite the temperature fluctuations, and we can use Equation 9.7 as a fair approximation to the true <u>barometric law</u> of Equation 9.5.

For Earth, the average scale height

$$H \sim \frac{\mathcal{R} \overline{T} R^2}{GM\overline{\mu}}$$

Equation 9.8

(putting $(R+h)^2 \sim R^2$ and averaging all other quantities) is about 8.4 km.

9.2. Measuring the Composition of Planetary Atmospheres

There are a number of methods by which we can determine the compositions of planetary atmospheres. By far the best is to send a probe into the atmosphere, carrying a direct sampler feeding a device such as a mass spectrograph which can analyse the atmosphere directly into its constituent species (with ambiguities if two species have the same molecular weight μ_i, or in practice if they have molecular weights which differ by small multiples such as $\mu_j = 2\mu_i$). Such an instrument on a slow descent can give us $N_i(h)$ with associated values of $T(h)$ and $\rho(h)$, permitting a full study of the atmospheric balance. We are rarely so fortunate, although we have incomplete information of this general type from the Venera missions to Venus and from the Viking missions to Mars. Fortunately, there are less direct, but also less precise, approaches that we can use from <u>flyby</u> (rather than <u>entry</u>) spacecraft, or from Earth.

a) Scale Height Method

By studying the refraction and absorption of radiation from distant targets as they pass close to the "limb" of a planet, the variation of density with height in the planetary atmosphere can be determined in a fairly straightforward way. Targets which have been used in practice are the

light from reasonably bright stars, and the radio transmissions from interplanetary probes during flyby missions. By these means it was established that Mercury and the Moon have negligible atmospheres, but that Venus, Mars and Jupiter have appreciable atmospheres with average scale heights of 14.9, 10.6 and $\sim$ 10 km respectively.
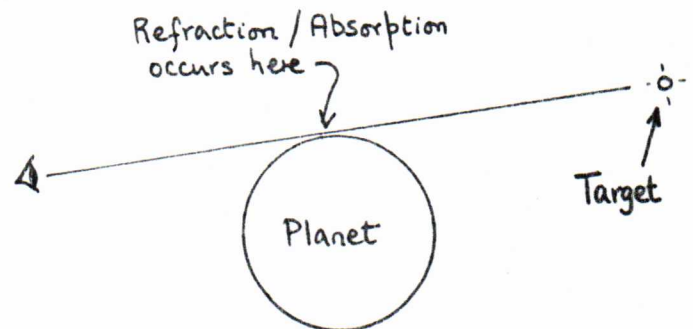
Knowing H, it is possible to deduce $\bar{\mu}$ in the planetary atmosphere from Equation 9.8, by making a reasonable assumption about $\bar{T}$, the mean temperature. If the planetary atmosphere is sufficiently dense, $\bar{T}$ may be close to T_e, the measured effective temperature in the infrared part of the planet's radiation spectrum. For Jupiter, the results give $\bar{\mu} \sim 4$, implying that the bulk of the atmosphere must be very light molecules such as hydrogen and helium. For Venus and Mars, we have more detailed information by other means.

b) The Spectroscopic Method
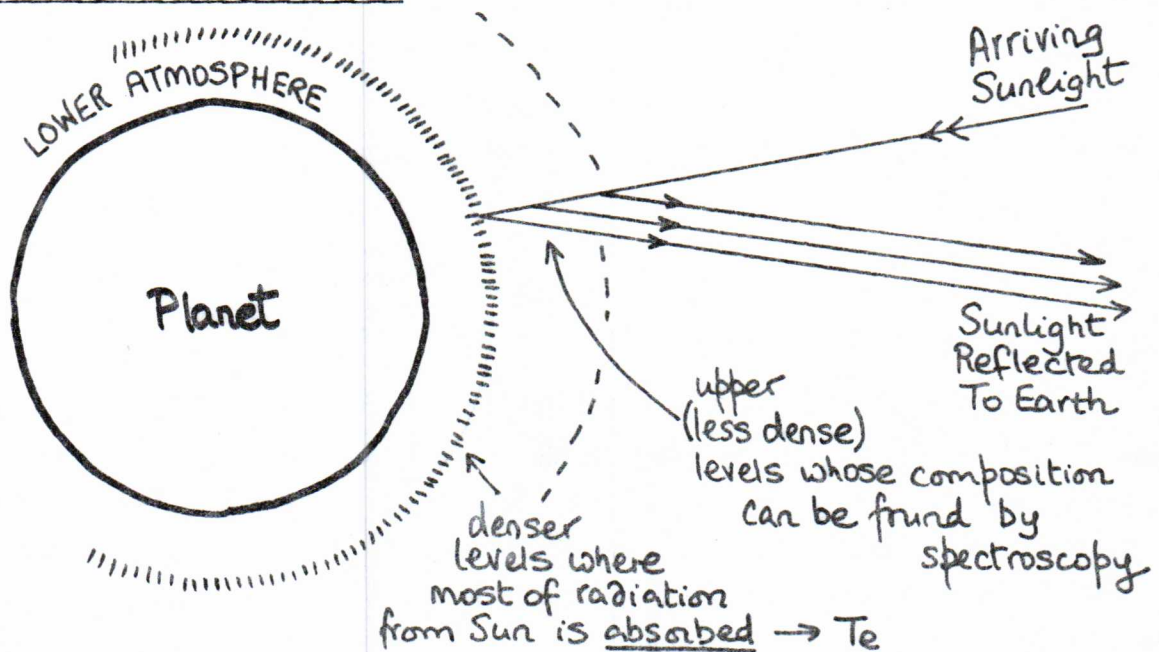
The spectroscopic method consists of identifying discrete-wavelength absorptions which are present in the sunlight reflected from planetary atmospheres but which are not present in the original sunlight spectrum. The reflected sunlight has passed twice through the more dilute upper levels of the planetary atmosphere, where the atomic and molecular energy levels are well-defined because the individual atoms and molecules are not interacting often with their neighbours.

A limitation of this method is that some atmospheric constituents are hard to detect because a) their absorptions occur mainly in wavelength bands for which we find it difficult to build good detectors, b) their absorptions are at wavelengths which are absorbed in the solar spectrum or c) their absorptions also occur in the Earth's atmosphere. Difficulty (a) is decreasing

with the advent of sensitive detectors for infrared and ultraviolet wavelengths where many molecules have strong emission and absorption bands. The effects of the Earth's atmosphere (c) are best reduced by making observations of the planets from above our atmosphere (for the ultraviolet and infrared) or in the visible by making use of the Doppler Effect which shifts wavelengths in planetary spectra away from the wavelengths of the same species in the Earth's atmosphere. The difficulty which cannot be got around however is (b)--the fact that sunlight itself contains many absorption features produced in the upper levels of the solar atmosphere. This makes it difficult to get good measurements of any atom or molecule in a planetary atmosphere if that species is also found in the solar atmosphere. This is not a problem for most molecules (which dissociate at solar temperatures) but is a serious difficulty if we are interested in some atomic species. It is a particularly acute problem for Jupiter, whose light atmosphere may have an atomic composition that is quite similar to that of the Sun.

9.3. Atmospheric Composition Data

For Mercury, the Mariner 10 probe detected a trace of helium, at an atmospheric pressure 2×10^{-15} of that at Earth surface. Only upper limits could be obtained for atomic or molecular hydrogen, and for oxygen, argon, neon and carbon dioxide. This "atmosphere" is so tenuous that it has no detectable effect on the planetary environment, and the surface of Mercury is in direct equilibrium with solar radiation.

For Venus, Venera (USSR), Mariner and Pioneer (U.S.) spacecraft documented a deep, dense atmosphere in which there is a deck of yellow cloud between 67 and 49 km above the solid surface. At the cloud level the atmospheric pressure is about 0.1 Earth atmospheres; at the surface the pressure is about 80 Earth atmospheres. The bulk atmosphere is predominantly carbon dioxide (97%), with variable traces of water vapour (0.1-2%) and < 3% of argon and nitrogen combined. Some very minor constituents have been identified: carbon monoxide at 50 parts per million (ppm), helium at 10 ppm, hydrogen chloride at 0.4 ppm, hydrogen flouride at 0.01 ppm and some sulphur dioxide and neon. The cloud deck is a mixture of sulphuric acid, water droplets and sulphur droplets; the water vapour is almost completely absent above this cloud deck, which acts as an effective drying agent. The dense atmosphere probably produces the heat transfer noted in Sec. 8.5, by heat convection

in the atmospheric circulation.

For Mars, the Viking lander missions gave the most definite data on atmospheric composition. The atmospheric pressure at Mars surface is about 7.7×10^{-3} Earth atmospheres and the major atmospheric constituent (95%) is again carbon dioxide, followed by nitrogen (2-3%) and argon (1-2%). Oxygen (0.1 - 0.4%) and water vapour (0.01 - 0.1%) are also significantly present. The thin atmosphere allows the night side of Mars to cool rapidly compared with Earth's.

Among the inner planets with appreciable atmospheres, Earth is the "odd planet out", as both of its neighbours have carbon-dioxide-rich atmospheres. The carbon-dioxide content of Earth's atmosphere is variable but approximately 0.03%; the major constituents are nitrogen (78.08%), oxygen (20.95%), argon (0.93%) and water vapour (up to 1%). The reasons for this "Earth anomaly" are intimately associated with the development of life on Earth, as we shall see later.

For Jupiter, the compositional mix is uncertain in detail as it is not yet clear at which levels various methods have effectively measured the atmospheric constituents. The species that have been detected are molecular hydrogen at an effective partial pressure of 67 Earth atmospheres, helium at an effective pressure of 34 Earth atmospheres, methane at 0.045 Earth atmospheres, and variable amounts of ammonia, at ~ 0.01 Earth atmospheres. Water vapour was detected in 1975, at $\leqslant 2 \times 10^{-5}$ Earth atmospheres partial pressure. Sundry hydrocarbons, including acetylene (C_2H_2) and ethane (C_2H_6), as well as phosphine (PH_3) and carbon monoxide (CO) have also been detected in small amounts.

For the distant planets Saturn, Uranus and Neptune, the information is scanty, but methane and molecular hydrogen appear as constituents of all their atmospheres, and ethane has been found in Neptune's atmosphere.

The coming years of exploration of planetary atmospheres by Pioneer and Mariner missions to Venus and Jupiter and by further Viking missions to Mars, will refine these data considerably, but some basic features are now clear:

1) Venus and Mars have primarily CO_2 atmospheres, unlike that of Earth, which is conspicuously lacking in carbon gases in comparison with its near neighbours.

2) All of the inner planetary atmospheres are hydrogen-poor in comparison with Jupiter, whose atmosphere not only contains hydrogen as a primary constituent, but also contains very hydrogenous compounds such as methane,

ammonia and the hydrocarbons in place of carbon dioxide and nitrogen in the inner Solar System atmospheres.

3) The division into "Terrestrial" and "Jovian" planets could also be made on criteria of oxidation state of the planetary atmospheres, if Jupiter is indeed typical of the outer planets.

9.4. Compositional Segregation in Earth's Atmosphere

In Section 9.1 reference was made to the fact that the composition of a planetary atmosphere might vary with height, so that the mixing ratio N_i of a given species is $N_i(h)$. We will review the main features of this variation in our own atmosphere to illustrate the phenomena which we should look for in other atmospheres.

First, chemical changes may take place when the atmosphere absorbs the ultraviolet wavelengths present in sunlight. On Earth, the nitrogen component of the Earth's atmosphere is little affected by this, but the oxygen, present at ground level as molecular oxygen O_2, can undergo photodissociation by ultraviolet light, producing highly reactive oxygen atoms

$$O_2 \xrightarrow{uv} O + O$$
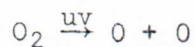
which can combine with normal oxygen molecules to form the triatomic ozone molecule O_3

$$O_2 + O \rightarrow O_3$$

Three factors control the rate of photodissociation of oxygen in our atmosphere--1) the number of ultraviolet photons available, 2) the number of oxygen molecules available, and 3) the efficiency with which a given photon can cause dissociation, which varies with the wavelength of the photon. Molecular oxygen is most strongly dissociated by radiation at wavelengths near 1450 $\overset{o}{A}$; at longer and shorter wavelengths the efficiency of its photodissociation decreases.

The second factor--number of available molecules--will be controlled mainly by the exponential law, so we can expect that at great heights there will be few photodissociation products because there are few molecules to be dissociated. Lower in the atmosphere, the number of photodissociation

products increases exponentially, until their number becomes limited by the first of our three factors. At greater depths in the atmosphere, the number of photodissociation products must decrease because most of the efficient photons available have already caused dissociations higher up and have thereby been removed from the sunlight reaching lower levels.

The result is the formation of "layers" of atomic oxygen and ozone in the upper atmosphere, as indicated schematically below:
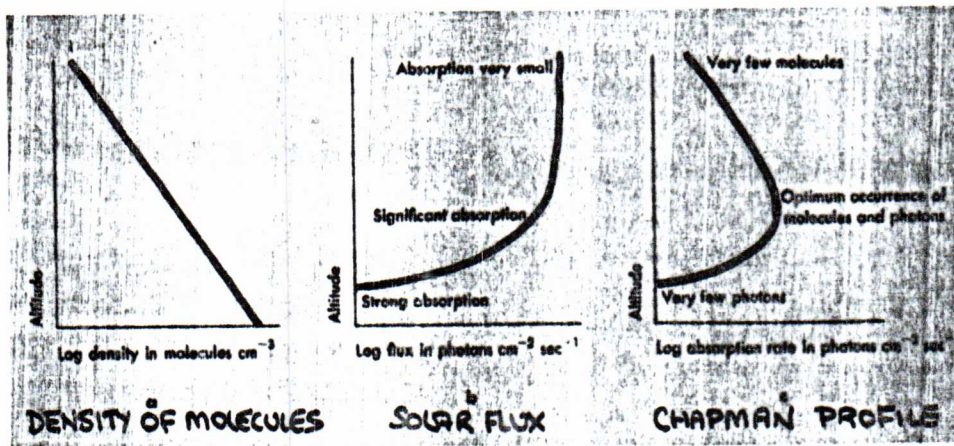
The Chapman profile. The rate of absorption of solar photons is proportional to the product of the density of absorbing molecules and the flux of photons (the logarithmic scales add: c = a + b — constant)

The characteristic "nose-shaped" curve of dissociation product density versus height is known as the "Chapman profile". It has the mathematical form

$$N(h) = N(h_o) \exp \{1 - z - \sec\chi \, e^{-z}\}/2$$

where $z = \dfrac{h-h_o}{H}$ and χ is the angle of incidence of the sunlight measured away from the vertical and h_o is the height at which the number of dissociation products $N(h)$ has its maximum value when $\chi = 0°$. H is the scale height of the atmosphere (assumed constant). For those who may be interested, the derivation of the Chapman profile is given in the notes in the Ring Binder.

The peak height of the atomic oxygen is ~ 100 km, and that of the ozone is ~ 25 km, in the Earth's atmosphere. The Chapman profile takes account of the fact that dissociation products can recombine to reconstitute the original molecule, and that this effect will be proportional to the square of the density

of the products, as a collision between two products is required to initiate a recombination (in fact a third "neutral" body is also needed in practice to carry away the excess energy released).

The observed distributions of atomic oxygen and of ozone in our atmosphere do not follow the simple theoretical profile (see diagram at right), due to mixing and diffusion. Mixing tends to make the relative concentrations of different species the same at all heights, i.e., to make $\bar{\mu}(h)$ more nearly constant with h. Diffusion tends to create an atmosphere in which each species obeys a barometric law with its own characteristic scale height H_i corresponding to putting $\bar{\mu} = \mu_i$ in our Equation 9.8.
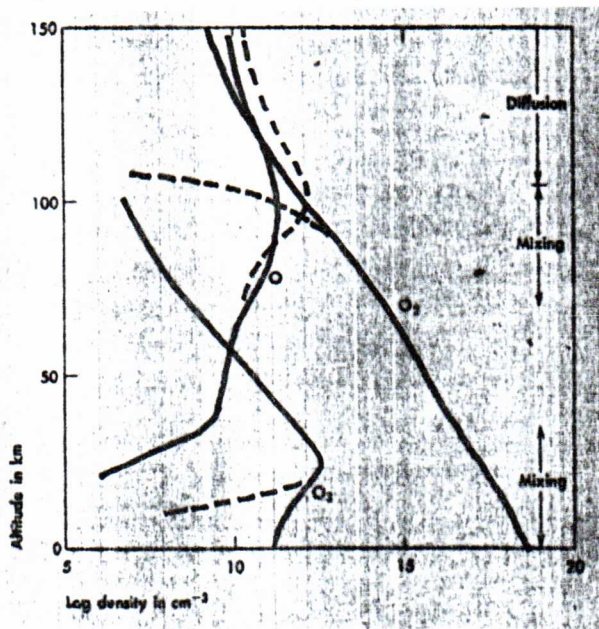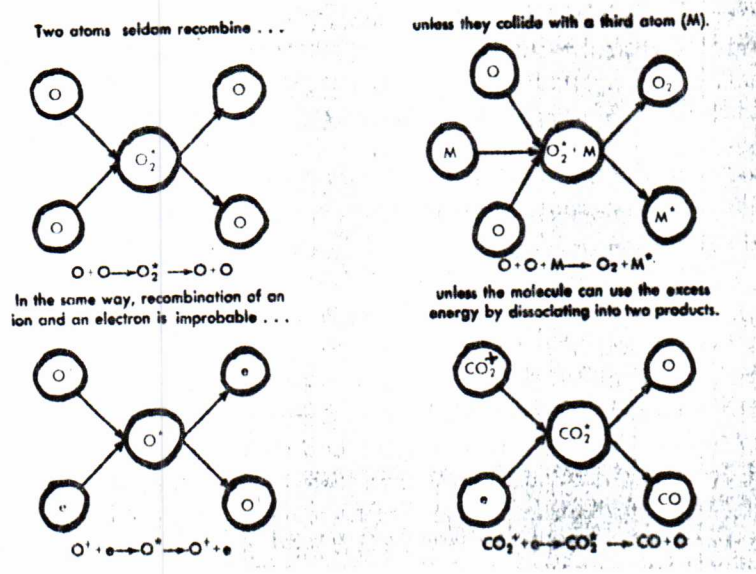

Dotted curves — Chapman profiles
Solid curves — Actual data

Where several physical processes compete, the fastest-acting wins. The photodissociation-recombination time scale is fastest where the concentrations of the reactants are highest, while mixing is governed by winds and meteorological variations which do not depend explicitly on density and thus on height. In general, mixing dominates at higher altitudes and photochemistry at lower altitudes, where concentrations are higher. Diffusion is competing with mixing everywhere and in the Earth's atmosphere becomes dominant above $\sim$ 105 km. We therefore find "layers" of dissociation products such as ozone at about the heights predicted by the Chapman profiles, but distorted by mixing. At very high altitudes we begin to see the composition segregation imposed by diffusion, with the species of lowest molecular weight μ having the largest scale heights H and thus predominating in the upper atmosphere.

Two atoms seldom recombine . . .   unless they collide with a third atom (M).

$O + O \rightarrow O_2^* \rightarrow O + O$

In the same way, recombination of an ion and an electron is improbable . . .   unless the molecule can use the excess energy by dissociating into two products.

$O + O + M \rightarrow O_2 + M^*$

$O^+ + e \rightarrow O^* \rightarrow O^+ + e$ $CO_2^+ + e \rightarrow CO_2^* \rightarrow CO + O$

9.5 Temperature Segregation and Ionization in Earth's Atmosphere

The uppermost levels of the atmosphere are exposed to the solar ultra-violet radiation without "screening" by the formation of photodissociation products. Above about 50 km, an important process is <u>photoionization</u>, whereby ultraviolet photons knock electrons out of neutral atoms to form <u>ions</u>, e.g.
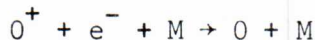
$$O + \text{photon} \rightarrow O^+ + \text{electron} (e^-)$$
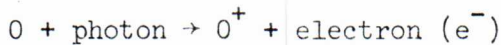
The <u>recombination</u> processes by which electrons are recaptured involve third bodies, which help to take up the excess energy originally deposited by the incident photon. At levels below $\sim$ 100 km, there are enough molecules around to assist recombination by processes such as

$$O^+ + e^- + M \rightarrow O + M$$

where the oxygen atom and the molecule M both take up the excess energy released at recombination (as kinetic energy). At high altitudes, where all densities are lower, three-body collisions are too improbable and recombination has to occur by a tortuous two-body route:

$$O^+ + O_2 \rightarrow O_2^+ + O$$

$$O_2^+ + e^- \rightarrow O + O$$
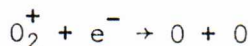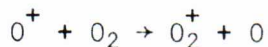
The <u>oxygen molecular ion</u> O_2^+ can dispose of the excess energy from the original photon by dividing it among the kinetic energies of its <u>two</u> dissociation products. This allows it to dispose of the energy in a way that is not avail-able to a simple oxygen ion, which is forced to

$$O^+ + e^- \rightarrow O^+ + e^-$$

in most collisions as there is no easy "sink" for the ionisation energy.

The result of these processes (and of others involving other ionic species) is the production in the upper atmosphere of an <u>ionised region</u> which is also <u>hot</u>. The ion density typically peaks around 300 km height and the ion density profiles are similar to the Chapman profile. The ionised levels are known as the <u>IONOSPHERE</u>.

The kinetic energy given to the various species during the recombination of the ions causes a very considerable increase in the local temperatures, up to $\sim$ 1000 K at $\sim$ 300 km.
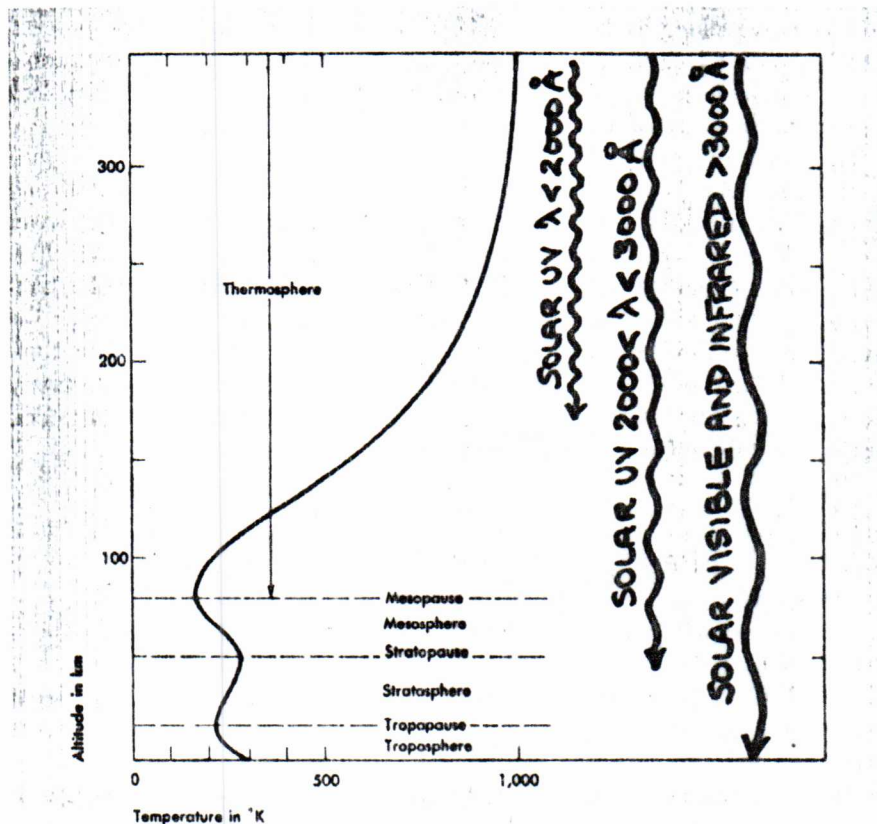
The variation of temperature with altitude at middle latitudes in the Earth's atmosphere. Layers of the atmosphere are called troposphere, stratosphere, mesosphere, and thermosphere; these are separated by features called tropopause, stratopause, and mesopause. The high temperature regions of the atmosphere result from the absorption of solar radiation at different wavelengths, as shown on the right.

The variation in temperature with height is shown in the preceding diagram. Note that by ~ 100 km the <u>density</u> in the atmosphere has fallen to $\sim 10^{-6}$ of its value at Earth-surface, so the biggest temperature variation affects only a very small fraction of the atmospheric mass.

It is conventional to divide the atmosphere into four components on the basis of the observed variations in temperature. The THERMOSPHERE, above ~ 80 km, has the temperature maintained by photoionization, mainly by wavelengths < 1000 Å in the arriving sunlight. Below that is the MESOSPHERE, reaching to ~ 50 km, where the temperature reaches a local maximum due to strong absorption by ozone of solar ultraviolet at wavelengths between 2000 and 3000 Å. The layer below 50 km is the STRATOSPHERE, which is considered to extend down to a level at ~ 15 km where the temperature reaches a local minimum. Below ~ 15 km there occurs the major absorption of the visible solar spectrum, producing the temperature equilibrium discussed in Section 8.6. The level below ~ 15 km is known as the TROPOSPHERE. The infrared effective temperature T_e is an average of temperatures in the troposphere and at the surface.

9.6 The Escape Velocity

Consider a small mass m moving with velocity v at distance d from a mass M, which we will suppose to be fixed in position. The <u>kinetic energy</u> of the small mass is

$$K = \frac{1}{2} mv^2$$

while its gravitational potential energy U_G is equal to the work that would be done bringing m from infinity to distance d from M:

$$\text{i.e. } U_G = \int_{\infty}^{d} \frac{GMm}{r^2} dr = -\frac{GMm}{d} \qquad \text{Equation 9.9}$$

The total energy of the mass m is $E = K + U_G$

$$E = \frac{1}{2} mv^2 - \frac{GMm}{d} \qquad \text{Equation 9.10}$$

Now, <u>if m is not acted on by forces other than M's gravity</u>, conservation of energy requires that E = constant during m's motion near M. If m is later at another distance d' from M, it must then be moving with a velocity v' such that

$$\frac{1}{2} mv'^2 - \frac{GMm}{d'} = E = \frac{1}{2} mv^2 - \frac{GMm}{d}$$

The mass m can escape from M's gravitational influence if it can get to $d' = \infty$ with <u>finite</u> residual velocity v'. Evidently the criterion for this to happen is

$$\frac{1}{2} mv'^2 \quad \text{just} > 0 \quad \text{at } d' = \infty$$

i.e. E just > 0

i.e. $\frac{1}{2} mv^2 > \frac{GMm}{d}$

$$v^2 > \frac{2GM}{d} \qquad\qquad \text{Equation 9.11}$$

Note that the criterion for escape is a criterion on the SIGN of the total energy E. If E is positive, escape is possible. If E is negative, escape is impossible.

The velocity

$$\boxed{v_e = \sqrt{2GM/d}} \qquad\qquad \text{Equation 9.12}$$

which is the critical case according to Equation 9.11 is known as the ESCAPE VELOCITY from distance d away from mass M. Note that it is specific to distance d.

Now suppose that M is a spherical planet of radius R. d is then the distance of m from the centre of M (see the diagram to the right).
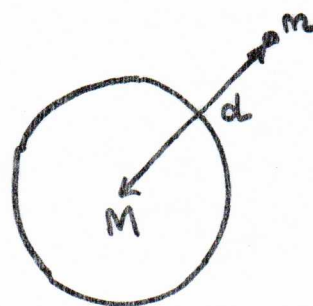
Now note that $v > v_e$ does not ensure escape. There is an additional <u>ballistic</u> criterion that m must not be moving with velocity v such that it will later strike the surface of the planet. If it does strike the surface, forces other than M's gravity may come into play (e.g. inelastic forces due to M's surface constitution) and so E is not necessarily conserved in m's future. Equation 9.11 must be satisfied if m is to have a chance to escape, but $v^2 > 2GM/d$ does not <u>guarantee</u> escape; m must also be moving in the right direction.

Now apply the same reasoning to a particle of gas in a planetary atmosphere.

If it is travelling with velocity v at a height h above the surface, it will escape from the planet's atmosphere if 1) $v^2 > \sqrt{2GM/(R+h)}$ and 2) it is travelling in a direction that will take it away from the planet's surface and 3) it does not collide with any other component of the atmosphere, e.g., another gas molecule or a Boeing 747 so that its velocity is altered in such a way as to violate (1) or (2).

9.7 Critical Height and the Exosphere

Condition (3) above requires that the escaping atom/molecule is initially at a level of the atmosphere where collisions between atoms/molecules are infrequent, which means that it must be at a height where the density is "sufficiently low". How low is "sufficient"? Suppose an atom of radius a is moving horizontally at a height h where the density is $\rho(h)$. If the gas particles are in fact randomly distributed we can determine the average collision rate by assuming them to be uniformly distributed. Then after the atom has travelled a distance d in the gas, it will have collided with any other similar atoms if their centres lie inside the volume $4\pi a^2 d$ (see diagram). Thus the expected number of collisions n_c in distance d would be, on average,

$$n_c = 4\pi a^2 d \cdot n(h)$$

gas particles

where $n(h)$ is the number density of atoms in the gas at height h, i.e.

$$n(h) = \frac{\rho(h)}{\overline{\mu}(h)u}$$
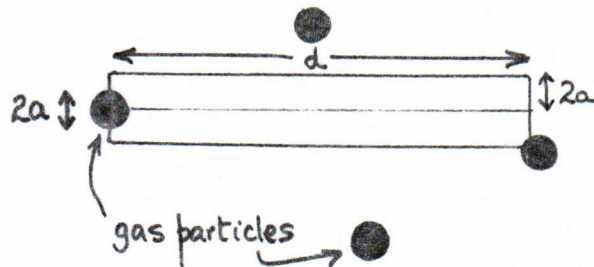
The distance d for which we can expect one collision ($n_c = 1$) is called the mean free path λ at height h, $\lambda(h)$. Clearly

$$\lambda(h) = \frac{1}{4\pi a^2 n(h)}$$

$$= \frac{1}{4\pi a^2 n(0)} e^{h/H} \qquad \text{in an exponential atmosphere.} \qquad \underline{\text{Equation 9.13}}$$

Equation 9.13 tells us how λ increases with height.

Now suppose an atom travels _vertically_ a distance dh. The expected number of collisions n_c in travelling height dh is $4\pi a^2 n(h)dh$. If the atom then travels from height h_1 to height h_2, over a distance in which $n(h)$ varies appreciably, the expected number of collisions would be:

$$n_c = \int_{h_1}^{h_2} 4\pi a^2 n(0)e^{-h/H}\, dh \qquad\qquad \text{Equation 9.14}$$

in an exponential atmosphere. _If h_1 is high enough_, then after the particle has travelled a reasonable fraction of $\lambda(h_1)$, it reaches a level h_2 where $\lambda(h_2) \gg \lambda(h_1)$ because of the variation in Equation 9.13, so that it can still continue upwards without making a collision. If h_1 is sufficiently high for _escape_, then the particle must be able to reach $h_2 = \infty$ with n_c still $\underset{\sim}{<} 1$ in Equation 9.14. This means that the CRITICAL HEIGHT h_c at which escape becomes possible is given by putting $n_c = 1$, $h_1 = h_c$ and $h_2 = \infty$ in Equation 9.14:

$$1 = \int_{h_c}^{\infty} 4\pi a^2 n(0)e^{-h/H}\, dh = 4\pi a^2 H n(0)e^{-h_c/H}$$

i.e. $n(h_c)$ must $= \dfrac{1}{4\pi a^2 H}$ at the critical height h_c, so that h_c is determined by the relation

$$\boxed{H = \frac{1}{4\pi a^2 n(h_c)} \qquad (= \lambda(h_c))} \qquad\qquad \text{Equation 9.15}$$

Equation 9.15 shows us that _the CRITICAL HEIGHT h_c is the level where the mean free path of a particle travelling horizontally is just equal to the scale height H of the atmosphere._ The atmospheric levels at $h > h_c$ are those from which particle escape is possible in the face of collisions, and these levels of a planet's atmosphere _are called the EXOSPHERE._

Obviously the critical height depends on the species of escaping particle to some extent; h_c is usually in the diffusion-dominated upper levels, where each different species μ_i has its own scale height H_i. It also depends on atomic/molecular size through the a^2 term in Equation 9.15. For oxygen and nitrogen in the Earth's atmosphere, h_c is ~ 550 km, i.e. well into the thermosphere where the ambient temperature T is significantly higher than the infrared effective temperature T_e.

† 9.8 The rate of escape from an atmosphere in thermal equilibrium

    We now have a proper criterion for escape of an individual particle from a planetary atmosphere: any particle travelling upwards at height h in the exosphere of a planet of mass M and radius R with velocity $v^2 > 2GM/(R+h)$ (from Equation 9.11) can escape.

    What remains is to relate this criterion to the environmental factors that determine the particle velocities. Theory predicts, and experiment verifies, that a gas whose particles are of mass m has a characteristic distribution of velocities when in thermal equilibrium at temperature T. This Maxwellian velocity distribution is such that the fraction f of all particles whose velocities lie between v and v + dv is

$$f(v)dv = 4\pi \left(\frac{m}{2\pi kT}\right)^{3/2} v^2 e^{-mv^2/2kT} \, dv$$

Equation 9.16

The form of this distribution is shown in the diagram below.
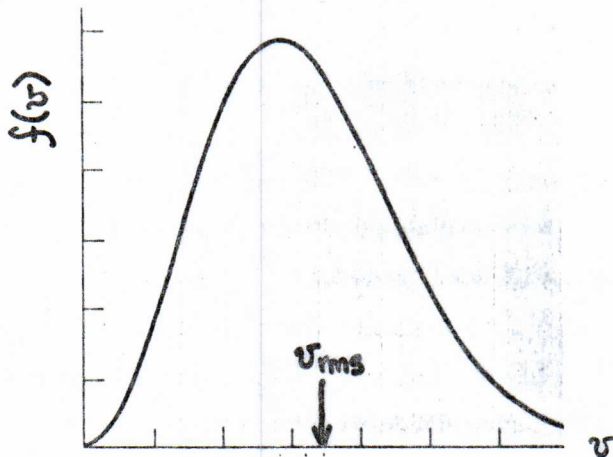
    It is a property of the Maxwellian velocity distribution that the average kinetic energy of all particles in a gas in thermal equilibrium at temperature T is

$$\overline{K} = \frac{1}{2} \overline{mv^2} = \frac{3}{2} kT$$

Equation 9.17

where k in Equations 9.16 and 9.17 is again Boltzmann's constant. The mean square velocity of particles of mass m in equilibrium at T is thus

$$\overline{v^2} = \frac{3kT}{m}$$

and so the root mean square velocity $v_{rms} = \sqrt{\overline{v^2}}$ (a convenient measure of "typical velocity") is

$$v_{rms} = \sqrt{\frac{3kT}{m}}$$

Equation 9.18

Can we simply substitute v_{rms} for v in our criterion for escape, and have a useful criterion for retention of a molecular or atomic species in a given planetary atmosphere? It turns out that this is too crude an approximation because the exospheric temperature is being maintained by solar radiation and so the solar energy supply maintains the shape of the Maxwellian velocity distribution. This ensures that there are always some particles with v >> v_{rms}. Even if only such faster-than-average particles escape, the average velocity of the remaining particles decreases, i.e. escape cools the exosphere. The absorption of sunlight offsets this cooling by continuously increasing the energy of the residual particles to maintain Equation 9.16 by compensating for the energy lost by escape. This means that the fraction of particles with velocities greater than v_{rms} remains constant so long as there is a supply of solar radiation. This in turn means that escape is significant even when v_{rms} is well below the escape velocity in the exosphere, because of the high-velocity "tail" in Equation 9.16. What is variable is the rate of escape. If only a very few very high-velocity particles can escape at any one time, the rate of escape will be low. If a large fraction of the particles can escape at any one time, escape will be rapid.

It is difficult to make exact predictions of the rate of escape of a given type of particle from a real atmosphere. The rate of escape may be limited in practice by the rate of diffusion of potential escapees into the planetary exosphere; escape may be encouraged for certain atomic species because they can participate in processes such as those discussed in Section 9.5 while others cannot, due to differences in the efficiency of ultraviolet absorption by different chemical species. It is useful to take a look at what would happen in an atmosphere that is fully in thermal equilibrium however, as this over-simple situation clearly illustrates the very strong dependence of the rate of escape on the ratio v_{escape}/v_{rms} for a given atomic species.

Suppose we consider an exosphere containing particles of mass m at temperature T. Because it is an exosphere, all particles travelling outwards with velocities v > v_e will escape. The particles will be travelling in random directions so the fraction travelling outwards at angles to the vertical between θ and $\theta + d\theta$ will
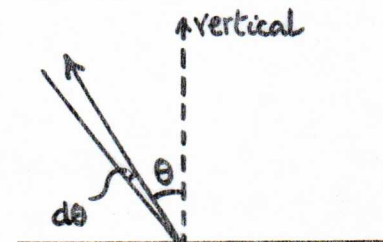
be given by $\frac{1}{2}$ sinθ dθ.   The rate at which particles of velocity v cross unit horizontal area per unit time will be n(h,v)vcosθ, where n(h,v) is the number of particles at height with <u>total</u> velocity v, and vcosθ is the <u>upwards component</u> of the total velocity. The <u>total number of escaping particles crossing unit horizontal area in unit time</u> is therefore

$$- \frac{dn(h)}{dt} = 2\pi \left(\frac{m}{2\pi kT}\right)^{3/2} n(h) \int_{v_{esc}}^{\infty} \int_{0}^{\infty} e^{-mv^2/2kT} v^3 \cos\theta \sin\theta \, dv \, d\theta$$

where we have now included all outbound particles with velocities from v_e to ∞ in the Maxwellian distribution at temperature T. If we put $n(h) = n(0)e^{-h/H}$ and $H = kTR^2/GMm$ from Equation 9.8, then it can be shown, after some mathematical manipulation, that

$$\dot{n} = \frac{dn(h)}{dt} = - \frac{kTn(0)}{mv_e} \left[\frac{mv_e^2}{3kT}\right]^{3/2} e^{-mv_e^2/2kT}$$

gives the rate of escape provided that diffusion upwards from below the exosphere can supply enough particles. As the <u>total</u> number of particles of mass m in the atmosphere is

$$n = \int_{0}^{\infty} n(0)e^{-h/H} \, dh = n(0)H$$

we have that

$$\frac{\dot{n}}{n} = - \frac{kTn(0)}{mv_e} \cdot \frac{1}{n(0)H} \left[\frac{mv_e^2}{3kT}\right]^{3/2} e^{-mv_e^2/2kT}$$

Putting $H = kTR^2/GMm$ (Eqn. 9.8) and $v_e = \sqrt{2GM/R}$ (R+h ∿ R in Eqn. 9.12) we find

$$\frac{\dot{n}}{n} = - \frac{v_e}{2R} \left[\frac{mv_e^2}{3kT}\right]^{3/2} e^{-mv_e^2/2kT}$$

which means that the number of particles in the atmosphere will decrease exponentially as

$$n = n_{(t=0)} \, e^{-t/\tau}$$

where $\quad \tau = \dfrac{2R}{v_e} \, s^{-3} \, e^{3s^2/2}$

and $\quad s = v_e/v_{rms} = \sqrt{\dfrac{2GMm}{3kTR}}$

Equation 9.19

Equation 9.19 is the key equation in understanding atmospheric escape. It says that the number of particles in the atmosphere declines by a factor e in a time τ that is _critically_ dependent on the ratio $s = v_e/v_{rms}$. The exponential $e^{3s^2/2}$ is a _very_ rapid function of s; its value for s = 1 is 4.48, for s = 3 its value is 7.29×10^5, and for s = 5 its value is 1.93×10^{16}. More realistic theories of escape (which do not assume that the atmosphere is isothermal with constant scale height) produce changes of order 2 in the expected value of τ, but the basic result can be illustrated by taking the Earth and Jupiter as examples in Equation 9.19:

1/e Lifetimes of Atmospheric Constituents with Given Values
of $s = v_e/v_{rms}$

	a) On Earth	b) On Jupiter
S = 1	$\sim 1^{hr} \, 25^{min}$	$\sim 3^{hr}$
2	$\sim 15^{hr} \, 50^{min}$	$\sim 1^{d} \, 9^{hr}$
3	~ 1 year	~ 2 years
4	1.5×10^4 years	3.1×10^4 years
5	5.7×10^9 years	1.2×10^{10} years
6	4.8×10^{16} years	9.8×10^{16} years

Clearly the values for $s \leq 3$ are very short compared with human historical time-scales, while the values for s = 6 are about a million times longer than the expansion time of the universe, as measured by the recession of the galaxies. Given that the estimates of the "age of the Solar System" from

radioactive dating are $\sim 5 \times 10^9$ years (see Section 10), we can say that if v_e/v_{rms} for a species of atom or molecule is $\lesssim 4.5$, then it will not be retained in that atmosphere for periods of time that are significant in Solar System history.

Earth's exosphere (at T = 1000 K) can retain all elements but hydrogen, for which τ is $\sim 10^5$ years. At ground level temperatures, all elements could be retained. We shall see later that atmospheric escape is much more significant for other bodies in the Solar System.

10) Radioactivity and the Time Scale of Planetary Development

The process of radioactive decay is related to the probability that a particle within an atomic nucleus can escape from the nucleus against the short-range nuclear attractive force. The potential energies of nuclear particles in the presence of this force are so large compared with other energy stores in nature that it is very unlikely that environmental factors outside atomic nuclei influence the probability of escape at all significantly. Under these circumstances the rate of radioactive decay among N nuclei of a given species is proportional to the number N itself, and the constant of proportionality, λ, is fixed by the nuclear physics and is not sensitive to environmental factors, i.e.

$$\frac{dN}{dt} = - \lambda N$$

so that the variation of N with time is

$$N(t) = N(0)e^{-\lambda t}$$

Equation 10.1

Equation 10.1 implies that N(t) halves itself during every time interval

$$\tau_{1/2} = \frac{\ln 2}{\lambda} = \frac{0.6932}{\lambda}$$

Equation 10.2

where $\tau_{1/2}$ is known as the half-life of the particular radioactive process.

If a radioactive decay transforms a parent nuclear species A into a product species B:

$$A \rightarrow B + \text{debris}$$

then
$$N_A(t) = N_A(0)e^{-\lambda t} \qquad \text{Equation 10.3}$$

$$N_B(t) = N_B(0) + N_A(0)(1 - e^{-\lambda t}) \qquad \text{Equation 10.4}$$

where $N_A(t)$ and $N_B(t)$ are the numbers of A and B nuclei in a given sample at time t and $N_A(0)$ and $N_B(0)$ are the "original" numbers. In principle we can combine chemical separation methods and mass spectroscopy to determine the ratio

$$R = \frac{N_B(t)}{N_A(t)}$$

now in any mineral sample. Using Equations 10.3 and 10.4, this ratio can be written as

$$R = e^{\lambda t}\left\{\frac{N_B(0) + N_A(0)}{N_A(0)}\right\} - 1$$

so that
$$\boxed{t = \frac{1}{\lambda}\ln\left\{\frac{N_A(0)(R+1)}{N_A(0) + N_B(0)}\right\}} \qquad \text{Equation 10.5}$$

This means that if we know the decay constant λ from laboratory studies of the $A \rightarrow B$ decay, the measured B:A ratio R in a given rock sample now would tell us the "age" t of the sample <u>if we knew the "initial" ratio $N_B(0)/N_A(0)$.</u> To make the mathematical statement of Equation 10.5 meaningful, we must be clear what is meant by the "age" of the sample.

The term "age" applied to a rock must be interpreted as the time elapsed since the last event during which chemical fractionation was possible throughout its volume, i.e. since the last process that could <u>reset</u> the ratio $N_B:N_A$ by bringing the rock into contact with a "reservoir" containing A and B or into contact with a possible "sink" which could absorb A or B from the rock. In practice most such fractionation processes involve significant heating of rocky material, and the term "age" will be almost the same as "time since last complete melting in the presence of other materials". In the case of a terrestrial igneous rock, this will be the time since last recrystallization from a magma,

unless the sample has been sufficiently stressed or reheated to have undergone significant local modification since this time. The practical procedures for selecting samples which have been chemically "closed" since their last recrystallization, or for correcting for minor "open-ness" in the system, are beyond the scope of this course, but can be studied, for example, in "Radiometric Dating for Geologists", edited by E.I. Hamilton and R.M. Farquhar (call number QE508.H26).

10.1 Radioactivities of Importance to Planetary Chronology

To be of assistance in planetary chronology, a radioactive decay process should have a half-life $\tau_{1/2}$ comparable to the planetary time-scale t that is being estimated. If $\tau_{1/2} \ll t$, then present abundances of the parent nucleus A will be very low and hard to measure accurately. If $\tau_{1/2} \gg t$, then the B:A ratio will change only very slowly and is thus an insensitive indicator of the passage of a time interval t. If $\tau_{1/2}$ is not of the same order as t, uncertainties of measurement or of the "initial" ratios $N_B(0):N_A(0)$ will lead to unacceptable uncertainties in t. As the events of planetary evolution generally develop on time scales measured in billions of years, the radioactivities of greatest interest to planetary chronology are the long-lived decays of uranium-238, uranium-235, thorium-232, rubidium-87 and potassium-40, listed in the Table below:

DECAY	HALF-LIFE $\tau_{1/2}$ (yrs)	COMMON MINERALS CONTAINING PARENT
$U^{238} \rightarrow Pb^{206} + 8He^4$	4.5×10^9	Zircon, Uraninite, Pitchblende
$U^{235} \rightarrow Pb^{207} + 7He^4$	0.71×10^9	-- as for U^{238} --
$Th^{232} \rightarrow Pb^{208} + 6He^4$	13.9×10^9	-- as for U^{238} --
$Rb^{87} \rightarrow Sr^{87} + \beta^-$	47×10^9	Biotite, muscovite, lepidolite, microcline
$K^{40} + e \rightarrow Ar^{40} + \gamma$ (orbital)	1.3×10^9	Biotite, muscovite, hornblende, glauconite

Of these decays, the rubidium-strontium process is the simplest and is becoming one of the most reliable chronometers for planetary evolution. The rubidium parent is found as a minor constituent in the micas biotite $(KHMg_2Al_2(SiO_4)_3)$,

lepidolite $(KLi_2Al(Si_4O_{10})(OH)_2)$, and muscovite $(KH_2Al_3(SiO_4)_3)$, and in the feldspar microcline $(KAlSi_3O_8)$. Unfortunately the strontium-87 product also occurs independently in nature and the Sr^{87} in a rock need not be produced entirely by rubidium-87 decays. We cannot put $N_B(0) = 0$ in Equation 10.5 but must attempt in effect to determine the original Sr^{87} component in a given sample. In this we are aided by the fact that another strontium isotope, Sr^{86}, occurs in nature but is not produced by any known radioactivities. This means that the amount of Sr^{86} in any sample does not change with time. We therefore have that

$$Rb^{87}(t) = Rb^{87}(0)e^{-\lambda t}$$

$$Sr^{87}(t) = Sr^{87}(0) + Rb^{87}(0)(1 - e^{-\lambda t})$$

$$Sr^{86}(t) = Sr^{86}(0)$$

where for brevity the isotope symbol is now used to represent its abundance in the sample. From these relations we have that

$$\frac{Sr^{87}(t)}{Sr^{86}(t)} = \frac{Sr^{87}(0)}{Sr^{86}(0)} + \frac{Rb^{87}(t)}{Sr^{86}(t)}(e^{\lambda t} - 1) \qquad \text{Equation 10.6}$$

Now a given rock sample (of given age t) will normally contain a number of chemically different rubidium- and strontium-bearing minerals which will have different initial rubidium-to-strontium ratios. Because Sr^{87} and Sr^{86} are chemically identical however, they will occur in the same relative proportions in such different minerals so that $Sr^{87}(0)/Sr^{86}(0)$ will be the same throughout a given rock sample. Equation 10.6 then means that a plot of Sr^{87}/Sr^{86} against Rb^{87}/Sr^{86} for the different minerals in a sample will be a line of slope $(e^{\lambda t}-1)$
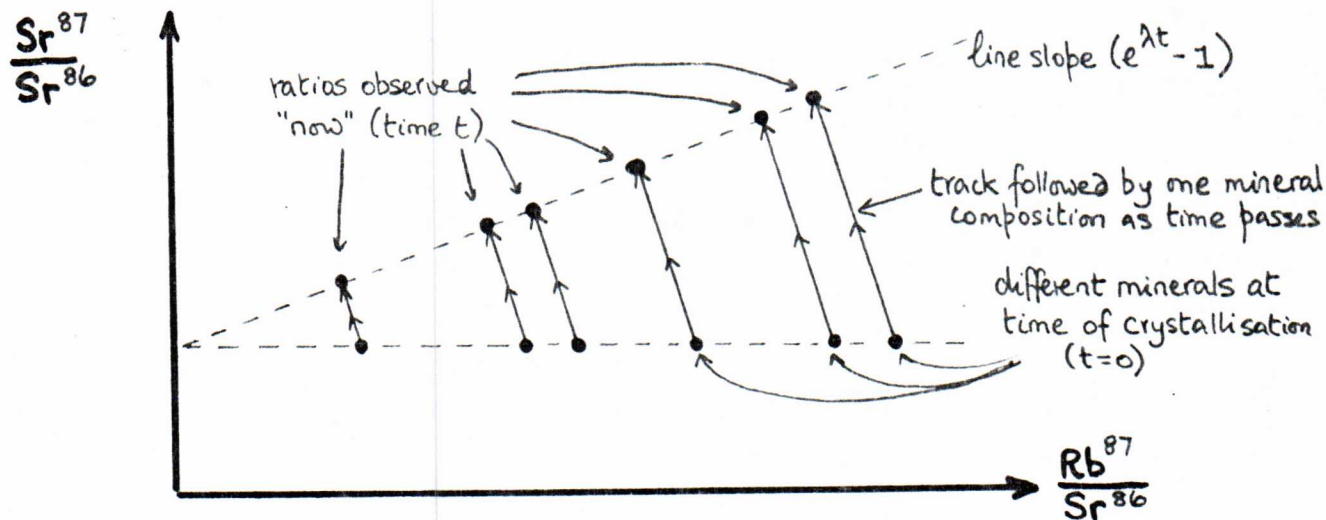
and intercept equal to the common ratio $Sr^{87}(0)/Sr^{86}(0)$. The technique applied is to make such a plot and then to use its measured slope and the known value of λ to find the time t since the event which fixed the initial compositions of the minerals. If the plot is not linear, then the sample may not have remained chemically closed throughout the time since last crystallization.

Similar techniques are applied in uranium-lead and thorium-lead dating, using the isotope Pb^{204}, which is not produced by known radioactivities, as the "control"-analogous to the use of Sr^{86} above. The uranium-"rich" minerals uraninite and pitchblende are relatively rare, so the occurrence of uranium as a $\sim$ 1 part per 1000 impurity in zircon $(ZrSiO_4)$ is of more general significance in chronology. The uranium-lead system is especially useful as the two isotopes U^{238} and U^{235} always occur together, so it is possible to cross-check "ages" obtained by $U^{238} \rightarrow Pb^{206}$ dating and by $U^{235} \rightarrow Pb^{207}$ dating. There is evidence however that solid-state diffusion of lead and uranium in minerals cannot be neglected, so these decays also have disadvantages as "planetary clocks" (see "Radiometric Dating for Geologists" for details).

Potassium-argon dating at first seems attractive because of the relatively high abundance of potassium (compared with rubidium and uranium) in the Earth's crust--but this advantage is offset by the fact that argon is a gas that is relatively easily released from rocks if they are heated or stressed, so allowance must usually be made for "open-ness" of the system to argon losses.

10.2. "Radioactivity Ages" for Solar System Bodies

On Earth, all continents contain some rock masses dated at up to 2.5 - 2.7 x 10^9 years since last major crystallization, although the majority of samples give significantly lower ages. The oldest rocks on the surface of the Earth have ages $\sim$ 3.7 x 10^9 yrs., and are found in Greenland, Rhodesia, Minnesota, Ontario and Northern Russia. The distribution of rock ages in Earth's crust implies that, although some rock masses have survived over 3.5 billion years of Earth history, major episodes of remelting have repeatedly reset the "radioactive clocks" over much of the surface. This is a result of the continuing activity of Earth's interior, which causes massive remelting and recycling of the crustal material.

This activity of the Earth prevents us from estimating by purely terrestrial observations the time that has elapsed since our planet as a whole separated from

the rest of the Solar System and became a "chemically closed" system--all we can say is that this event must have occurred more than 3.5 billion years ago. We can however estimate this "separation age" of the Earth by comparing data on the abundances of lead isotopes on Earth with those in meteorites. Meteorites are samples of the rocky Solar System debris whose orbits intersect Earth's and which become available to us after entering the atmosphere. About 7% of all meteorites reaching Earth's surface consist mainly of iron and the remainder are composed mainly of silicate minerals. The iron meteorites contain so little uranium and thorium that the relative abundances of their lead isotopes must represent abundances which are unaffected by radioactive decay and so are effectively "primeval" samples of the lead isotope ratios in the material which formed the Solar System.

If we then determine the average ratios of U^{238}, U^{235}, Pb^{206} and Pb^{207} to Pb^{204} in the Earth's crust, we can use the meteoritic lead abundances to correct for the "initial" Pb^{206} and Pb^{207} in the Earth's composition. This allows us to estimate the length of time over which lead of radioactive origin has been produced from uranium in the Earth by using Equations 10.3 and 10.4. A variety of estimates has been made by this means, giving an average of $(4.65 \pm 0.15) \times 10^9$ years for the "age of the Earth".

Recrystallization ages for the meteorites themselves, based mainly on rubidium-strontium and uranium-lead dating, generally agree with this age estimate for the Earth. We therefore have consistent evidence for a widespread segregation of Solar System material between ~ 4.5 and 4.8 billion years ago.

The Apollo (USA) and Luna (USSR) moon-landing missions returned rock samples to Earth which have been used to begin the study of lunar chronology. The samples from the lunar highlands all indicate that a major crystallization event took place there between 4.3 and 4.4 billion years ago, followed by a period of major disturbance and resetting of the radioactive clocks which ended between 3.9 and 4.0 billion years ago. The samples from the smooth lunar "maria" indicate that these extensive plains were formed between 3.2 and 3.9 billion years ago, after the main highland units were in place. The most striking feature of the ages of the lunar samples is the complete lack of crystalline rocks with ages less than 3.16 billion years. This contrasts sharply with the situation on Earth, where very few surface rocks have survived for this long. We can now be sure that the entire lunar surface is an ancient relic which preserves features of the early history of the Solar System that have long since been erased in the more turbulent environment of the Earth.

11) The Organisation of the Solar System

We found in Sections 4 and 6 above that certain gross properties of the Solar System planets are organised with distance from the Sun:

1) the two main types of planet - Jovian and terrestrial - are found at different distances from the Sun, the massive low-density Jovians being located exclusively in the outer Solar System while the less massive but high-density terrestrials are located exclusively in the inner Solar System.

2) the mean density of the terrestrial planets decreases with distance from the Sun, when we allow for the probable effects of gravitational self-compression.

This suggests that the relative abundances of heavy and light chemical elements might vary in a systematic way with distance from the Sun. As the timing of the planetary orbits shows that the Sun contains 99.98% of the known mass of the System (see the Table on p. 12), it is important to relate the variation of planetary compositions to that of the Sun itself.

11.1 The Solar and Planetary Compositions

Analysis of the absorption lines in the Solar spectrum tells us the chemical composition of the dilute upper levels of the Solar atmosphere, which lie immediately above the denser levels where the solar black-body spectrum at T_{eff} = 5780 K (p. 28) is formed. The Solar atmosphere contains traces of all chemical elements which would to detectable by our spectroscopic methods in a gas at $\sim$ 6000 K, but a handful of chemical species make up the bulk of the composition. The following Table lists the dozen most abundant elements in the Solar atmosphere.

PRINCIPAL ELEMENTS IN THE SOLAR COMPOSITION

Symbol	Element	Relative # Abundance %	Relative # Abundance $(Si = 10^6)$
H	Hydrogen	93	3.18×10^{10}
He	Helium	6.5	2.21×10^9
O	Oxygen	0.06	2.15×10^7
C	Carbon	0.035	1.18×10^7
N	Nitrogen	0.011	3.74×10^6
Ne	Neon	0.010	3.44×10^6

Symbol	Element	Relative #% Abundance	Relative #(Si = 10^6) Abundance
Mg	Magnesium	0.003	1.06×10^6
Si	Silicon	0.003	1.00×10^6 (defined)
Fe	Iron	0.002	8.30×10^5
S	Sulfur	0.0015	5.0×10^5
A	Argon	0.0002	8.5×10^4
Al	Aluminum	0.002	7.2×10^4

The question of whether or not this composition represents the bulk composition of the Sun cannot be answered directly, as we have no means of probing the deeper layers of the Solar atmosphere. Calculations of the hydrostatic equilibrium of a Solar mass of this composition show however that the gravitational compression of the central regions of the Sun would raise the material there to a density and temperature at which thermonuclear fusion of hydrogen to helium could occur. The energy release from hydrogen-to-helium fusion can account for the observed luminosity of the Sun, and furthermore can have supported this luminosity with the Sun in its present state of compression for at least the ~4.6 billion years estimated as the "age of the Earth" in Section 10.2 above. In other words, the observed luminosity, mass and size of the Sun are consistent with the assumption that the above composition indeed represents the bulk composition of Solar material, except for a central core region where the helium abundance should be enhanced due to hydrogen fusion over the last few billion years.

The solar composition listed above therefore represents our best estimate of the composition of most of the mass of material in the Solar System. The crucial question in relation to the planets is then: do the planets have compositions which can be simply related to this Solar mixture? The low mean densities of the Jovian planets imply that their compositions must be generally similar to the hydrogen--and helium-rich Solar mix, as is also suggested directly by the observations of the atmospheric composition of Jupiter itself (see Section 9.3 above). It is certainly reasonable to examine models of the origin of the Jovian planets which would attempt to form them from material of essentially Solar composition.

On close inspection, this hypothesis is also promising for the terrestrial planets. If the solar composition were stripped of highly volatile elements, such as hydrogen and helium, and of volatile compounds such as water, hydrocarbons, etc.

the residual composition would be rich in magnesium, silicates, and iron, which are just the materials we find to be abundant in the bulk Earth composition model. This conclusion is reinforced by mineralogical analysis of meteorite material, which gives the results shown in the Table below (averaged over all classes of meteorite):

AVERAGE METEORITIC COMPOSITION

Mineral Component	% by Weight	Total
SiO_2	38	
MgO	24	62%
FeO	12	74%
Fe	12	86%
FeS	6	92%
Al_2O_3	3	95%

We can provisionally conclude that both the terrestrial and Jovian planetary abundances might be derivable from the Solar composition if a suitable mechanism could be found for removal of volatiles from that composition in the inner Solar System. We will examine possible mechanisms later in the course.

11.2. Organisation of the Planetary Motions

The idea that the planets might have their origin in common with one another and with the Sun is hinted at by the compositional organisation of the Solar System. It is also hinted at strongly by the organisation of the motions in the System.

First, the planets all orbit around the Sun in the same sense, and close to a common plane. This means that if an outside observer saw the Earth orbiting around the Sun in the clockwise direction, then he would see all of the other planets orbiting in the same direction. In the table on pg. 68 are listed the angles of inclination (i) of the other planetary orbits to the "plane of the Earth's orbit" (known as the ECLIPTIC). The only planets whose orbits are inclined by more than 3 1/2° to Earth's are Mercury and Pluto.

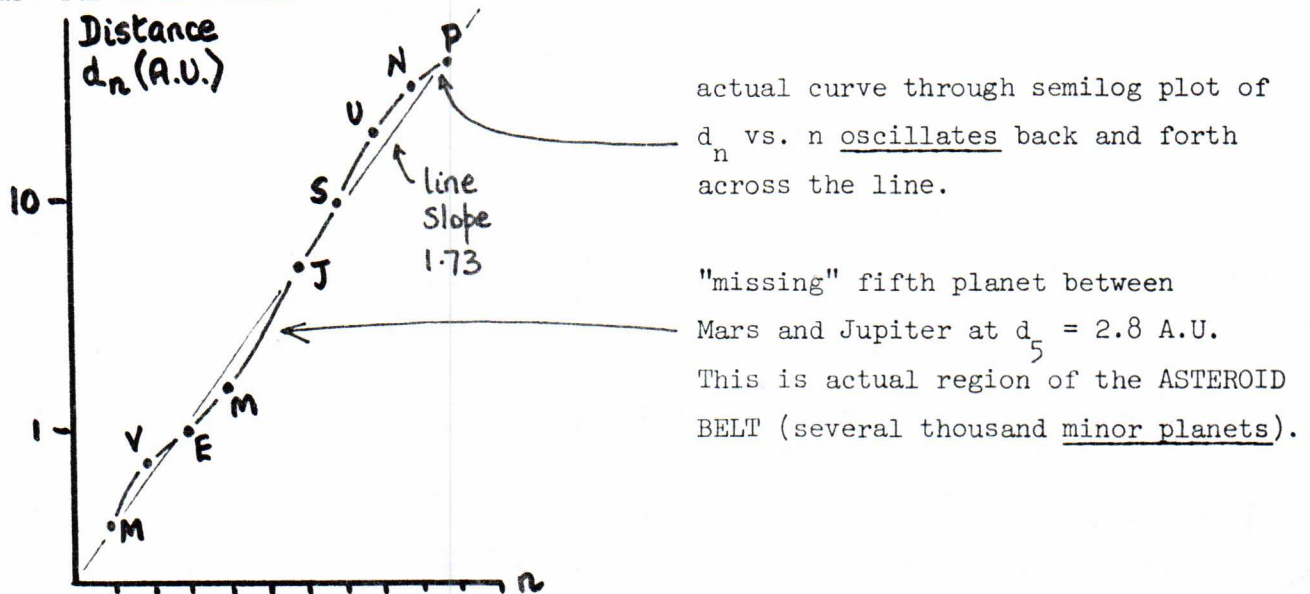| Planet	Orbital Inclination to the Ecliptic (i)	Eccentricity of Orbit (e)
Mercury	7° 1' 15"	0.2056
Venus	3° 23' 40"	0.0068
Earth	--	0.0167
Mars	1° 51' 00"	0.0934
Jupiter	1° 18' 17"	0.0485
Saturn	2° 29' 22"	0.0557
Uranus	0° 46' 23"	0.0472
Neptune	1° 46' 22"	0.0086
Pluto	17° 10'	0.250

Second, the planetary orbits do not deviate substantially from being circles. The "eccentricity" (e) listed to the right in the preceding Table is a measure of the departure from circularity of the orbit:

$$\frac{1+e}{1-e} = \frac{\text{Furthest distance of planet from Sun}}{\text{Closest distance of planet to Sun}}$$

The orbits are in fact ellipses with the Sun at a focus of every ellipse. This fact was of the utmost importance in leading to a theory of universal gravitation in the Seventeenth Century, but the ellipses, apart from those of Mercury and Pluto, are sufficiently close to being circles that we can treat many aspects of planet formation in circular-orbit approximation in a first-order theory.

Neither the near-coplanar arrangement of the orbits nor the near-circularity of the orbits would be expected if the planets were a random collection of interstellar debris encountered by the Sun on its travels through our Galaxy. Instead, the organisation of the planetary motions suggests a common dynamical ancestry for the matter which formed the planets. Close association with the Sun is also indicated by the fact that the solar rotation is in the same sense as the planetary orbital motions and that the solar equator is only 7° 15' from the plane of the Earth's orbit (and so is close to, but not exactly in, the mean plane of all the planetary orbits).

A final orbital regularity is the fact that the planetary orbits are fairly regularly spaced, so that the ratio $(d_{n+1}:d_n)$ (where d_n is the mean distance of the n'th planet from the Sun) is 1.73 ± 0.2 through the Solar System. This orbital regularity was greatly reinforced by the discovery of the belt of minor planets, or asteroids, between Mars and Jupiter, around the region corresponding to n = 5 on this relationship (taking n = 1 for Mercury, n = 2 for Venus, etc.). The regularity of the planetary spacings has been expressed in a variety of ways in the history of planetary studies, and is generally known as the Titius-Bode Relation. The diagram below shows a plot of $\log (d_n)$ versus n. If the ratio $(d_{n+1}:d_n)$ were precisely constant, this plot would be a straight line. The real plot oscillates back and forth around the line of slope 1.73, showing that the spacing organisation is not quite a simple ratio factor, but that the spacings are far from random.

actual curve through semilog plot of d_n vs. n oscillates back and forth across the line.

"missing" fifth planet between Mars and Jupiter at d_5 = 2.8 A.U. This is actual region of the ASTEROID BELT (several thousand minor planets).

The satellite systems of the planets also show considerable organisation of their motions. The inner satellites of Jupiter, Saturn and Uranus, and both satellites of Mars, all have orbits within $\sim 1°$ of their planets' equators, and circulate around their planets in the same sense as their planets rotate, in orbits of low eccentricity (e $\lesssim 0.02$). There are exceptional satellites however, notably Earth's Moon, whose orbit is at $18° - 28°$ from Earth's equator at different times, and has e = 0.055. The outer satellites of Jupiter and Saturn, and the two satellites of Neptune, also break one or more of the above rules. The trend however is for the satellite systems to be organised like "miniature Solar Systems", and this also applies to orbit spacings ($d_{n+1}:d_n \sim$ constant being the rule in the satellite systems of the outer planets.)

In summary, there is much circumstantial evidence both from compositions and motions in the Solar System that the planets have an origin in common with one another and with the Sun. We shall now attempt to trace a theoretical picture of this origin.

12) Our Galaxy - The Milky Way

On a clear moonless night away from city lights we can see a hazy band of light across the sky on a great circle around the Earth. This band of light is shown by even a modest instrument, such as binoculars or a small telescope, to consist of a vast number of individual stars - it is in fact, the plane of the flattened distribution of stars and gas we call our home 'galaxy', or the Galaxy (as in the Earth).

12.1. Galactic Structure

The Galaxy is a loosely spiral arrangement of stars which is about 100,000 light-years (i.e. about 10^{18} km) across. All of the material in it is in orbit around a dense central region called the NUCLEUS. As in the Solar System, the orbital velocities vary with distance from the centre, so the Galaxy does not "rotate" as a solid body, except in its central regions where the star density is high. The outer parts of the Galaxy, which contain our Sun, revolve around the nucleus with a pattern of velocities similar to that in the Solar System (Equation 2.1). The Sun is about 30,000 light-years from the nucleus and circles once around the Galaxy in about 250 million years. From this and Equation 2.3 we can infer a total mass for the Galaxy of about 10^{11} solar masses. As the mass of the Sun is a "typical" stellar mass we conclude that our Galaxy may contain about one hundred billion stars.

Globular Clusters

Sun's Location

Disc

Nucleus

100,000 Light Years

Side View

Sun's Location
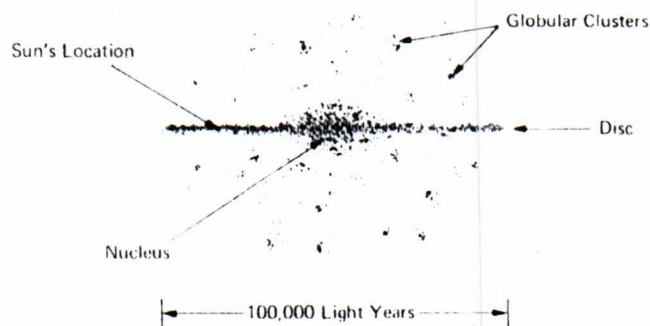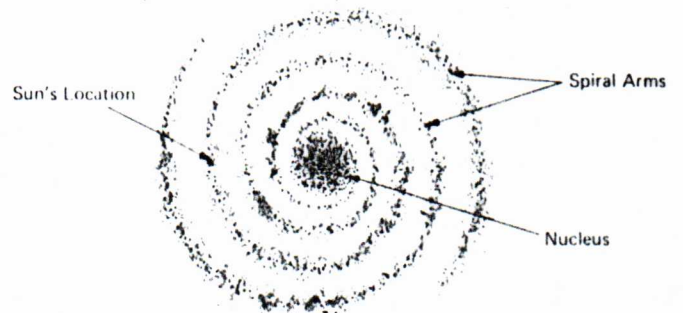
Spiral Arms

Nucleus

Top View

The spiral structure of the Galaxy is flattened to a disk (see diagram) whose thickness near the Sun is about 6000 light-years. The spiral arms extend outwards from the nucleus (the actual pattern being less organised than that shown schematically in the diagram). The Sun is located in the inner edge of a spiral arm; when we look in what we call the constellation of Cygnus we look nearly along the tangent to this arm and hence see a richer field of stars than we do at right angles to this direction.

The average distance between stars in the solar neighbourhood is about 5 light-years; the Sun's actual nearest neighbour is the Alpha Centauri triple system, whose distance is about 4.3 light-years from the Sun. It is important to remember that the distances between stars are enormous compared with the sizes of the stars themselves; the distance to Alpha Centauri is about 30,000 solar diameters. This is the reason for the use of light-years (or parsecs, which are a unit related to the procedure for measuring stellar distances and equal to 3.26 light-years) for quantifying stellar distances. On the scale of the Galaxy as a whole, individual stars are as individual atoms are to Man: Galaxy/Star is a scale factor similar to Man/Atom.

12.2. Star Clusters and Gas Clouds (Nebulae)

Throughout the disk of our Galaxy we find noticeable clumps or groups of stars which are much closer together than average. These groups are known as star clusters. The clusters in the disk tend to lie in the spiral arms, contain from about 50 to about 1000 stars, and do not have very symmetric stellar layouts or strong concentrations of stars to their centres. These are called OPEN CLUSTERS; about 1055 are presently documented.

In contrast, there exists another class of star cluster, the GLOBULAR CLUSTERS. These are distributed through a volume of space extending far above the disk but still symmetric about the galactic nucleus; this spheroidal region is known as the "halo" or "corona" of our Galaxy. Globular clusters, of which about 125 are known, can contain 100,000 or even 1,000,000 stars and show pronounced spherical symmetry and central condensation. Their diameters range from 50 to 500 light-years while open clusters are smaller, generally less than about 30 light-years in diameter. Estimates of the ages of stars in clusters based on theories of their thermonuclear equilibrium place most of the globular clusters among the oldest known structures in the Galaxy, up to 12 or more billion years old. The open clusters contain much younger stars in general, with ages as recent as 100 million years or so.

The disk region of the Galaxy also contains many regions where interstellar matter is made visible by starlight. Some open clusters, such as the bright Pleiades Cluster, contain REFLECTION NEBULAE, which are a thin veil of dusty material which scatters (reflects) the light of the embedded stars. Thus the spectrum of a reflection nebula mimics the spectra of the stars it contains. Reflection nebulae are typically a few light-years across and contain interstellar material at a density of order a few times 10^{-20} kg/m^3.

The regions around individual hot stars or small clusters sometimes glow with a predominantly reddish light, which on spectral analysis provides a rich emission-line spectrum dominated by the red Balmer line of hydrogen. These EMISSION NEBULAE are gas clouds fluorescing under the action of the ultraviolet light from the hot stars which they contain. Their composition, like the solar composition, is predominantly hydrogen and helium with an admixture of the heavier elements which can amount to as much as 5% of the total composition. They are examples of interstellar gas clouds rendered visible by the action of the stars they contain, i.e. by the excitation of the fluorescence process.

The emission nebulae are often crossed by dark lanes of absorption, sometimes gathered into prominent dark knots called DARK NEBULAE. These are concentrations of interstellar dust grains which are opaque to starlight or to the light from the emission nebulae.

The emission nebulae provide spectacular evidence for the continued presence in the galactic disk of interstellar gas of the right composition to make stars as we know them through the Sun. It is important to realise that we SEE these gas clouds because they contain illuminating stars. Far more are known by radio-astronomy methods, which can detect gas clouds which do not contain stars and are too cool to emit significant amounts of visible light. These are detected by their 21-cm wavelength radio radiation, which is produced by electron spin 'flips' in the ground state of the hydrogen atom. The invisible clouds greatly outnumber the visible emission nebulae, and it is their properties which we shall use as starting-points for discussion of star formation and planet formation.

For a further review of the observed properties of stars and interstellar gas, read Wood, Chapter Six, p. 133–156.

13) The Collapse of Interstellar Gas Clouds

Optical and radio astronomy tell us that interstellar gas clouds do exist among the stars. The random thermal motions of the gas particles tend to disperse the interstellar clouds, while their self-gravity tends to make them contract onto themselves. If a cloud is in equilibrium, these opposing tendencies must be balanced. The problem of stability of an interstellar cloud has common ground with the problem of escape of planetary atmospheres--if an individual gas particle is in the exosphere of the cloud travelling outwards with a velocity exceeding the escape velocity, the particle will escape. In the planetary escape problem however we can distinguish an atmosphere of escapees from the dense stabilising mass of the planet. In an interstellar gas cloud, the escapees and the stabilising mass are the same system--the "atmosphere" is the entire cloud mass. Under these circumstances the criterion for stability can be put in a more useful form than our earlier criterion for atmospheric retention.

13.1) Jeans Criterion for Gravitational Instability

The criterion for gravitational stability of a gas mass is known as the Jeans Criterion, after the physicist J.H. Jeans who studied the problem in 1928. An exact analysis needs advanced mathematical methods--we can no longer make the simplifying assumption that the acceleration due to gravity is nearly constant throughout an extended gas cloud. An approximate analysis, which gives almost the same result as more sophisticated treatments, can be based on considering the energy balance of a gas cloud, as follows.

If the total kinetic energy present in the thermal motions of the particles could provide the work necessary to separate all the particles to infinity against the gravitational attraction of the cloud, then a real cloud will be able to expand significantly, whatever the details of the velocity and collision histories of individual particles. An approximate condition for dispersal of a warm gas cloud would then be

$$\text{Total K.E. (cloud)} + U_G(\text{cloud}) > 0$$

It will be briefer to refer to the first term as the THERMAL ENERGY of the cloud, henceforth U_T. A regime of approximate equilibrium between gravity and thermal motions (with perhaps only slow dispersal) should be expected if

$$U_G \underset{\sim}{\sim} - U_T$$

An exact treatment shows that the proper stability criterion is in fact

$$U_G = - 2 U_T \qquad\qquad \text{Equation 13.1}$$

To calculate U_T, we simply multiply the average kinetic energy per particle (from Equation 9.17) by the total number of particles in mass M of the gas:

$$U_T = \frac{3}{2} kT \cdot \frac{M}{\mu u} \qquad\qquad \text{Equation 13.2}$$

where μ is the mean molecular weight of the gas, T is its temperature (assumed constant throughout it) and u is the atomic mass unit.

To calculate U_G, suppose that the cloud is spherical and of uniform density ρ. U_G represents the total work which would be necessary to dismantle the cloud entirely--to a set of infinitely separated particles. To calculate that, we proceed as follows:

1) Recall that the gravitational potential energy U_G of a mass m at distance d from a mass M is (by Equation 9.9):

$$U_G = - \frac{GMm}{d}$$

2) Recall that in a spherically-symmetric body a mass element at radius r experiences no net attraction to the material at larger radii, and is attracted to the material at small radii AS IF that material were concentrated at the centre of the spherical body.

3) Now consider the gravitational potential energy dU_G of a thin spherical shell radius r, thickness dr, in a spherical body of constant density ρ (see diagram)

$$dU_G = - \frac{GM(r)m(r)}{r}$$

where M(r) = total mass INSIDE radius r,

i.e. $M(r) = \frac{4}{3} \pi r^3 \cdot \rho$
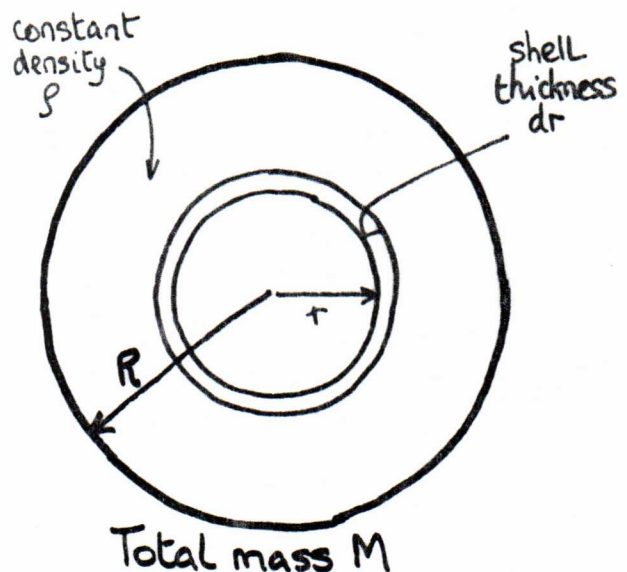
and $m(r)$ = total mass in the shell = $4\pi r^2 dr \cdot \rho$

Hence $dU_G = -(\frac{4\pi\rho}{3})^2 G \cdot 3r^4 \cdot dr$

4) To find the total potential energy U_G of the entire spherical mass, total mass M, radius R, we add all contributions dU_G from all spherical shells $r = 0$ to $r = R$, by integration:

$$U_G = -(\frac{4\pi\rho}{3})^2 G \cdot 3 \int_0^R r^4 dr = -(\frac{4\pi\rho}{3})^2 G \frac{3}{5} R^5$$

$$= -\frac{3}{5} G (\frac{4\pi\rho R^3}{3})^2 \frac{1}{R}$$

i.e. $\boxed{U_G = -\frac{3}{5} \frac{GM^2}{R}}$ Equation 13.3

Then $U_G = -2 U_T$ when

$$M = 5 (\frac{k}{uG}) \frac{T}{\mu} R$$

This relation between the mass M and radius R of an "equilibrium cloud" is less convenient for many purposes than one between M and the density ρ. We can eliminate R in favour of ρ by substituting

$$R = (\frac{3M}{4\pi\rho})^{1/3}$$

which after some rearrangement gives the expression

$$M^2 = \frac{375}{4\pi} (\frac{k}{uG})^3 (\frac{T}{\mu})^3 \frac{1}{\rho}$$

in equilibrium. Putting in numerical values for the constants,

$$\boxed{M = 7.5 \times 10^{21} (\frac{T}{\mu})^{3/2} (\frac{1}{\rho})^{1/2} \quad kg}$$ Equation 13.4

Given μ, T and ρ, clouds with M greater than this will contract under gravity and clouds with smaller M disperse.

We should not expect the masses, compositions, temperatures and densities

of real interstellar clouds to conform exactly to the relationship in Equation
13.4, for many reasons. Real clouds will not be spherical, or of uniform density
and temperature. The argument behind Equation 13.4 also takes no account of
possible bulk motions in the clouds, or of non-gravitational forces such as those
due to the weak interstellar magnetic fields. Nevertheless, Equation 13.4 turns
out to represent a creditable guess at the range of stable structures in the
interstellar medium, and identifies a basic problem in star, and hence planet,
formation.

The heat sources available in interstellar regions (starlight, cosmic-ray
particles, shock waves from stellar explosions) maintain the dilute interstellar
gas, which is mainly hydrogen, at temperatures typically 50 to 100 K. At these
temperatures the hydrogen is mainly molecular so that $\mu = 2$. The gas densities
are typically in the range 10^{-20} to 10^{-18} kg/m^3. For these ranges of values
Equation 13.4 predicts stable masses of a few hundred to a few thousand solar masses,
and cloud radii of order a few tens of light-years (one light-year is the distance
travelled by light in one year, i.e. 9.46×10^{12} km). These are indeed typical
interstellar cloud masses and sizes. BUT NOTE THAT CLOUDS WITH ONLY ABOUT ONE
SOLAR MASS WILL NOT BE STABLE AT THESE DENSITIES AND TEMPERATURES. One solar mass
is so far below the stabilising mass that CLOUDS WITH ONLY ONE SOLAR MASS WILL
DISSIPATE AT NORMAL INTERSTELLAR TEMPERATURES AND DENSITIES. Equation 13.4 thus
tells us that collapse of a one-solar-mass cloudlet would require severe pre-
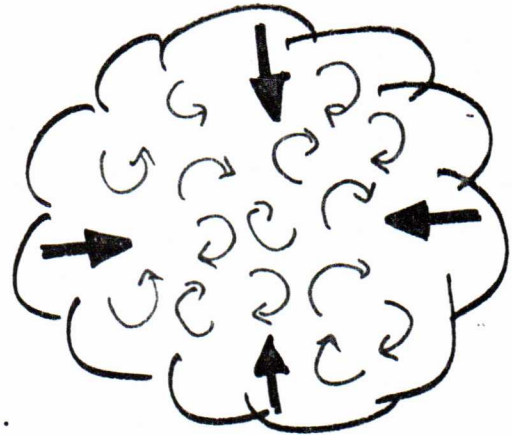cooling or precompression relative to "normal" interstellar conditions.

Compression of an interstellar cloud is possible by shock waves from stellar
explosions, or by passage of the cloud into a spiral arm of the galaxy. Varia-
tions in heat input are also possible, but none of the possible fluctuations
can overcome the basic discrepancy between the "solar" mass and the self-stabilising
mass. This means that isolated stars like our Sun cannot form directly from the
observed interstellar clouds. In any plausible scenario, the initial stages of
gravitational collapse in the interstellar medium must involve entire clouds whose
total masses would provide many stars—i.e. star clusters.

13.2) Fragmentation Into Cloudlets

Real interstellar clouds are not in fact uniform but contain density
fluctuations and turbulent eddies. The density fluctuations do not initially

affect the overall stability of the cloud, according to Equation 13.4, but their presence is crucial to what happens later.

As the overall contraction of a cloud proceeds, the individual subfluctuations are themselves compressed. While they remain transparent the temperature increase that results from their compression can be radiated away to their cooler surroundings. The effect of this is that the density of the subfluctuations steadily increases while their temperature remains more or less constant. This means that as the collapse of the cloud continues, ρ increases in the subfluctuations while T generally does not. Referring back to Equation 13.4, we see that the equilibrium mass M in any subfluctuation therefore decreases as time passes. The actual mass of a subfluctuation is essentially constant however, so in time the mass of each subfluctuation becomes greater than the equilibrium mass at its temperature and density. When this occurs, the subfluctuation can begin to collapse locally, independent of the fate of the rest of the cloud.

The steady decrease of the equilibrium mass, coupled with the presence of initial inhomogeneities, leads to fragmentation of the cloud into separate, independently contracting cloudlets. These cloudlets will generally be rotating, due to the turbulence of the original cloud. So even if the overall cloud does not rotate on average, the separate cloudlets will rotate. The process of fragmentation is believed to be the key step in the formation of a star cluster, as the equilibrium mass finally falls to values in the range of stellar masses.
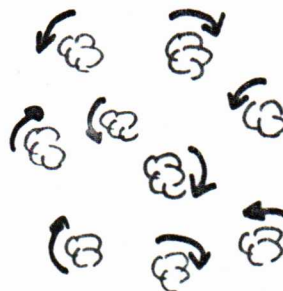
Infall K.E. → converted into turbulent K.E. → halts gross contraction.

13.3) Angular Momentum and Flattening to a Disk

Once a cloud has begun to collapse independently, the main force on it is its own gravity, rather than gas pressure from the rest of the cloud. Because its gravity is directed towards its own centre, it can not exert any couple (torque)

on itself and so its total <u>angular momentum</u> must remain constant as it collapses. For an individual gas particle of mass m travelling in a circular orbit of radius r around the cloudlet centre at velocity v, the angular momentum around the rotation axis is

$$L = mvr = m\omega r^2$$

As the cloudlet contracts, r for a given particle must decrease on average, so that its velocity v must increase. This means that <u>the cloudlets spin faster as they contract</u>. Referring back to Section 5.2, recall that the effect of rotation is to <u>decrease</u> the effective self-gravity of a mass in its rotation equator but to leave unchanged its effective self-gravity in the polar direction. This behaviour was discussed in Section 5.2 in relation to the oblateness (flattening across the poles) of the planets. We must now examine its consequences for the development of a gravitationally-contracting cloudlet.

Basically, nothing can stop the gravitational collapse of a cloudlet in the direction down the rotation axis until the cloudlet reaches a density at which it becomes opaque to its own radiation. While it remains transparent, the temperature increase produced by gravitational compression can be radiated away. This prevents the gas pressure in the cloudlet from providing support against the cloudlet's gravity. Collapse down the axis therefore proceeds essentially unimpeded.

Collapse across the equator is however limited by the fact that only <u>part</u> of the cloudlet's self-gravity is available to produce contraction. While the cloudlet remains nearly spherical, this part will be given by Equation 5.4:

$$\text{Inwards force per unit mass} = \{\tfrac{4}{3}\pi\rho G - \omega^2\}r$$

where ρ is the mean density of the cloud. This inwards force becomes zero, i.e. collapse across the equator is halted, when

$$\omega^2 = \tfrac{4}{3}\pi\rho G$$

which, using Equation 13.5 occurs at a radius given by

$$r^4 = \frac{3L^2}{4\pi G\rho m^2}$$

where L is the original angular momentum of the particle. A particle moving in the equator with initial angular momentum L will spiral inwards as the cloudlet contracts, until it reaches the radial distance r where Equation 13.6 becomes satisfied. The particle is then in a stable orbit around the rotation axis, and will not approach closer to the axis.

Because the collapse down the rotation axis cannot be halted in this way, the cloudlet must flatten to a disk whose elongation is in the original rotational equator. The kinetic energy acquired by the matter falling downwards into the disk will tend to increase the temperature of the cloudlet by collisions, but will be radiated away so long as the disk remains transparent.

Another property of the collapsing cloudlets is that they begin to sort particles in space according to their angular momentum around the rotation axis. Consider which particles can reach the central part of the cloudlet-disk. Particles fall freely down the rotation axis because they have essentially no angular momentum about this axis; particles reaching low values of r by collapsing across the equator must begin with low values of L, from Equation 13.6. Particles starting near the centre will have little angular momentum because of their low values of r (initial). These considerations all mean that the angular momentum per unit mass will be smallest near the centre of the disk and greatest towards to the edge.

The final state of an initially uniform collapsing cloudlet will therefore be a flattened disk containing a central condensation of relatively low angular momentum per unit mass, surrounded by particles which are no longer spiralling inwards, but have come to precisely the orbital equilibrium described in Section 2, in which the gravitational attraction of the matter near the centre of the disk exactly provides the centripetal acceleration necessary to maintain each particle's orbit.
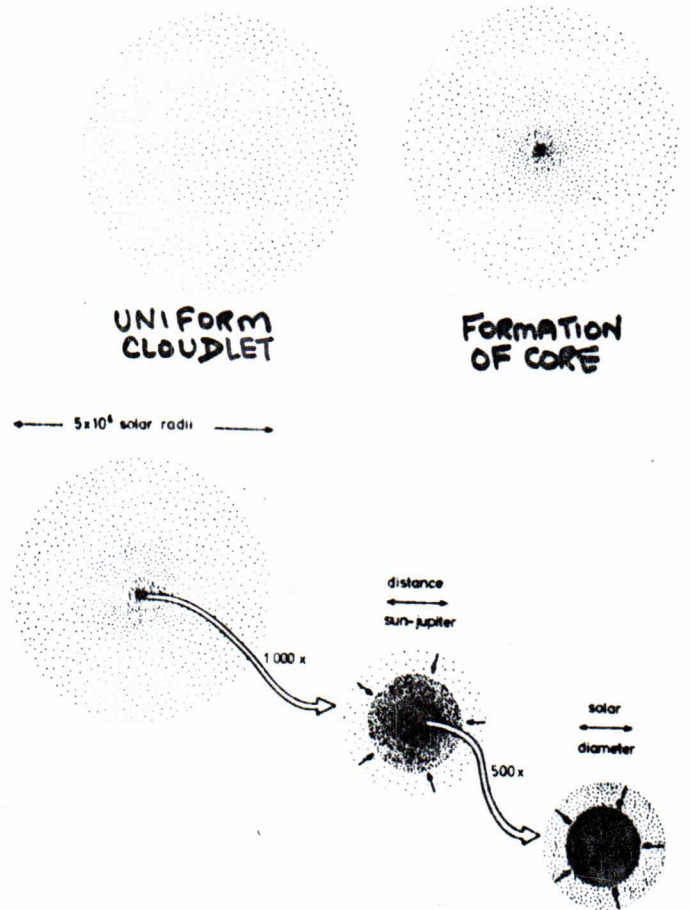
14) Star Formation

The work done by gravity on the cloudlet fragments during their collapse is converted into kinetic energy of infall and of random motions (i.e. into increasing temperature). The randomised (thermalised) energy can be radiated away to the cooler environment as long as the mass concentrations in the cloudlets remain transparent to their own radiation. There will come a point in each cloudlet however when its denser parts become opaque to their own radiation--i.e. when the mean free path of a photon becomes shorter than the characteristic scale of the system due to the increasing density. Once part of a cloudlet becomes opaque, its

radiation is partly trapped and so further collapse will begin to <u>heat</u> the dense region, and thus eventually to increase its gas pressure.

Once opacity becomes important and the temperature rises in the densest parts of the cloudlet, gas pressure can play a role in slowing the collapse towards an eventual hydrostatic equilibrium. <u>Gas pressure first slows the collapse in the central core of the cloudlet,</u> then material which is still falling in from the less dense regions must be brought to a halt at the boundary with the equilibrated core (see diagram to the right). This forms a shock front at the surface of the core where the energy of infall is converted by collisions into random kinetic energy (i.e. the work done by gravity is fully converted into heat). The core becomes increasingly compressed and heated by the continuing infall and its pressure and temperature rise substantially.

UNIFORM CLOUDLET

FORMATION OF CORE

5 x 10⁶ solar radii

1 000 x

distance
sun-jupiter

500 x

solar
diameter

14.1) The Onset of Nuclear Fusion

As the temperature in a cloudlet's hot core rises, the kinetic energies brought into collisions by the gas atoms there become high enough to cause ionisation, and the core becomes <u>thermally ionised</u>. In hydrogen, the ions are bare protons, so fully ionised hydrogen is in effect two intermingled gases, one of free protons and the other of free electrons. This ionisation occurs near $T \sim 4000$ K and contributes greatly to the opacity of the gases, as free electrons are very effective scatterers of radiation.

At sufficiently high temperatures, proton-proton collisions enter a new physical regime which is crucial to the evolution of the core. The diagram on pg. 81 shows the mutual potential energy of two protons as a function of their separation:
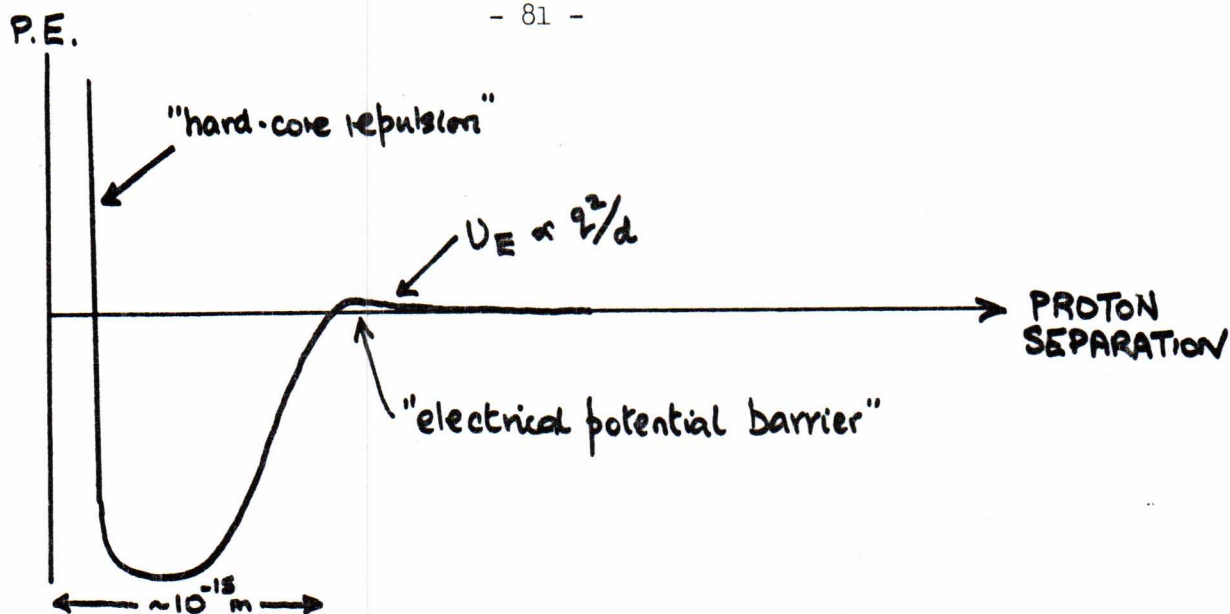
The significance of these processes stems from the fact that some mass disappears in this chain of events. The net result of the chain of reactions is the conversion of four protons into a He^4 nucleus, two positrons, two neutrinos, and some γ-radiation. In this conversion the mass of the products is less than that of the reactants by about 0.029 atomic mass units, and an equivalent amount of energy is released; the mass loss Δm and energy release ΔE are related by the Einstein relationship

$$\Delta E = \Delta m \cdot c^2$$

The energy is released to the gas in the kinetic energy of the products and in γ-rays, which will be reabsorbed; in either case it contributes to maintaining the temperature of the gas. Only the ∿ 2% of the released energy which is given to the neutrinos escapes from the gas, because neutrinos have almost no interactions with other matter. The factor c^2 in the $E = mc^2$ relationship makes the energy release from nuclear fusion one of the most potent energy supplies known to physics, and once the fusion processes begin the core can easily resupply the energy lost by radiation from its surface. This means that the core can stop contracting, and can hold itself in hydrostatic equilibrium by maintaining the temperature everywhere inside it with the energy provided by nuclear fusion. This fusion-supported equilibrium is the intrinsically stellar process, and once it has begun we can say that a star has formed in the contracting cloudlet.

Fusion can hold a hydrogen star whose mass is equal to the Sun's in hydrostatic equilibrium for ∿ 10 billion years, so once such a star forms in a cloudlet it dominates the energetics of the cloudlet from that time onwards. After star formation has occurred, the remaining contraction goes on in the presence of a steady energy supply from the star, which stabilises the temperature distribution throughout the disk. This temperature stabilisation proves to be the crucial factor in the development of a planetary system.

Starting by assuming a uniform spherical interstellar cloudlet of mass about one solar mass and an initial density of ∿ 10^{-16} kg per m^3, R.B. Larson showed in 1969 from computer simulations of the gravitational collapse that ∿ 400,000 years elapse before an opaque core is formed. The star's lifetime while supported by nuclear fusion should therefore be much longer than the time taken to develop the protostar as an opaque core, and we can therefore expect the time taken for formation

P.E.

"hard-core repulsion"

$U_E \propto q^2/d$

PROTON SEPARATION

"electrical potential barrier"

$\sim 10^{-15}$ m

At large separations, their interaction is dominated by the mutual repulsion of their positive electrical charges--which gives a positive potential energy term which increases as the reciprocal of their separation:

$$U_E \propto + \frac{q^2}{d}$$

At separations $\leqslant 10^{-15}$ m however the proton-proton interaction is dominated by the powerful but short-range <u>nuclear force</u>. If two protons come closer than the peak of the "electrical barrier" their future is controlled by this nuclear force, under which their identity as protons may not be preserved. To provide the kinetic energy necessary to overcome the electrical barrier, temperatures of order a few million K are necessary. Once the core reaches these temperatures, proton-proton collisions can initiate <u>nuclear fusion reactions</u> there:

$$H' + H' \rightarrow D^2 + \beta^+ + \nu$$
$$D^2 + H' \rightarrow He^3 + \gamma$$
$$He^3 + He^3 \rightarrow He^4 + 2H'$$
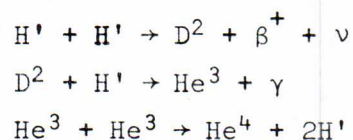
In the first reaction, the two protons become a proton and a neutron bound together to make the heavy isotope of hydrogen known as <u>deuterium</u> (D^2). The positive charge formerly carried by one of the protons appears on a positively-charged electron (or 'positron', β^+) and a neutrino (ν) is also emitted. A later collision between the deuteron so produced and a third proton can incorporate the proton in a nucleus of the light isotope of helium, He^3. Collision between two He^3 nuclei can finally form the exceptionally stable nucleus He^4, which contains two neutrons and two protons; the two "spare" protons are ejected and can participate in further fusion processes.

of a star to be much less than its eventual lifetime as a radiating object.

14.2) Multiple Stars

Before considering the theories describing the formation of planets from the disk of gas accompanying a newly-born star, we must mention an alternative course of evolution which appears to be followed in many stellar systems. If the cloudlet from which a star is forming is not very uniform initially, it is possible that collapse of the cloudlet produces several mass concentrations which can become cores, i.e. that the processes we have previously described occur in more than one mass centre, to produce several stars. When such multiple stars form, the original angular momentum of the cloudlet ultimately resides in the revolution of the newly-formed stars around their mutual centre of mass. This forms multiple stars whose interstellar separations should be comparable with the dimensions of the disks which accompany single stars. We will return to this aspect of multiple star systems later, but note here that astronomical observations indicate that about 50% of all stars near the Sun are in fact multiple systems in which the interstellar distances are often of the same order as the Sun-Jupiter separation in the Solar System.

15) Evolution of a Preplanetary Disk

All modern theories of formation of the Solar System now presume that the formation of the Sun took place about 5 billion years ago, and that this event was accompanied by the formation of a thin orbiting gas disk by the processes outlined in Section 13.3. The chemical composition of this gas disk would have been the same as that of the Sun. The goal of the theories is to explain how the planets could have formed by accumulation of solids within this gas disk.

15.1) Equilibrium of the Gas Disk
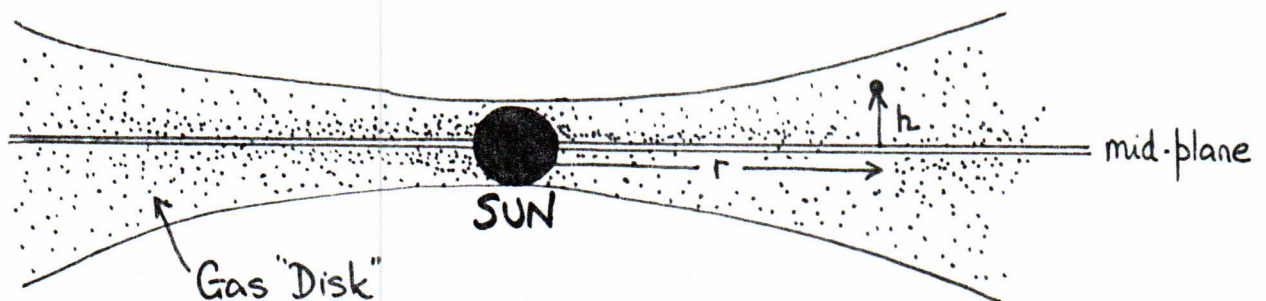
Without going into details of the calculations, we can examine how to specify the physical conditions in the gas disk once the central star has stabilised its energy output. First we consider <u>force balance</u> in the disk.

The expression for the total gravitational force F_g on a mass m at distance r, height h in the disk will depend in detail on the ratio of masses in the central star and in the disk. <u>We do not know what this ratio was in the early Solar System,</u> but can place a limit on it by noting that the gas disk must have contained at least enough heavy elements to form the known heavy-element mass of the planets. If the disk of gas was originally of solar composition, the <u>minimum</u> mass of the disk must have been about 2% of the solar mass. This minimum estimate presumes that no heavy elements were lost from the system during the processes that formed the planets. If losses occurred, then the original disk mass must have been > 2% of the solar mass.

<u>In a "minimum-mass" model we can neglect the mass (and hence the gravity) of the disk itself</u> and write the gravitational force as that due to the Sun alone:

$$F_g = \frac{GMm}{(r^2+h^2)} \qquad \text{towards the Sun} \qquad \qquad \text{Equation 15.1}$$

which has two components:

$$F_{in} = \frac{GMmr}{(r^2+h^2)^{3/2}} \qquad \underline{\text{inwards}} \text{ towards the Sun}$$

and

$$F_{down} = \frac{GMmh}{(r^2+h^2)^{3/2}} \qquad \underline{\text{downwards}} \text{ towards the mid-plane of the disk.}$$

When the disk is in equilibrium, F_{in} must be in balance with the centripetal force needed to maintain the orbital motions and with the gas pressure at every distance d:

$$\text{i.e.} \quad F_{in} = m\left(\frac{1}{\rho} \frac{dP}{dr} - \omega^2 r \right) \qquad \qquad \text{Equation 15.2}$$

where P and ρ are the gas pressure and gas density at distance r from the Sun. The much smaller force component F_{down} must be in balance with the gas pressure alone:

$$\text{i.e.} \quad F_{down} = -m\left(\frac{1}{\rho} \frac{dP}{dh} \right) \qquad \qquad \text{Equation 15.3}$$

Because of the need to balance contraction towards the disk plane with gas pressure,

the disk temperature must be maintained to preserve force balance. As the disk will be radiating to interstellar space, maintenance of its temperature requires an energy supply, to compensate for the energy losses by radiation from its surface. This energy supply is provided by the central star. To maintain the force balance there must simultaneously be a thermal balance whereby the luminosity of the star controls the temperature distribution throughout the gas disk, and thus controls the gas pressure at any given density via $P = \rho RT/\mu$ (Equation 9.2). Computation of this thermal balance involves a thorough analysis of the energy transfer by radiation and convection throughout the gas disk; in this analysis it is very important to include the ability of the gas to absorb the star's radiation, and to calculate how this gas opacity will vary with density and temperature. The treatment of the opacity is beyond the scope of our course (see for example "Numerical Models of the Primitive Solar Nebula" by A.G.W. Cameron and M.R. Pine, Icarus, vol. 18, p. 377 (1973)); it involves computing the absorption at all wavelengths by all the important atomic and molecular species in the gas disk at any given temperature.

If we do not adopt a "minimum-mass" disk model, F_{in} and F_{down} must also include the gravity of the gas disk as well as that of the star. The principles of the calculation remain the same, but the computation of the force balance becomes much more difficult, as F_{in} and F_{down} now depend on the density distribution $\rho(r,h)$ in a more complex fashion. Until fast digital computers became available, realistic computations of the equilibria of massive gas disks were prohibitively difficult.

In either the "minimum-mass" model in which self-gravity of the disk is neglected, or in a model which assumes an appreciable disk mass, the procedure for computing the conditions in an "equilibrium disk" is essentially to guess a trial distribution of mass and angular momentum through the disk, then to examine the force and thermal imbalance in this guess; from the imbalance, it is possible to specify the adjustments to the trial distribution which are needed to bring the model closer to equilibrium. Because variations in P, ρ and T with r and L also adjust the force and thermal balance equations, these adjustments must be repeated many times until the procedure converges on an equilibrium description in which force balance and thermal balance are obtained simultaneously. If the initial guess is too far from being a realistic distribution, there is no guarantee that the adjustments will converge to a solution. (The procedures needed are described in detail by Cameron and Pine in the article referred to above.)

The detailed results depend of course on the assumed disk mass and on the assumed stellar luminosity. Some general features of the equilibrium disks emerge from all models however. These are:

1) Temperature and density in the disk decrease with r and with h--i.e. the gas disk is hotter and denser closer to the star.

2) The scale height of the disk increases with r, i.e., the disk is thicker further from the star.

For a minimum-mass disk and a "solar" star, the conditions near the mid-plane of the equilibrium disk at the orbital distance r_E of the Earth are T $\sim$ 600 K, $\rho \sim 10^{-6}$ kg m^{-3} and scale height H $\sim 10^7$ km (i.e. $r_E/15$). The following graphs show the variation of properties in a minimum-mass disk around the Sun. The density variation is given in a form which is useful later on, removing the height variation by taking the density projected onto unit area of the disk plane; this projected density σ(kg m^{-2}) is known as the surface density of the disk. Note that the surface density falls off abruptly beyond about 30 A.U.; this compares very favourably with the data on the planetary distances in Section 1.6 and planetary masses in Section 4--30 A.U. corresponds to the orbit of Neptune, beyond which we know of only the undersized planet Pluto. The equilibrium disk therefore has about the right size to correspond to the known Solar System.
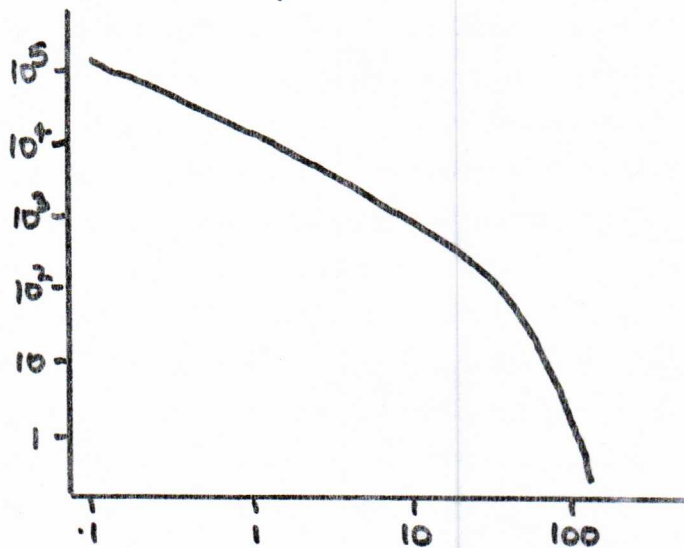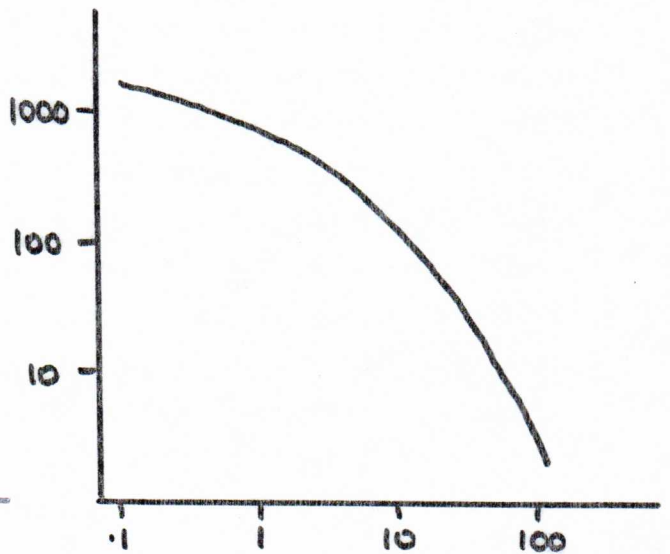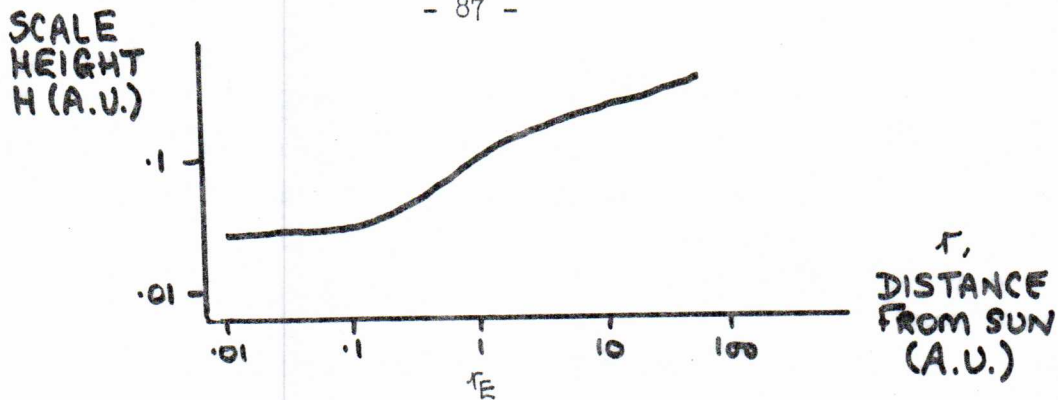
Note that besides having the right <u>scale</u> to form a planetary system such as our own, the equilibrium disk already exhibits the gross kinematics of the Solar System---confinement towards a common plane, and an overall predominant direction of motion around the central Sun. Because the Sun itself has condensed from the same original cloudlet, its rotational equator and rotation direction are also related to the mean plane and orbital direction in the gas disk.

15.2) The Formation of a "Pebble Disk"

We can now begin to trace the development of planets from the equilibrium gas disk. Solid particles will condense out of the gas disk if the partial pressure of a constituent of the gases exceeds the vapour pressure which can be in equilibrium with the solid phase of the material at the temperature concerned. Chemical equilibrium theory can predict <u>which</u> solids will be in equilibrium with a gas disk of solar composition; we will first look at the physical aspect of the process, which leads to the formation of a "pebble disk".

The initial solid particles will be microscopic grains formed randomly by collisions between the appropriate species of atoms and molecules in the gas disk. Consider a solid grain, radius a, density ρ_s, in gas of density ρ_g at temperature T. The gas atoms/molecules will travel at mean velocity $\bar{v}$. In a short time interval Δt the mass of gas colliding with the surface of the grain will be approximately

$$\rho_g \cdot 4\pi a^2 \cdot \bar{v}\Delta t$$
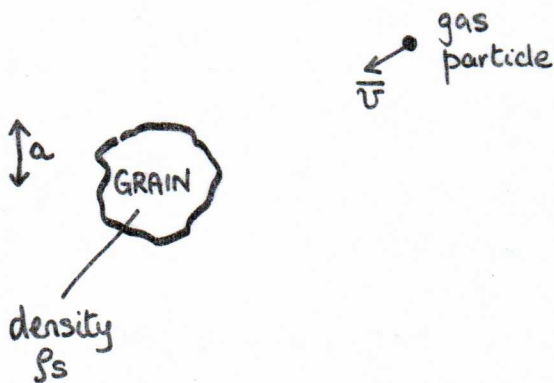
Only a fraction f of this mass will consist of atoms or molecules which are chemically correct for incorporation into the grain. So the mass Δm of CHEMICALLY APPROPRIATE matter striking the grain in time Δt will be

$$\Delta m = f\rho_g \cdot 4\pi a^2 \cdot \overline{v}(f) \cdot \Delta t$$

where $\overline{v}(f)$ is the average velocity of the CHEMICALLY APPROPRIATE atoms and molecules in the gas at temperature T.

The grain will exist only if the temperature T is below the condensation temperature of the solid at the partial pressure of the vapour in the gas disk, so the mass Δm becomes added to the grain, forming a new layer whose thickness Δa must satisfy

$$\Delta m = 4\pi a^2 \rho_s \Delta a$$

Hence the rate of growth of the grain is $\dfrac{\Delta a}{\Delta t} = \dfrac{f\rho_g \overline{v}(f)}{\rho_s}$

This favours the growth of grains at high densities, and of materials which are abundant in the gas provided the temperature is below the condensation temperature. Low-density materials will grow faster because their greater surface areas will sweep up more gas in a given time. Provided the temperature remains below the condensation temperature, condensation is faster in warmer gas because the average particle velocity is higher in warmer gas. (Remember average K.E. $= \dfrac{3}{2} kT$.)

As ρ_s/ρ_g will be $\sim 10^9$, gas pressure cannot support the solid grains. Collisions between the gas particles and the grains can only slightly change the momentum of the grains, which must therefore fall to the mid-plane of the disk. For the relevant grain sizes, which will be much less than the mean free path of the gas molecules, it can be shown that the gas will exert a drag force on a grain which varies with the velocity of the grain:

$$F_{drag} \sim \pi a^2 \rho_g \, c_s v_s$$

Here c_s is the SOUND VELOCITY in the gas and v_s is the grain velocity. As before, a is the grain radius.

The grains will therefore fall at the velocity at which this drag force balances the gravitational force F_{down} towards the mid-plane of the gas disk. To a fair approximation we can write

Calculation of the Gas Drag Law (Notes, page 88).

The gas molecules in the undisturbed gas move randomly at the mean velocity $\bar{v}$; as their motions are random their bulk velocity is zero. When a solid grain moves through the gas at the velocity v_s, the gas molecules strike the surface of the grain and reflect from it. To estimate the drag exerted by the gas on the solid grain we presume that after reflection the gas molecules move randomly relative to the moving grain, i.e. that the mass Δm_g of gas which strikes the grain in time Δt acquires a bulk velocity equal to the velocity v_s of the grain. The drag force on the grain will be given by the rate of decrease of momentum of the grain, which by conservation of momentum must equal the rate of increase of the bulk momentum of the reflected gas molecules, or

$$\frac{\Delta p}{\Delta t} = \frac{\Delta m_g \cdot v_s}{\Delta t}$$

We can relate Δm_g to Δt by noting that the mass of gas $\Delta m_g = \frac{4}{3}\pi a^3 \rho_g$ which could be contained in the volume of a grain of radius a would be swept up in a time $\Delta t = 2a/\bar{v}$ if the mean free path in the gas $\lambda \gg a$. The drag force on a grain is therefore

$$F_{drag} = \frac{\Delta p}{\Delta t} = \frac{4}{3}\pi a^3 \rho_g \cdot v_s \cdot \frac{\bar{v}}{2a}$$

$$\text{i.e. } F_{drag} = \frac{2}{3}\pi a^2 \rho_g \cdot v_s \cdot \bar{v}$$

This drag law, valid when $\lambda \gg a$, is known as Epstein's Law (Epstein, Physical Review, __23__, 710 (1924)). As $\bar{v} = \sqrt{2kT/m}$ in a Maxwellian velocity distribution of molecules of mass m, and the velocity of sound

$$C_s = \sqrt{dp/d\rho} = \sqrt{kT/m}, \quad \text{we can put} \quad \bar{v} = \sqrt{2} C_s, \quad \text{and so have derived the result given on p.88 to the accuracy}$$

$\sqrt{2} \sim 1.5$.

Most currently fashionable models of the formation of the ~~planets~~ Solar System require mutual collisions
gravitational interactions
~~/~~ ~~_____~~ ~~_____~~ to build up/small "seed" ~~_____~~ objects into the final planetary bodies.

While many details of these processes remain controversial, the general character of a successful
theory of planetary
accumulation
~~theory~~ is emerging, and this will be reviewed here.

The first step in planetary accumulation is thought to be ~~GRAVITATIONAL INSTABILITY~~ of
preplanetary   small
~~rather primordial gas distant each nouds from~~ the condensation of/solid particles from the preplanetary gas

disk. Solid particles will condense ~~from~~ out of the disk if the partial pressure of a constituent of the gases

exceeds the vapour pressure which can be in equilibrium with the solid phase of the material at the

prevailing temperature. The chemistry of this process will be studied in Section   ; chemical

equilibrium theory can predict which solids will appear ~~th~~ where in a gas disk of solar

composition. Here we will be concerned with the physical aspect of the process, which leads to the

formation of a "pebble disk".

Settling to the midplane under Epstein Drag

$$F_{drag} = \pi a^2 \rho_g v_s c_s \quad (\lambda \gg a)$$

$$F_{drag} \sim F_{grav} \implies v_s = \frac{4}{3} \cdot \frac{GM_\odot}{r^3} \cdot \frac{\rho_s}{\rho_g} \cdot \frac{ah}{c_s}$$

while

$$a = \frac{f \rho_g \bar{v}_f t}{\rho_s}$$

$$\to v_s = -\frac{dh}{dt} = \frac{4}{3} \cdot \frac{GM_\odot}{r^3} \cdot \frac{f \bar{v}_f}{c_s} \cdot th$$

$$-\frac{dh}{h} = Kt\, dt \qquad \left(K = \frac{4GM_\odot f \bar{v}_f}{3 r^3 c_s} \right)$$

Solution of this is $\qquad h = h_0 e^{-Kt^2/2} \qquad$ for particle at $h = h_0$ at time $t = 0$

This is __exponential decrease in height__ (__infinite__ time to reach mid-plane itself)

Characteristic time of fall is $\sqrt{2/K} = \sqrt{\dfrac{6 r^3 c_s}{4 GM_\odot f \bar{v}_f}}$

So __typical__ grain size at mid-plane is

$$\frac{f \rho_g \bar{v}_f}{\rho_s} \sqrt{\frac{6 r^3 c_s}{4 GM_\odot f \bar{v}_s}}$$

$$= \frac{\rho_g}{\rho_s} \sqrt{\frac{3 r^3 c_s f \bar{v}_f}{2 GM_\odot}}$$

Substitute typical values:

It is then possible to examine which possible solids will condense out of these gases. If the concentrations of gases in the mix exceed those in equilibrium with a solid, the gaseous materials will condense into the solid phase until their partial pressures come to equilibrium with the solids; this causes readjustments in all of the preceding calculations of gas-gas equilibrium. The computer can then "follow" the evolution of the solids from the gas disk, adjusting the gas equilibrium to account for the compositional changes as time passes.

A further stage in the calculations is to examine modifications to the gravitational and thermal equilibrium of the gas disk which result from its chemical evolution and the precipitation of the pebble disk. Changes in the physical equilibrium of the gas disk, if sufficiently rapid, may prevent the chemical equilibrium of the pebble disk going to completion for some species of solid.

The table below shows a typical variation of condensates with temperature at inner-disk gas pressures (10^{-4} to 10^{-3} atmospheres total pressure):

MATERIAL (FORMULA)	CONDENSATION TEMPERATURE	TEMPERATURE OF DISAPPEARANCE
Corundum (Al_2O_3)	1758	1513 ($\rightarrow$ spinel)
Perovskite ($CaTiO_3$)	1647	1393
Melilite ($Ca_2Al_2SiO_7$ + Ca_2MgSiO_7)	1625	1450 ($\rightarrow$ diopside)
Spinel ($MgAl_2O_4$)	1513	1362 ($\rightarrow$ anorthite)
Iron/nickel alloy (Fe-Ni)	1473	below 700, + H_2S $\rightarrow$ FeS
Diopside ($CaMgSi_2O_6$)	1450	below 600, + H_2O $\rightarrow$ oxides
Forsterite (Mg_2SiO_4)	1444	
(Ti_3O_5)	1393	1125 ($\rightarrow$ rutile)
Anorthite ($CaAl_2Si_2O_8$)	1362	
Enstatite ($MgSiO_3$)	1349	
Eskolaite (Cr_2O_3)	1294	
Cobalt metal (Co)	1274	
Alabandite (MnS)	1139	
Rutile (TiO_2)	1125	
Alkali feldspars $(Na,K)AlSi_3O_8$	1000	
Troilite (FeS)	700	

MATERIAL (FORMULA)	CONDENSATION TEMPERATURE	TEMPERATURE OF DISAPPEARANCE
Tremolite (Ca silicates + H_2O)	600	
Magnetite (Fe_3O_4)	600	
Serpentine (Enstatite + H_2O)	425	
Ice (H_2O)	175	
Hydrated ammonia (NH_4OH)	150	
Hydrated methane ($CH_4 \cdot 8H_2O$)	120	
Methane ice (CH_4)	80	
Inert gas solids	50	

Lewis shows that Mercury's bulk density matches that for equilibrium condensation near 1400 K. The composition at these temperatures is dominated by metallic iron, forsterite and diopside because the higher-temperature silicates, perovskite and corundum are formed from elements that are less abundant in the gas composition. The mean atomic weight of these condensates is about 35. In fact it is unlikely that a planet will be formed exclusively from material representing a unique condensation temperature, as this will correspond to a unique distance r in the gas disk. A planet is more likely to form from material condensing out in a zone of distances around the final distance of the planet from the Sun. If we conclude however that 1400 K correctly indicates the mean condensation temperature of the material which formed Mercury, we can use the disk equilibrium calculations to predict the corresponding mean condensation temperatures for the other terrestrial planets; these prove to be $\sim$ 900 K for Venus, $\sim$ 600 K for the Earth, $\sim$ 450 K for Mars. The significance of the Mercury "temperature calibration" is that it eliminates much of the temperature uncertainties that remain in the disk calculations due to uncertainties in the initial disk mass.

Lewis's main result is that equilibrium condensation at the mean temperatures inferred for the other planetary zones would produce pebbles in each zone whose bulk properties would closely match those of the planets now found there.

In the Venus zone, at 900 K, the pebbles would contain all of the Mercury-zone materials, plus a much greater fraction of silicates such as enstatite, and the alkali-metal feldspars. Also important for our final understanding of Venus is the incorporation of some materials containing sulphur. The mean atomic weight is $\sim$ 26.5.

In the Earth's zone, an important difference is that the iron content of the pebbles will be increasingly bound up in iron compounds rather than in the metal, also that some water enters the expected pebble composition. The water is not incorporated as free water at temperatures as high as 600 K, of course; it appears as water of crystallization within the mineral structures. The mean atomic weight of the condensates in the Earth's zone is very close to that in the Venus zone, and is $\sim$ 27.

At the distance of Mars, the lower condensation temperature of 450 K means that water incorporation is more important than it was at the distance of Earth, and the iron occurs mainly in oxides and sulphides rather than as the metal. Mass for mass, the pebbles in the Martian zone would contain more water and less iron than these in the Earth and Venus zones. The mean atomic weight of the Martian condensates would be $\sim$ 25.

The curve in the figure to the right shows the variation of the predicted mean density of the pebbles with the condensation temperature. The points plotted are the <u>decompressed</u> mean densities of the terrestrial planets, from Section 6.1. The point representing each planet has been plotted at the temperature in the gas disk at its present orbit, if the Mercury zone is taken to at T = 1400 K, as discussed above. The value of Mercury in providing this temperature calibration is made very evident by this diagram--the steep slope of the expected density-temperature curve in the Mercury zone makes Mercury's mean composition a very sensitive "thermometer" for an equilibrium model.

ρ_s (kg/m^3)

Predicted Mean Density of Equilibrium Pebble Condensate

Mercury

Venus Earth

Mars

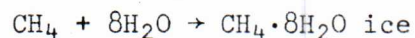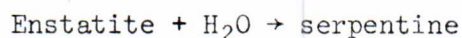Condensation Temp (°K)
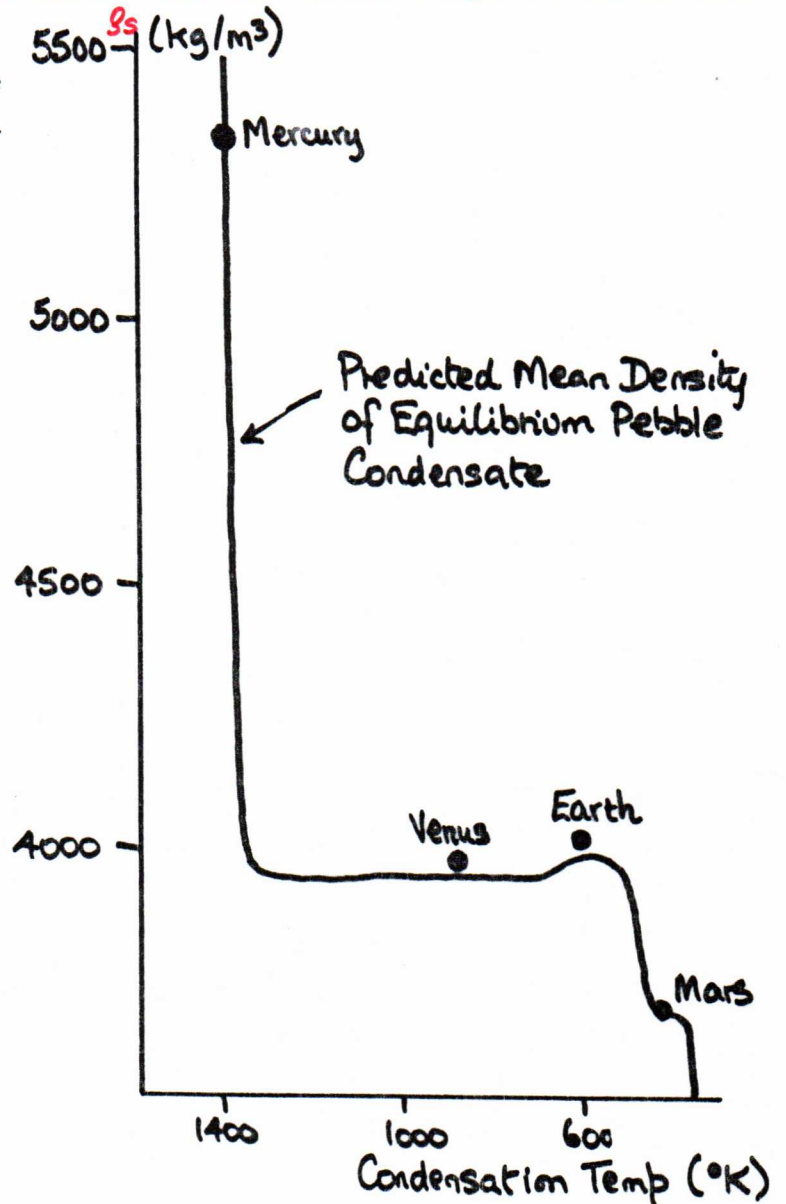
Obviously there is a <u>very</u> good fit between the predicted curve and the planetary data. This means that equilibrium condensation from a gas disk will produce a disk of pebbles whose composition varies with distance from the Sun exactly as required to make planets with <u>bulk</u> compositions matching those of the terrestrial planets. Furthermore, this variety of composition corresponds to condensing an increasing fraction of the gaseous material with increasing distance from the Sun; this feature will help to explain why Venus and the Earth are more massive than Mercury, but about as massive as one another. What is initially mysterious on this picture is the low mass of Mars, which should contain an even greater fraction of the original gas composition; we shall explain the "stunted growth" of Mars later (Section 16.4).

Beyond Mars there are lower gas temperatures, and several <u>major</u> differences in the condensation processes. In the <u>gas</u>, there is greater incorporation of hydrogen into chemically reduced species, e.g.

$$CO + 3H_2 \rightarrow CH_4 + H_2O$$
$$N_2 + 3H_2 \rightarrow 2NH_3$$
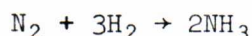
while in the solid condensates we encounter processes such as

$$Fe + H_2O \rightarrow FeO + H_2$$

$$Enstatite + H_2O \rightarrow serpentine$$

$$H_2O \text{ gas} \rightarrow H_2O \text{ ice}$$

$$NH_3 + H_2O \rightarrow NH_4OH \text{ ice}$$

$$CH_4 + 8H_2O \rightarrow CH_4 \cdot 8H_2O \text{ ice}$$

$$CH_4 \text{ gas} \rightarrow CH_4 \text{ ice}$$

These lead to incorporation of the abundant hydrogen, oxygen, carbon and nitrogen into ICE PEBBLES. This increasing condensation of light elements lowers the <u>expected</u> average density of the solids to ~ 1500 kg/m^3. Note that the observed mean densities of the Jovian planets are <u>less</u> than this; their decompressed mean densities would presumably be even less . The gross differences between the Jovian and terrestrial planets can now be accounted for in the equilibrium--condensation model as follows:

1. Hydrogen in the terrestrial region appears only in water of crystallisation of ROCKY material in pebbles.

of the gas will have condensed into solids. <u>The mass of the pebbles inside the region</u>
will therefore be

$$M_D = \pi \frac{D^2}{4} \cdot f \, \sigma_g$$

where σ_g is the surface density of the gas disk at distance r from the Sun (see p. 86).

Under what circumstances will the motions of the pebbles within the region
of diameter D be controlled by the gravitational attraction of the mass M_D, rather
than by the velocity differences resulting from the variation in r over the region?
This is equivalent to asking: Under what circumstances will the mutual attrac-
tions of the pebbles in the region of diameter D overcome the velocity DIFFERENCES
which they have because they are all separately in orbit around the Sun?

Suppose we consider a single pebble,
mass m, whose velocity <u>relative to the</u>
<u>AVERAGE velocity of the region</u> is v',
and whose distance from the centre of the
region is s. <u>Approximate</u> the gravitation-
al attraction of the region, mass M_D, for
this pebble by the gravitational attrac-
tion of a POINT mass M_D at distance s,
i.e. $F = GM_D m/s^2$ towards the centre of
the region. Our question is now: can
this force keep the pebble within the
region against its tendency to drift
out because of its relative velocity v'?
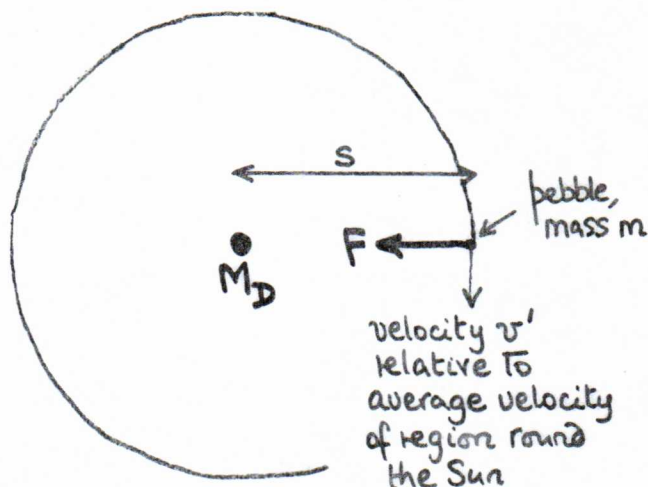
Obviously the "local" gravitational attraction F will be strong enough to
"restrain" the pebble if it could provide the centripetal acceleration that would
be necessary to keep the pebble in a circular orbit of radius s around M_D, given
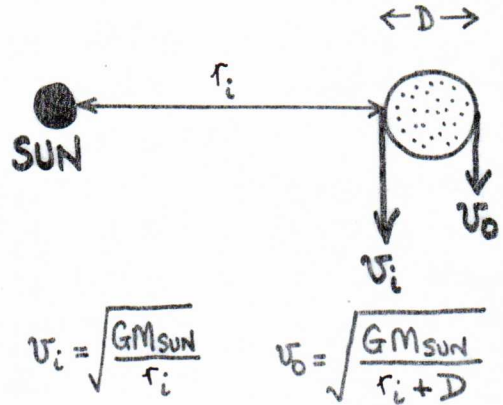the pebble's velocity v' relative to M_D.

NOTE: We are talking about the <u>relative</u> motion of the pebble and M_D
here. Both would share the orbital motion of the region around the Sun; the
question is whether or not "local" gravity will keep the pebble in orbit around
M_D while M_D in turn orbits the Sun.

Evidently the critical condition is $F \geq mv'^2/s$

i.e. $\quad \dfrac{GM_D}{s} \geq v'^2$ $\qquad\qquad\qquad\qquad$ Equation 16.1

Now we use the fact that the velocities are controlled by the Sun, through equation 2.1. By reference to the diagram, you will see that the total velocity in orbit is greatest (v_i) for pebbles on the inside of the region, and (v_o) least for those on the outside, relative to the Sun. Thus v' is greatest for those pebbles closest to, and furthest from, the Sun relative to the centre of the region. This pattern of velocities means that we should put S = D/2 to test the worse case in Equation 16.1. v' depends on D and the rate of change of v with r:

$$v_i = \sqrt{\frac{GM_{SUN}}{r_i}} \qquad v_o = \sqrt{\frac{GM_{SUN}}{r_i + D}}$$

$$v' = \frac{D}{2} \cdot \frac{\partial v}{\partial r} = \frac{D}{2} \frac{\partial}{\partial r}\left(\sqrt{\frac{GM_{SUN}}{r}}\right) = \frac{D}{2} \cdot -\frac{1}{2}\sqrt{\frac{GM_{SUN}}{r^3}}, \text{ doing the differentiation } \frac{\partial}{\partial r}$$

i.e., for orbits around the Sun, $v'^2 = \frac{D^2}{16}\frac{GM_{SUN}}{r^3}$

Equation 16.2

Substituting for M_D, s and v' in Equation 16.1, we get the condition

$$G \cdot \pi \frac{D^2}{4} \cdot f\,\sigma_g \cdot \frac{2}{D} \geq \frac{D^2}{16}\frac{GM_{SUN}}{r^3}$$

i.e. $\boxed{\dfrac{8\pi f \sigma_g r^3}{M_{SUN}} \geq D}$

Equation 16.3

This means that if D is too large, the spread in velocities arising from all the different orbits will keep the pebbles apart despite their mutual gravitation. If D is less than a critical size, the pebble mass inside diameter D can coalesce into a single massive object, or "planetesimal". Note that D will be larger further from the Sun, as both r and f will increase. Eventually, the fall-off in gas density σ_g will reduce D at large distances from the Sun.

In the terrestrial neighbourhood, where the Mg-Fe-Al-Si-S group of elements define the solid composition, f = 0.0035. At Earth orbit, where $\sigma_g = 1.5 \times 10^4$ kg/m^2 in a minimum-mass disk, D will be about 2200 km. Note that this is the size of the region within which pebbles can "fall together", NOT the size of the final

planetesimal. For D = 2200 km in the terrestrial neighbourhood, M_D would be about 2×10^{14} kg. If we assume that the planetesimals at Earth orbit will have mean densities of $\rho_s \sim 4000$ kg/m^3 (the "decompressed" mean density of the Earth) then masses of 2×10^{14} kg would be spheres about 5 km in diameter. About 20 billion of these must be accumulated to make an Earth-sized planet. There is still much uncertainty about the processes that must have dominated this accumulation (see Wood, Chapter Seven).

16.2) Growth From Planetesimal to Planet

The self-gravity of the pebble disk could cause the contraction of regions about 2200 km across near Earth's present orbit, to form planetesimals with masses of order 2×10^{14} kg. A torus 2200 km wide around Earth's present orbit would contain a <u>total</u> pebble mass

$$M_P = 2\pi r_E \cdot f\sigma_g \cdot (2200 \text{ km}) = 1.1 \times 10^{20} \text{ kg, substituting values}$$

At the end of the processes we have discussed so far such a torus would therefore contain about 500,000 individual planetesimals a few km across. These would all be in orbit around the Sun at slightly differing distances, and so would drift away from and towards one another with small relative velocities $v' \sim 0.1$ metres/sec. The calculation by which we estimated the size of the self-stabilising regions also shows that the low-velocity encounters brought about in this band of planetesimals by their differential motions around the Sun will permit their further slow accumulation under gravity. The time scale for this accumulation can be estimated roughly from the time needed for one complete circuit of the Earth's orbit at the RELATIVE velocity v', which is about 270,000 years. Any torus of planetesimals with similar orbits and width $\sim D$ in 16.1 would be able ultimately to aggregate into a few large masses as a result of such slow relative motions. If every such torus produced <u>one</u> massive object, these objects would be $\sim 10^{20}$ kg in mass, which at terrestrial decompressed densities would have diameters of order 350 to 400, comparable to those of the larger asteroids in today's Solar System.

It is not clear whether such a "slow collision" mechanism based on differential orbital motions, or some other process, represents the dominant stage in build-up of solid bodies. Possibly some density fluctuations which were present from the outset grow more rapidly than the "smoothed" theory would predict, to form

the "seeds" around which planetesimals finally congregate to form planets. Gas drag on the planetesimals could also encourage the collapse of clumps of them faster than a simple consideration of the orbital motions would lead us to expect. Large-scale convection in the gas disk may also modify the simple processes described above. In any event, some combination of these effects must have brought about the formation of substantial PROTOPLANETS which then nucleated the final stage of the build-up of the final planetary mass by ACCRETION from a disk of planetesimals onto a small number of gravitationally-dominant objects.

A fairly precise treatment of accretion processes CAN be made, but it is mathematically messy and a simplified approach can demonstrate the essentials. There is a good discussion of accretion in an article by Wiedenschilling, Icarus, vol. 22, p. 426 (1974).

Suppose a protoplanet of mass M, density ρ and radius R accretes planetesimals of mass m from orbits larger than its own by an amount b. The initial relative velocity of the planetesimals with respect to the protoplanet is v'. Then from Equation 16.2, we have, putting b = D/2,

$$v'^2 = \frac{b^2}{4} \frac{GM_{sun}}{r^3}$$

where r is the protoplanet's distance from the Sun.
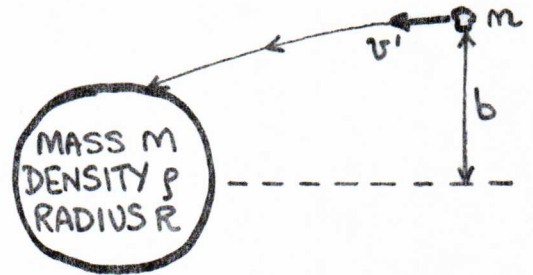
The arriving planetesimals will impact the protoplanet's surface at a variety of angles, depending on the initial value of b (see diagram below). The
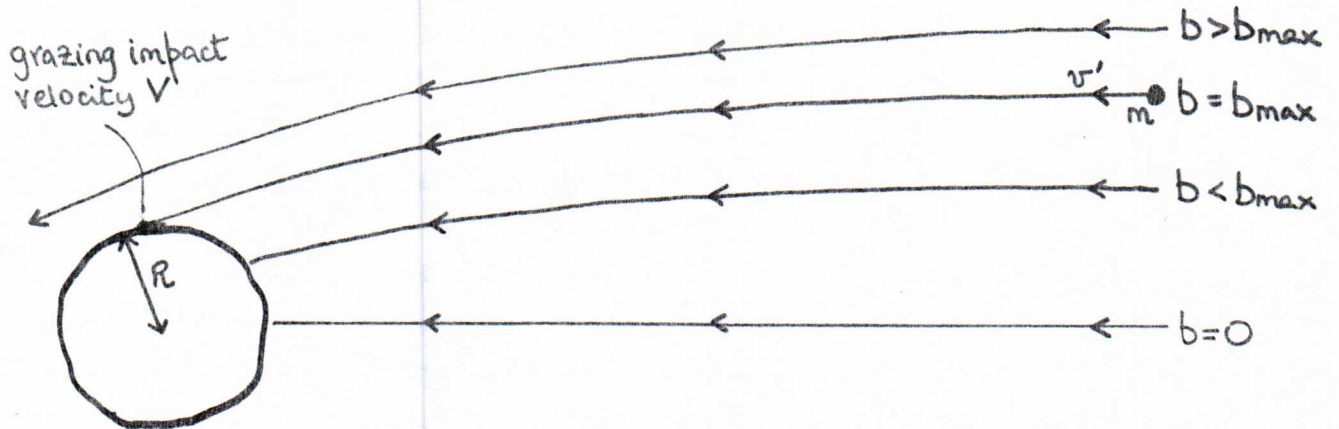
greatest distance b_{max} from which a planetesimal would JUST hit the protoplanet's surface, i.e. striking tangentially, will be greater than the radius R of the protoplanet because of the gravitational "pulling-in" of the planetesimal

motions by the protoplanet. Suppose that planetesimals arriving from b_{max} impact the protoplanet at velocity V.

The gravitational force on the planetesimal due to the protoplanet acts through the centre of the protoplanet, and so cannot change the angular momentum of the planetesimal about an axis through the centre of the protoplanet. The initial angular momentum was $mv'b_{max}$; the final angular momentum is mVR. So <u>CONSERVATION OF ANGULAR MOMENTUM</u> gives $mv'b_{max} = mVR$.

The increase in kinetic energy of the planetesimal as it falls towards the protoplanet must come from the change in its gravitational potential energy. If it arrives from a large distance (relative to R), its initial gravitational potential energy will be negligible, so

<u>CONSERVATION OF ENERGY</u> gives $\frac{1}{2} mV^2 - \frac{GMm}{R} = \frac{1}{2} mv'^2$

Knowing that $v'^2 = GM_{SUN} b_{max}^2 / 4r^3$, and putting $M = \frac{4}{3} \pi \rho R^3$, the above equations can be rearranged to give the key result which expresses the extent of the gravitational focussing:

$$b_{max}^2 = \frac{1}{2} R^2 \left(1 + \sqrt{1 + \frac{128 \pi \rho r^3}{3 M_{SUN}}}\right)$$

Equation 16.4

The protoplanet will be able to "sweep up" every planetesimal coming past it in solar orbit at $b \leq b_{max}$ - its "capture area" for orbiting planetesimals is πb_{max}^2.

Note that b_{max} bears a fixed proportion to R, once the density and location of the protoplanet are known, <u>if the velocities remain essentially those of circular orbits.</u>

The original amount of material in orbits within $\pm b_{max}$ of the Earth NOW would have been about 3×10^{22} kg. This is still only about 1/200 of the actual final mass of the Earth; this indicates the significance of MULTIPLE ENCOUNTERS between protoplanets and planetesimals. "Near misses", i.e. encounters with $b > b_{max}$ may still perturb the passing planetesimal into an orbit which brings about a later accretion collision. Thus <u>the "near misses" can gradually bring more distant material within "accretion range" of a growing protoplanet.</u> Another possibility is that some "near misses" greatly deflect a planetesimal without capturing it, leaving it in a very eccentric orbit which may cross the orbits of several protoplanets; once large protoplanets are present this provides a mechanism

for "cross-contamination" of the Solar System chemistry. "Near misses" must ultimately be responsible for bringing most of the rocky debris into orbits from which it eventually falls onto one of the final planetary bodies.

16.3) The Rate of Accretion (Circular-Orbit Model)
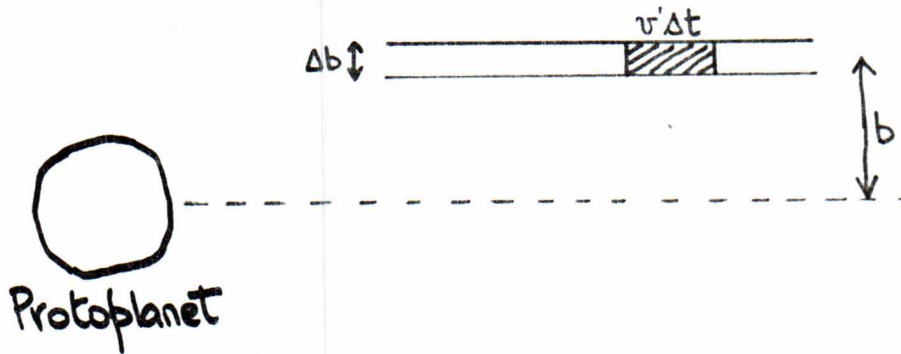
To estimate the rate of accretion, we consider masses arriving from orbits differing in radius from the protoplanets by b, with relative velocities $v'(b)$. In a time interval Δt, all masses in a length $v'(b)\Delta t$ along the protoplanet's orbit will enter the "spatial window" leading to eventual accretion. The mass accreted from the orbits in the range b to $(b + \Delta b)$ in the interval Δt must then be

$$\Delta M = f\sigma_g \cdot v'(b)\Delta t \cdot \Delta b$$

where f = fraction of gas condensed

σ_g = original surface density of the gas disk at that distance from Sun,

so
$$\frac{\Delta M}{\Delta t} = f\sigma_g v'(b)\Delta b$$

The total mass accretion rate (on both sides of protoplanet) is:

$$\frac{dM}{dt} = 2 \int_0^{b_{max}} f\sigma_g v'(b)db$$

While most of the matter is in circular orbits $v'(b) = \sqrt{\dfrac{GM_{sun}}{r^3} \cdot \dfrac{b^2}{4}}$. After multiple near misses have become commom, $v'(b)$ will become much more chaotic than it is for pure circular orbits. Detailed theoretical models have to predict the evolution of $v'(b)$ on a statistical basis as the accretion of the planets progress. This leads to a mathematically complex problem where b_{max} itself depends on $v'(b)$, and the analysis has to be handed over to digital computers to become tractable. For Earth, the pure circular-orbit model would only give the protoplanet

access to ~ 0.5% of the Earth's actual final mass. This demonstrates that the
later chaotic stages of the real process are important. In treating the later
stages of accretion, Wiedenschilling, in Icarus, vol. 27, 161 (1976) uses the concept
of an accretional "feeding zone" of each planet. This "feeding zone" is the flatt-
ened torus around a protoplanet's orbit from which it can ultimately accrete all
planetesimals/pebbles, etc., taking account of the scattering of material away from
its initial orbits by the protoplanets themselves. The final "feeding zones" of the
present-day planets must have overlapped, in general, for if they did not, then either
other planets would have formed in between the actual ones or bands of debris would
remain between the planets now. On the other hand, the final planets must be
outside one another's "feeding zones". If they were not, these planets would
continue to accrete each other. This last constraint can ultimately account for
the non-random spacing of the planetary orbits. It has been shown that the type
of spacing now found minimises the time-averaged gravitational interaction among
the planets.

Theoretical and numerical models have been used to attempt to estimate for
each planet its ACCRETION TIME, i.e. the time taken to grow from 1% to 99% of its
final mass. Typical models in which the feeding zones meet the above requirements
predict the Earth's accretion time to be ~ 2×10^8 years. The accretion time of
Mercury would have been faster, perhaps ~ 5×10^7 years, while that of Mars would be
slower, perhaps ~ 10^9 yrs., due to its low mass and its competition with Jupiter for
material between their two protoplanets.

If we look at the density distribution among the planets now with the feeding-
zone idea in mind, i.e. smear each planet's mass out over its "feeding zone" during
accretion stage, then the following original density distribution is implied for
the Solar System material which actually condensed to form planets:
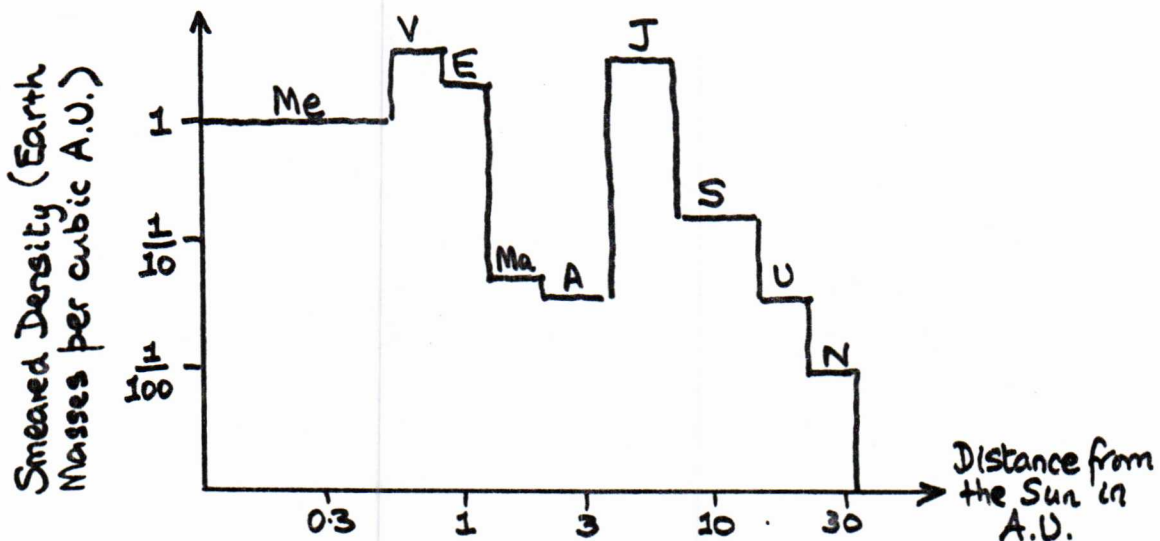
The 100:1 "jump" in the smeared density between the asteroids and Jupiter is much too great to be due to the increased condensation fraction f in outer Solar System. This implies that Jupiter has probably "poached" on material that would have been accreted by Mars and the asteroids had the massive Jovian protoplanet not formed early on. The fact noted in Section 15.3, that Mars is less massive than might be expected from condensation sequence calculations, can therefore be explained if Jupiter's protoplanet formed rapidly and depleted the Martian accretion "feeding zone" of material that could have added to the mass of Mars.

(Note that Earth's gravitational attraction at Mars is only 16% of Jupiter's (both at closest approach to Mars), so that Earth is less likely to be a cause of the Martian "stunted growth".)

The minor planet, or asteroid, belt may also be material that was prevented by Jupiter from accreting into a planet, because of the tides raised in this material by Jupiter's gravity.

17) An Unsolved Problem - Dissipation of the Gas Disk

The gas disk from which the solids condensed is not present in the Solar System now, and two lines of evidence suggest that it dissipated quite early in the history of the planetary system.

First, as we noted in Section 15.1, the variation in gas density and temperature outwards from the Sun produces a pressure gradient in the gas such that gas pressure decreases with distance from the Sun. This gas pressure gradient provides a small outwards force on any volume of gas which slightly reduces the effect of the solar gravity as the volume orbits around the Sun. From Eq. 15.2, the gas at distance r from the Sun will orbit at a velocity v_g given by

$$\frac{v_g^2}{r} = \frac{GM_{Sun}}{r^2} + \frac{1}{\rho_g}\frac{\partial P}{\partial r}$$

where ρ_g is the gas density at distance r and $\partial P/\partial r$ is the gas pressure gradient. As P decreases with increasing r, $\partial P/\partial r$ will be negative, so the orbital velocity of the gas will be slightly less than that of the solid particles, which do not exert a bulk pressure and so obey the law $v^2 = GM_{Sun}/r$. This velocity difference $(v - v_g)$ means that the GAS ALWAYS EXERTS A DRAG ON THE SOLIDS. As discussed in Sec. .o 15.2, gas drag on the pebbles produces a force proportional to the cross-sectional

area of the solid, to the gas density, and to the velocity difference between the
solid and the gas. The deceleration produced by this force is inversely proportional
to the mass of the solid, i.e. inversely proportional to its volume. Gas drag is
proportionally more important for the small solid pieces than for the large, but
as the accumulation to larger and larger solid sizes takes progressively longer
and longer times, the effects of gas drag are never completely negligible. As long
as the gas disk is present, the gas drag will produce a slow inwards drift of the
pebbles, planetesimals and protoplanets towards the Sun. It is important to check
that gas drag will not in fact make all of the solids fall into the Sun before they
could accumulate by the processes we have described; this has been done, for example,
by Goldreich and Ward (see Section 16.1). It is equally important that the gas disk
eventually be removed, or else the "finished planets" would ultimately spiral into
the Sun.

Second, the gravitational attraction of the finished planets would allow them
to accrete significant gaseous atmospheres directly from the gas disk. Their attrac-
tion would lead to permanent increases in the gas density around them, with extensive
transition regions between "atmospheres" orbiting with the planet and "free gas"
in the disk producing the drag. The problem of gravitational retention of gases by
the planets was treated in Section 9.6.

Because the root-mean square velocity v_{rms} is smaller at any temperature
T for the more massive molecules, escape proceeds fastest for the lightest gases
and slowest for the most massive. The atmospheres which the planets could have
retained directly from the gas disk would therefore be depleted from the solar
composition in the lightest elements, by amounts which would depend on the equili-
brium temperature, final mass and size of the planets.

Calculation shows that the finished Earth could not have retained hydrogen for
appreciable times in an atmosphere accumulated directly from the gas disk, but would
have retained heavier species. Chemically reactive species might have been removed
from the atmosphere by subsequent processes of planetary evolution, but the chemi-
cally inert gases, such as helium, neon, argon, krypton and xenon would have been
difficult to remove. Thus the abundances of the heavier inert gases in the Earth
environment now should be an indicator of the extent to which the Earth did retain
a gaseous atmosphere from the gas disk.

Krypton (most common isotope mass 84 a.m.u.) and Xenon (most common isotopes
masses 129 and 132 a.m.u.) are apparently present in the Earth environment at only
$< 10^{-6}$ of their abundances relative to silicon in the solar composition. If they had
been retained in an atmosphere formed from the gas disk by Earth's gravity they should

be present at something much more like the solar ratio to silicon. This implies that BY THE TIME GAS CAPTURE WAS POSSIBLE IN SIGNIFICANT AMOUNTS THE GASES WERE NO LONGER THERE TO BE CAPTURED, in the terrestrial zone. This means that the gas disk must have been dispersed, and any gases previously captured by the Earth re-dispersed, by some process by the time that Earth's accumulation was mostly complete, i.e. by about 10^8 yrs after the formation of the pebble disk.

The Jovian planets Jupiter and Saturn have evidently retained large amounts of hydrogen and helium however. The mean density of the light ices is much too high to match the observed low mean densities of these giant planets, especially when gravitational compression is taken into account. So the bigger Jovian planets must have captured an appreciable fraction of their final mass from the gas disk directly; the gas disk must still have been present in their zone when they accumulated to the point where capture became possible. The accretion-time estimates suggest that the gas disk remained in the Jovian zone about 10^7 yrs.

In summary, the evidence suggests that the gas disk material was no longer in the inner Solar System about 10^8 years after solid condensation began, but remained in the outer Solar System about 10^7 years after condensation. Removal of the gases by about $10^{7.5}$ years after formation of the pebbles would also permit build-up to planets without significant loss of material into the Sun due to gas drag. In fact, gas drag actually helps the accretion process in the early stages by braking any unusually high-velocity material; because the drag force is velocity-dependent it tends to reduce velocity differences among the solid particles and thus favours accretion over dispersion.

The mechanism for dissipation of the gases is controversial, however; it is sometimes referred to, a little cynically, as the "magic broom" because it is obviously necessary in the context of all versions of the disk-accretion theory yet it has to be provided in the right amount and at the right time to do its job, rather "magically" in relation to our detailed understanding of the processes involved.

A common suggestion for the "magic broom" is a hypothetical T Tauri phase of the early Sun. T Tauri stars are a class of star in gas clouds where star formation is thought to be occurring. They have diffuse envelopes, emission-line spectra which indicate the presence of hot gas close to the star, and relative displacements of details in their spectra implying that material towards the observer has significant velocities of outflow from the star. It is thought that they are expelling mass at rates up to 10^{-5} or 10^{-6} solar masses per year, and that the "near-side" mass flow

accounts for some of their spectral peculiarities. The T Tauri stars are thought to be stars of intermediate mass, like the Sun, so that the mass flow must be quite short-lived.

T Tauri stars are believed to be systems in which a significant fraction of the star's luminosity is converted into bulk motions of its outer envelope when convective motions in its outer layers approach the local sound speed. Under these conditions a variety of mechanical waves can be set up which "overheat" the stellar gases and thus expel them, setting up a "stellar gale" which would sweep aside the residual gas disk around the star. The Sun's present high-temperature "corona" and the gentle mass outflow from it known as the "solar wind" might be less vigorous forms of the same phenomena. We do not know whether all stars of approximately solar mass undergo a T Tauri phase towards the end of star formation, but several hundred T Tauri stars are well-documented and it is not a rare phenomenon. If the Sun did undergo a T Tauri phase, it would certainly have removed whatever atmospheres the inner planets would have accumulated prior to that time, as well as the gas disk.

An alternative proposal for the "magic broom" involves the fact that electrons and ions in the warm inner regions of the gas disk would not be free to travel across the Sun's magnetic field. Magnetic forces would bring about a struggle between gravity, which tries to maintain the orbital velocities, and the solar rotation, which tries to force the ions and electrons to rotate with the solar field. This situation has been studied by Hoyle and Alfven; a possible result is that the Sun's spin is slowed down by the interaction whilst the gas velocities are increased. The ultimate effect is expulsion of the gases from the inner Solar System and "braking" of the solar rotation. The final process of detachment of the gases and equilibration of the system is not completely clear, however and the models for the "magic broom" which are based on magnetic phenomena must be regarded as extremely speculative.

A satisfactory explanation of the "magic broom" remains an outstanding problem in Solar System research.

17.1) Argon in Earth's Atmosphere

The abundant isotopes of argon in the "solar composition" are argon-36 and argon-38. These are very depleted in Earth's composition relative to the non-volatiles. Argon does however comprise almost 1% of the Earth's atmosphere:

about 99.5% of this argon is argon-40. This isotope can be formed by <u>radioactive</u> <u>decay</u> of potassium-40, which could be present in the terrestrial-zone pebbles:

$$K^{40} + \text{orbital electron} \rightarrow Ar^{40} + \gamma\text{ray} + X\text{ray}$$

This strongly suggests that our atmosphere's present argon content was probably formed by radioactive decay of potassium in the Earth's primeval <u>solid</u> mixture. This would account for the unusual argon abundance, but leaves us with the question: How did the bulk of the planetary atmospheres form, if not from capture of gases from the original gas disk?

18) The Gross Evolution of Terrestrial Planets

This evolution has been reviewed by W.M. Kaula, "The Seven Ages of a Planet", in <u>Icarus</u>, vol. 26, p. 1-15 (1975). The <u>main</u> stages of planetary evolution he describes are:

1. heating by accretion, core formation and radioactivity.
2. formation of crust, and plate tectonics.
3. terminal volcanism.
4. quiescence.

18.1) Thermal Evolution of Planets

Five processes play a part in determining the thermal history of a planet. They can be divided into two surface heating mechanisms, and three bulk (volume) mechanisms. The detailed temperature distribution within a planetary body at any moment will depend on the interplay between all five processes.

<u>The first process is heating of the surface by absorption of solar radiation</u> at or near the surface. This determines the LONG-TERM surface temperatures once accretion is essentially complete and once the atmospheric composition and structure have stabilised. The basic features of this process were discussed earlier; note that if the interior of the planet is at a temperature different from that of the surface, heat will travel to or from the surface by thermal conduction, and conductive loss or gain of energy by the surface must be included in detailed calculations of planetary thermal equilibrium. Precisely, any square metre of the planetary surface obeys

$$(1 - A)\, S(t) = \epsilon\sigma T_s^4 + K\frac{\partial T}{\partial R} \quad [\epsilon = \text{emissivity of actual surface relative to black-body at temp } T_s]$$

where $S(t)$ is the solar illumination in watts/m^2 as a function of time t (allowing for changing presentation towards the Sun), T_s is the surface temperature, K the thermal conductivity of the planetary material and $\partial T/\partial R$ is the gradient of internal temperature T with radius R in the planet. K is about 3 W/K/m for silicate minerals, so conduction will be less than 1% or so of radiation from the surface once the temperature gradient $\partial T/\partial R$ is less than about 1 K/m.

The second process is heating of the surface by the thermalisation of kinetic energy brought in by the impacting bodies during the accretion of the planet. Each body falling onto the planetary surface releases energy as heat which is a) radiated into space, b) used to raise the surface temperature, c) used to raise the temperature of the impacting body material, d) used to perform chemical and physical changes in the impacting material, e) conducted into the planetary interior. For a planet, mass M, radius R during accretion the above relation becomes:

$$(1-A)S(t) + \rho_i\left[\frac{v'^2}{2} + \frac{GM}{R}\right]\frac{\partial R}{\partial t} = \epsilon\sigma T_s^4 + \rho_i\left[C(T_s-T_i) + L\right]\frac{\partial R}{\partial t} + K\frac{\partial T}{\partial R} \qquad \text{Equation 18.1}$$

where ρ_i and T_i are the density and temperature of the infalling matter, v' is the relative velocity between the infalling matter and the planet when the matter is far from the planet, C is the specific heat of the infalling matter and L the latent heat associated with any significant phase changes in the infalling matter as a result of its impact. $\partial R/\partial t$ is the rate of growth of the planetary radius R with time t as accretion proceeds, related to the rate of growth of the planetary mass M by

$$\frac{\partial M}{\partial t} = 4\pi R^2\, \rho_i\, \frac{\partial R}{\partial t}$$

The complexity of this situation is shown in order to illustrate the variety of factors to which the surface temperature T_s is sensitive at any time t during the accretion process. In practice the most significant parameter which will vary between different descriptions of the accretion process is $\partial M/\partial t$ throughout the accretion, i.e. $\partial R/\partial t$ in equation 18.1. In Wiedenschilling's models (Icarus, vol. 27, 161 (1976)) $\partial M/\partial t$ reaches a peak of about 10^{-8} M.yr^{-1} about 1/3 of the accretion time through Earth's accretion. The maximum $\partial R/\partial t$ is a few cm/yr and the surface

temperature is never more than 100 K higher than it would be due to solar radiation alone. Individual impacts, of course, would produce massive temperature increases around the point of impact due to the strong localisation of the energy release; the impact of a typical planetesimal on Earth's surface would release an amount of energy equivalent to several million megatons of TNT being detonated. However, Weidenschilling's calculations suggest that the GLOBAL rise in surface temperature would be rather small. Accretional heating of the surface is only very significant in models which contrive to accrete the planets much more rapidly than the 10^8-yr time scale suggested by Weidenschilling's calculations. None of the accretion models which deals with the evolution of the relative velocities v' due to "near-misses" produces very short accretion times, but an unconventional model by Cameron (Icarus, vol. 18, 407 (1975)) in which planets form in a turbulent early phase of the gas disk does lead to accretion times of only a few thousand years. Cameron's model would form very hot (surface temperatures of several 1000 K) planets in the inner Solar System.

The third process is adiabatic compression of the planetary material, i.e. the action of the final planet's gravity in compressing the planet to a mean density greater than that of the original planetesimals which formed it. For the Earth, gravity compressed material from a mean density of 4030 km/m^3 to a mean density of 5520 kg/m^3 according to the interior models discussed earlier; the gradual decrease in gravitational potential energy $-3GM^2/5R$ as R decreased due to Earth's self-compression would be expected to produce temperature increases of several hundred Kelvin throughout much of the planet. The compression is most severe towards the centre of the planet, and temperatures could reach well over 1000 K there. Note that adiabatic compression heats the interior of the planet, not the surface, and will set up a temperature gradient $\partial T/\partial R$ in which T decreases with R throughout most of the planet. Thus accretion warms the planet from the outside inwards while adiabatic compression heats it more strongly from the inside outwards.

The fourth process is heating of the planetary volume by energy released in radioactive decay processes. The transformation from a radioactive parent nucleus to a stable product nucleus is accompanied by the release of energy to the outside world as the kinetic energy of the products. If the planet is initially of uniform composition, this heat source acts equally throughout its volume; the heat released must be transported to the planetary surface to be radiated into space. Thus the energy supply due to radioactive heating increases as the volume of the planet, whereas the energy sink due to radiation increases as the surface area, other factors being equal. Radioactive heating thus has greater effect on the

temperatures of large objects than of small ones. The radioactivities in the planetesimals could not have appreciable effect on their temperatures because of the relatively large proportion of surface to volume in a km-sized body. The same radioactivities accumulated into the final planet, whose size is several thousand km, will be several thousand times more effective in raising the planetary temperature throughout its interior.

The efficiency of a radioactive process in heating a planet depends on the energy released per radioactive process, and the rate of the radioactive processes (the half-life of the particular decay or decays). Short-lived radioactivities (half-lives less than about 10^8 years) will release most of their energy while the material is in the interstellar gas cloud, the gas disk, or the planetesimals. The radioactivities of interest as planetary heat sources are long-lived decays with half-lives of order 10^9 years or more.

The long-lived radioactive species which are sufficiently abundant in Earth rocks to have significant thermal effects are uranium-238, uranium-235, thorium-232 and potassium-40. The Table gives the energy release per decay, the half-life, and the energy release rate per kg per year for these four radioactivities:

ISOTOPE	ENERGY RELEASE PER NUCLEUS	HALF-LIFE (yrs)	ENERGY RELEASE PER KG PER YEAR
U^{238}	76×10^{-13} joule	4.5×10^9	2970 joules
U^{235}	72	7.1×10^8	18000
Th^{232}	64	1.39×10^{10}	820
K^{40}	1.1	1.3×10^9	940

The importance of these radioactivities in heating the Earth depends on the relative abundance of the different species, and on time; as the shorter-lived isotopes decay the contribution of their radioactivity to the thermal balance of the Earth declines exponentially. The radioactive nuclei should originally have been distributed uniformly through the volume of the Earth, according to accretional theory. Estimates of the importance of radioactive heating based on the abundances of uranium, thorium and potassium expected in the original Earth vary due to uncertainties in the abundances, but it is a feature of all calculations that radioactive heating superimposed on adiabatic compression should bring a large fraction of the planetary volume close to or above the melting temperature of iron, between 1800 and 2800 K depending on the pressure. In a less massive planet these conditions

would not be attained so readily, due to the greater efficiency of cooling at smaller sizes and to the lesser initial adiabatic compression. It is thought that iron melt-down conditions would have been reached in the Earth after about $1-2 \times 10^9$ years, at levels a few thousand km below the surface.

The fifth heating process is core formation, or gravitational segregation of denser materials towards the centre of the Earth and "floating" of lighter materials towards the surface. This releases further gravitational potential energy, estimated to be about 2.5×10^6 joules/kg from an initially homogeneous Earth to the present chemically-differentiated Earth. Geochemical evidence suggests that this process has concentrated minerals containing the radioactive species towards the Earth's surface, so it is extremely difficult to make definite predictions about the details of the thermal history at different levels. Once melting has partially occurred due to radioactive heating and adiabatic compression, core formation leads to further rises in temperature in the interior and so the process "feeds on itself" towards the final state in which we now find the Earth, differentiated into liquid iron-rich core, mantle and light thin crust. Melting probably occurs first at intermediate depths, and proceeds rapidly throughout the interior as dense materials drip towards the centre, but the details depend significantly on properties of materials at high pressures which are not easy to estimate.

What is clear from all of the above is that an initially homogeneous, "cold-accreted" object in the inner Solar System should not remain homogeneous or cold, but should develop internal activity whose scale and duration may be strongly influenced by the total mass (size) of the object. This activity can play an important role in planetary evolution.

18.2) Crustal Differentiation, Tectonism and Outgassing

At temperatures somewhat below melting, crystalline substances under stress undergo a slow steady flow, or CREEP. This behaviour has a complicated dependence on temperature and pressure and also on the rate at which ambient conditions change; the same material may flow under slowly changing stresses yet fracture if the same changes occur rapidly. The study of the flow and deformation of solid matter is known as RHEOLOGY; the rheological properties of matter under the conditions that are likely to be important in planetary interiors are generally extrapolated from laboratory studies of rocks and minerals. It is generally believed that a layer of weakness occurs in the top 50-150 km of Earth's mantle known as the ASTHENOSPHERE or RHEOSPHERE. In this layer (which

is a seismic low-velocity region) the mantle material behaves like an extremely viscous fluid whose viscosity is strongly temperature-dependent. This property leads to large-scale convection of heat towards the surface by plumes of hotter-than-average material in which creep motions of order one cm/yr occur. These uprising plumes locally uplift the surface over them to form large "domes" topped by volcanoes, such as the Haggar massif in the central Sahara, or the Hawaiian islands. On Earth, we know that this activity in the upper mantle is intimately connected with three processes of major importance in determining the conditions at the planetary surface: differentiation, tectonism and outgassing.

Crustal Differentiation

As a result of the effective fluidity of the upper mantle, lighter basalt and more Si-Al-rich MAGMAS are squeezed upwards to or near the surface where they are depressurised and cooled, producing a thin crust of materials lighter than the average upper-mantle material. In this fashion the elements aluminum, potassium, sodium, calcium, carbon, nitrogen, hydrogen and helium (produced by radioactive decays), as well as smaller amounts of other elements, have been concentrated towards the surface. The radioactive elements uranium and thorium have also been squeezed upwards despite their large atomic weights; their atomic sizes are too large for them to fit easily into the structures of the denser Mg-Fe-rich (mafic) materials in the mantle, whereas they can quite readily be incorporated into the more open lattice structures of the materials forming the crust. Provided the upper levels of a terrestrial planet become warm enough for creep to be important, we should expect some differentiation of lighter materials and the heavy radioactives towards the surface.

On Earth the crust is further differentiated into a layer of Si-Al-rich (sialic) rock typically about 35 km thick at the continents, and a thinner layer of more mafic material typically about 8 km thick on the ocean floors, both over-laying a more mafic substrate in the upper mantle. Radioactive dating shows that substantial volumes of sialic crust had been formed as long as 3.5 billion years ago, and it seems likely that formation of a differentiated crust was well advanced 0.5 to 1×10^9 years after the beginning of the accretion processes on Earth, probably before core formation had begun on a large scale. Oceanic crust is still being formed by the second process we will consider.

Crustal Differentiation on Earth (Schematic)
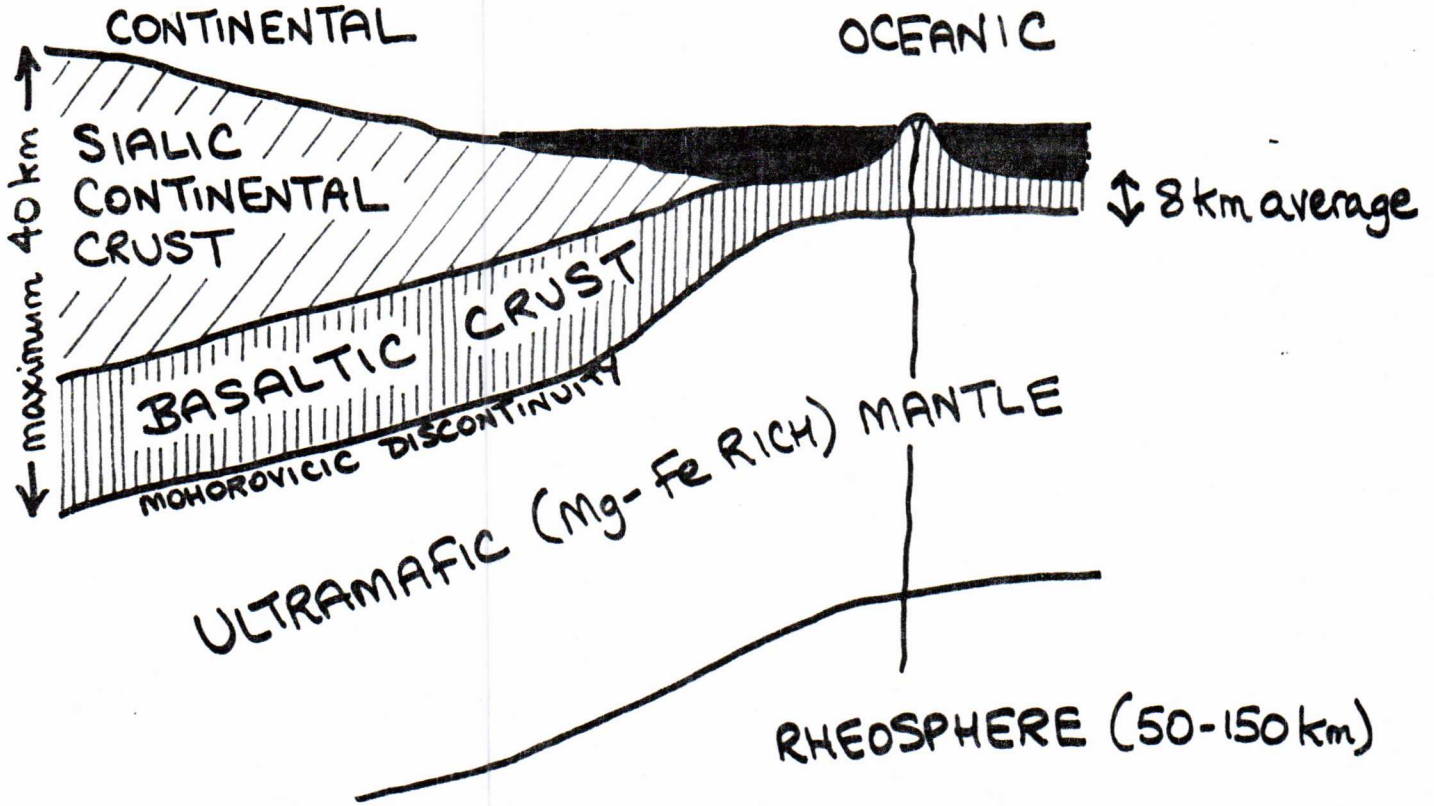
Sialic Crust "Granite"

Si 73%

Al 13%

Na,K 7%

Fe,Mg 3%

Ca 2%

Others 2%

Mean Density ~ 2700

Oceanic Crust "Basalt"

Si 49%

Fe,Mg 18%

Al 16%

Ca 8%

Na,K 5%

Others 4%

Mean Density 2800 - 3000

Tectonism

Earth's crust is broken into large PLATES in relative motion. The plates move apart from rifts at which mantle material is injected into the surface layer, creating "new" crust. The continents generally sit amidst and above plates of oceanic crust whose motions produce what is called the "continental drift". Where moving plates collide, one is pushed under the other back into the mantle at what

is called a <u>subduction</u> zone. Collisions between plates of the thin oceanic crust on Earth produce the deep ocean <u>trenches</u> and volcanic <u>island arcs</u> such as the Aleutians and are sites of deep earthquake activity. Collisions involving the edges of continental masses produce large-scale <u>mountain-building</u> along the continental boundary, as in the Andes. Where adjacent plates slide against one another there are <u>transform</u> faults such as the San Andreas in California. The entire dynamic system is known as global <u>tectonism</u>.

The ultimate driving mechanism for tectonism on Earth is still **uncertain**. It is not clear whether present convection in the mantle maintains tectonism or is maintained by it, through the sinking of the plates at their edges into lighter underlying material in the rheosphere. It is also not clear whether <u>large-scale</u> convection carries the plates around at their velocities of $\sim$ 1 cm/yr or whether the plates simply "slide downhill" under gravity away from the domes over the most important mantle plumes.

What is clear is that <u>magmatism and tectonism now act together on Earth to recycle differentiated crustal materials back through the upper mantle</u>. Subduction carries differentiated material back into the rheosphere, where it can be reincorporated into magmas and be re-injected elsewhere on or near the surface. Such recycling will cease only when the interior of the Earth cools sufficiently to "shut down" creep processes in the upper mantle. A less massive planet formed of the same materials as Earth would have had relatively more efficient cooling and so might never have reached a period of tectonism, or if it had one it might have been relatively short.

Outgassing

<u>The magmas rising from Earth's rheosphere leak fluid and gaseous phases as pressure is released at or near the surface</u>, in volcanic gases, geysers and fumaroles. The mixture of volatile materials exhaled from the magmas depends both on the composition of the magma and on the effective temperature at which the release occurs, and the detailed composition of gases exhaled from Earth's interior today varies from site to site. <u>The major constituent</u> (about 75% by volume) however <u>is water vapour</u>, and it is clear that <u>outgassing of the mantle could have released enough water</u> (embodied in the original planetesimals as water of crystallisation) <u>to have produced the present ocean mass</u>. The next most abundant gas in today's exhalations <u>on average</u> is carbon dioxide, followed by nitrogen and sulphur dioxide. The relative proportions of these gases vary from site to site, but carbon dioxide usually
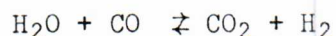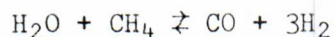
predominates by between two and thirty to one. Other gases emitted as minor constituents of modern outgassing are carbon monoxide, sulphur trioxide, hydrogen, argon, helium, methane, ammonia, chlorine, hydrogen chloride, hydrogen fluoride and hydrogen sulphide.

It is therefore clear that outgassing of a terrestrial planet's interior can provide a "permanent atmosphere" rich in carbon dioxide, nitrogen, argon, etc. which could be retained by the planet's gravity. Indeed, the major constituent of the Venus and Mars atmospheres - CO_2 - is the major component of the dry gas exhaled from Earth today. But on Earth itself, CO_2 is now only a trace component of the atmosphere, leaving nitrogen as the principal constituent (78.1%). Furthermore, the molecular oxygen which is over 20% of our atmosphere now is not present in outgassed vapours. What has happened on Earth to cause this?

18.3) Hydrogen and Oxygen in Earth's Atmosphere

The modern mixture of gases exhaled by the Earth may not correspond in detail to those which formed the first permanent atmosphere by outgassing of its interior. Recycling of the crustal materials through tectonism and magmatism brings material enriched in "modern" crustal composition through the outgassing system. It may be that a truer estimate of the initial atmospheric composition of the outgassing Earth would be that obtained by heating mafic and ultramafic rocks, such as those thought to comprise the bulk of the upper mantle, under laboratory conditions which might simulate magmatic processes. When this is done (e.g. Fanale, Stembridge and Horowitz, Advances in Astronautical Sciences, vol. 25, 165 (1969) and references therein) it is found that while the dry gas is still about 50% carbon dioxide, as much as 40% of the composition can be molecular hydrogen, and methane can account for 3 or 4%.

This suggests that Earth's first permanent atmosphere may have been more hydrogenic than the present one. Molecular hydrogen released by outgassing will escape from Earth's gravity rapidly compared with the billion-year time scale for planetary maturity. Not only that, but chemical equilibria involving atmospheric hydrogen, such as

$$H_2O + CH_4 \rightleftharpoons CO + 3H_2$$
$$H_2O + CO \rightleftharpoons CO_2 + H_2$$
$$\text{(effectively} \quad 2H_2O + CH_4 \rightleftharpoons CO_2 + 4H_2)$$
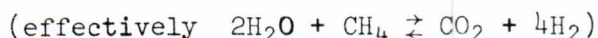
will be driven towards the carbon-dioxide-forming direction as a result of the steady
escape of hydrogen from the environment. Thus loss of hydrogen would gradually
convert hydrogenated species such as methane into carbon dioxide. Atmospheric
weathering of the surface would then increasingly favour the formation of more
highly oxidised minerals, and recycling would gradually increase the carbon-dioxide
content of the outgassed vapours and decrease their content of hydrogenated materials.

There is no plausible combination of temperatures and magmatic compositions
that would lead to the release of significant amounts of molecular oxygen by out-
gassing; the oxygen is taken up into water molecules, carbon dioxide and monoxide,
sulphur dioxide and trioxide, and nitrous or nitric oxide, under all reasonable con-
ditions. Thus Earth today, unlike the other terrestrial planets, shows strong evi-
dence in its atmospheric composition for an agency other than outgassing.

There is strong evidence from the chemistry of certain Earth-surface
materials that the oxygen content of the Earth's atmosphere has grown from $\sim 1\%$
of the present atmospheric content since a time about 1.4 to 1.8 billion yrs before
the present. The main evidence stems from the oxidation state of iron-bearing
minerals present in certain sediments, which is a sensitive indicator of the level
of oxygen prevailing in the atmosphere at the time the mineral composition was
established.

There are certain sands and gravels formed by sedimentary processes which
contain iron, gold and uranium minerals in addition to the quartz grains which
make up the bulk of their composition. These sediments are much studied because
of the economic significance of the gold and uranium content; they are known as
pyrite sands with gold and uranium reefs. From the roundness of their grains,
sorting by grain size, and other properties it is believed that they have undergone
repeated weathering, erosion, transportation and sedimentation processes which
would have brought their contents into contact with the contemporary atmosphere.
Their uranium content allows an estimate of their ages since achieving final compo-
sitional stability, using the techniques of radioactive dating (Section 10).
Their ages can also be estimated by dating igneous rocks which are interbedded
with them, which is often a more reliable procedure in the face of various
processes which can transport radioactive elements and their byproducts in sediments.
It is found that in pyrite sands whose compositions were established more than
1.8 billion years ago, the bulk of the iron is in the form of the pyrite (FeS_2)

which gives the sediments their name, and the uranium is mainly uraninite UO_2.
Neither of these minerals is stable under prolonged exposure to the modern oxygen-
rich atmosphere, and it is estimated that the atmospheric oxygen must have been
less than 1% of the present level at the time these sediments were finally deposited.
In contrast to these "ancient" sands, their modern counterparts contain iron as
magnetite Fe_3O_4 and uranium in the more oxidised form U_3O_8. These indicate that
in more recent times the oxygen content increased sufficiently to destabilise
the "ancient" mineral forms of iron and uranium in these sands.

Much of the early fossil record occurs in sediments interleaved with iron-
rich deposits called banded iron formations; these are sometimes used as low-grade
iron ores and are also well studied. Older deposits contain iron as siderite $FeCO_3$,
pyrite FeS_2 and magnetite Fe_3O_4 as well as in silicates. Younger deposits contain
iron as hematite Fe_2O_3. While this evidence is not as "clean" as that from the pyrite
sands, where the older sediments contain iron mainly as an anoxygenic mineral,
the banded iron formations give evidence for increasing oxidation of the iron at
the time of deposition, going from times 2 billion years ago to times a little
more than 1.4 billion years ago.

Further evidence is provided by the so-called "red beds", which are silts
containing a small proportion of iron as ferric oxide, normally the mineral hematite.
The red beds are thus formed with iron in its most oxidised form, and they occur
only in formations whose ages are estimated to be less than about 1.4 billion years.

The detailed interpretation of this information from the rock record depends
not only on the proportion of molecular oxygen in the atmosphere at the time
these formations were laid down, but also on understanding of the rates of trans-
port and deposition of the materials during the genesis of these formations. It
is therefore difficult to be exact in specifying what the oxygen level was at a
precise time in the past. However, the general trend from pyrite to magnetite to
hematite over the time frame 1.8 to 1.4 billion years before the present is generally
thought to correspond to the crossing of the 1% level (relative to present atmos-
pheric oxygen) during this interval. This would mean that the oxygen level was
"low" for about 50% of the time since the formation of the large masses of sialic
continental crust.

The evidence therefore suggests that a full understanding of the Earth's
bulk atmospheric composition depends on identifying a process which could have
occurred in a primitive environment containing a warm differentiated crust, water
and an atmosphere rich in carbon dioxide and possibly hydrogenated molecular

species such as methane, but lacking <u>free</u> oxygen. The process must be in that sense anoxygenic, but must ultimately lead to the evolution of large amounts of free oxygen in the Earth environment. The process is thought to be life itself.

INDEX