COSMOLOGY AND ASTROPHYSICS

AN ELEMENTARY INTRODUCTION

Pierre Darriulat, Hà Nôi, 2004

CONTENT

1 Introduction

Our knowledge of the universe has made spectacular progress in the past forty years or so, in particular following the discovery of the cosmic microwave background (CMB) that gave convincing evidence in favor of the big bang model. With the advent of space astronomy, new wavelengths became available for exploration as illustrated in Figure 1.1. At the same time, progress in particle physics made it possible to imagine realistic scenarii of the very first minutes of the life of the universe while technical advances in computing have open the door to detailed simulations of the dynamics of very complex gravitational systems of relevance to the evolution of structures over a broad range of scales.

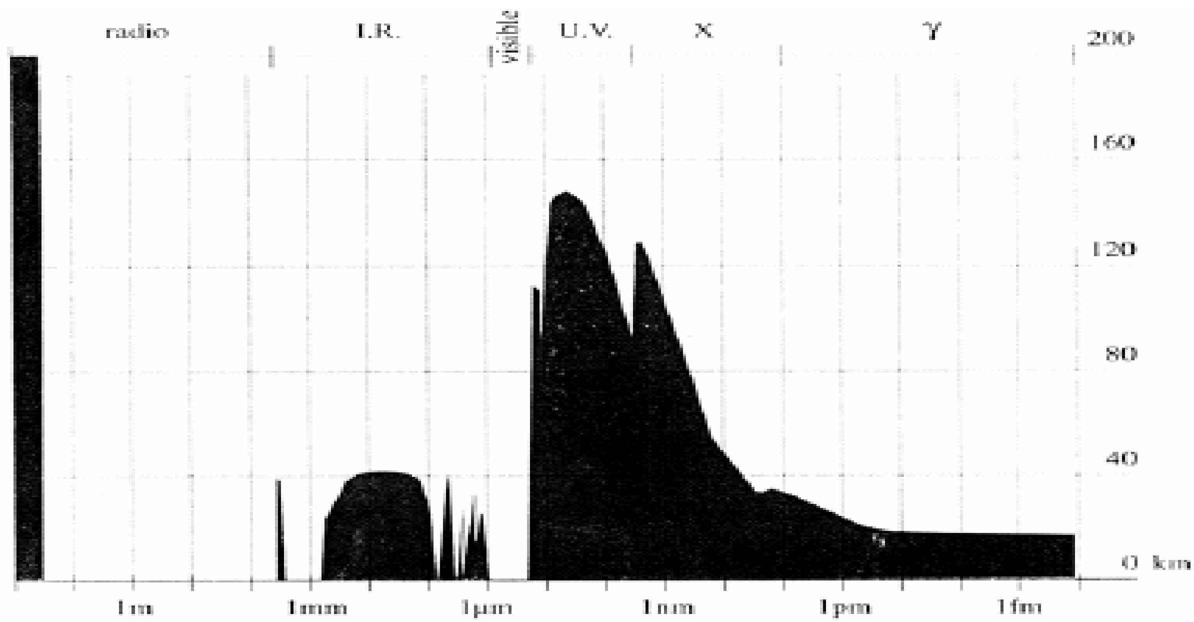


Figure 1.1 : The altitude (in km) at which one has to climb to observe electromagnetic radiations as a function of their wavelength. The visible window separates the radio, microwave and infrared regions on the left from the ultraviolet, X and gamma regions on the right.

The present notes are meant for physics students wishing to learn about this rapidly developing field of science. Their only ambition is to provide sufficient bases for enabling them to listen usefully to more detailed lectures on cosmology and astrophysics. As is often the case for a rapidly developing field of science, many branches of physics contribute to present day's astrophysics: general relativity, particle physics, nuclear physics, plasma physics and magnetohydrodynamics, atomic and molecular spectroscopy. The student who reads the present notes may be ignorant of some of these. In order to give him some idea of what they talk about, rudimentary summaries of their essential results are available in the Appendix.

Much of the current description of the universe can be (superficially) understood from very elementary considerations, many of which require only Newtonian mechanics. In particular a detailed knowledge of general relativity is in most cases unnecessary: at the larger scales the very simple Friedmann regime of a homogeneous and isotropic universe is adequate and at local scales the Newton approximation is in most cases sufficient. Moreover, it turns out that the universe seems to be flat, which highly simplifies the formal description of its dynamics. However, the student wishing to be introduced to general relativity can do it most profitably by reading Einstein's seminal paper (Annalen der Physik, XLIX (1916) 769) that is available in English (in particular in Stephen Hawking, On the shoulders of Giants). The basic ideas underlying the theory are summarized in the appendix.

The field evolves very fast and many earlier arguments have now become obsolete: newcomers should be aware of this fact and not loose time on the study of arguments that have become nearly irrelevant if they wish to catch up quickly with the current developments (I am thinking in particular to old textbooks that are unaware of the extremely rapid developments of the last three decades or so).

Let us briefly recall the main qualitative features of the big bang phenomenology (Figure 1.2): fourteen billion years ago, the universe "started" in a big bang that resulted in an expansion that is still going on today with galaxies recessing from each other with relative velocities proportional to their distances. The proportionality factor is the Hubble "constant", today of the order of 70 km/s/Mpc (1pc*[parsec]* ~ π ly*[light years]* ~ π 10^{16} m). The details of what happened immediately after the big bang are conjectured from what we know of particle physics; in particular from the idea that the symmetries of forces that we observe today, strong, weak and electromagnetic, are the relic of a higher symmetry that was once realized and subsequently broken down. The regime of this so called grand unification theory (GUT) is expected to have taken place at a mass scale of ~ π 10^{16} GeV. Shortly after the big bang, particles and antiparticles annihilated leaving a hot,

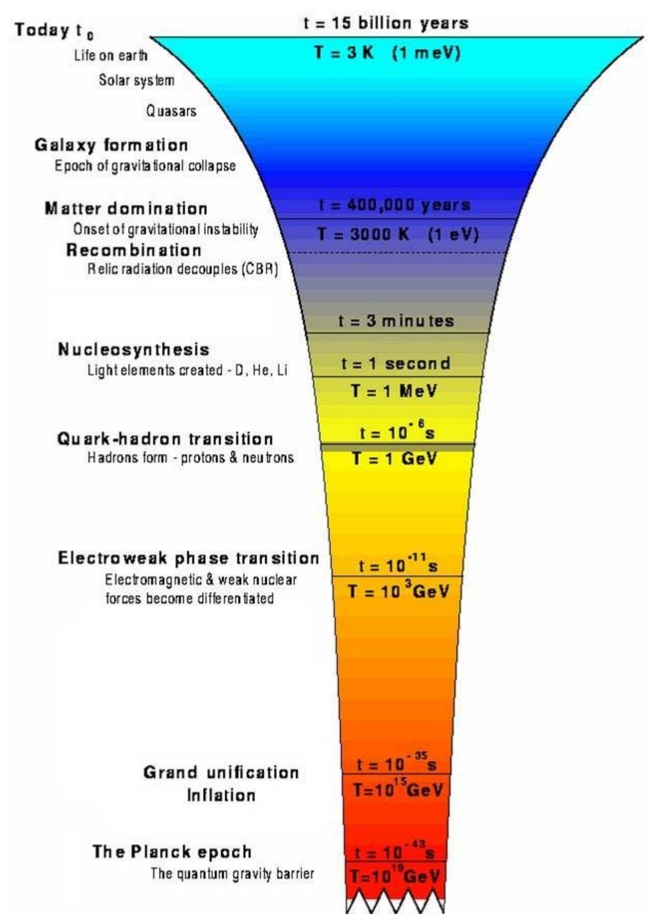


Figure 1.2 : A schematic summary of the history of the universe

high density gas of photons and neutrinos together with a slight excess of matter particles, electrons and nucleons. The evidence that the universe is globally electrically neutral is very strong. Today, on average, it contains about a billion photons per proton and ~3 protons per cubic meter. Quarks and gluons had condensed into hadrons after a few microseconds and nucleons condensed in turn into nuclei, essentially H, He and Li, after a few minutes. For nearly 400 000 years the temperature was high enough for electrons, photons and ionized nuclei to cohabit in thermal equilibrium. Most of the energy in the universe was carried by photons, one says that the universe was radiation-dominated. Then electrons and nuclei combined into atoms, and a few 10^5 years later the energy density of matter overcame that of radiation: the universe evolved from a radiation-dominated to a matter-dominated state. As soon as electrons and nuclei combined into atoms, the universe became transparent to photons as testified today by the presence of a uniform and isotropic CMB with a temperature of 2.7K. After an era of "dark ages" came the cosmic "renaissance", matter starting to condense under the effect of gravity, stars forming, lumped into galaxies that subsequently clustered further into larger structures, clusters, super-clusters, walls and voids. The formation of stars was accompanied by a re-ionization of their atoms. Typical galaxies like our Milky Way formed some 5 billion years after the big bang. The basic steps of stellar evolution, following gravitational condensation of molecular clouds, include a hydrogen burning era during which gravitational and thermal energies are in equilibrium, followed by a gravitational collapse. The final state depends on the star mass, with a broad spectrum spanning from white dwarfs to neutron stars, pulsars and black holes. The collapse sometimes results in a gigantic temperature increase at the surface, a supernova explosion. The present lectures aim at telling this story in more details.

2 Some elementary results of (mostly) Newtonian mechanics.

2.1 Force exerted by a sphere

The force exerted by a spherical shell of uniform density ρ, radius R, thickness dR, on a point mass μ at a distance a from its centre is $\boldsymbol{F}=\int d\boldsymbol{F}_1+d\boldsymbol{F}_2$.
Here (Figure 2.1) the integral is over the spherical shell and the indices 1 and 2 refer to

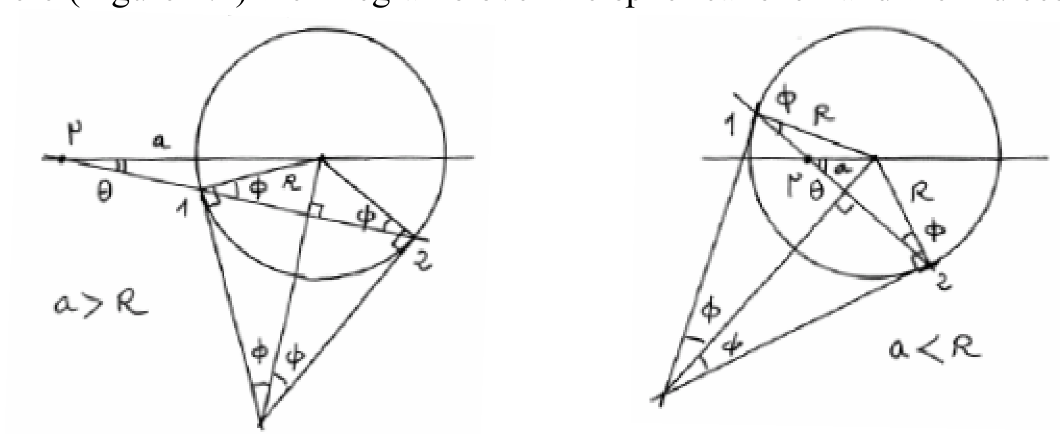


Figure 2.1 : Force exerted on a point by a spherical shell

the two surface elements dS seen from the mass μ under a same solid angle, $d\omega= \sin\theta\, d\theta\, d\varphi=dS/r^2$.

Writing G for the Newton gravity constant,

$dF=(G\mu/r^2)\, \rho\, (r^2 d\omega/\cos\Phi)\, dR = G\, \mu\, \rho\, dR \sin\theta/\cos\Phi\, d\theta\, d\varphi$.

The amplitude of the force is the same for 1 and 2 as r has disappeared and the angles Φ are equal; but for $a<R$ the forces have opposite directions and therefore cancel. The simplicity of this result is a consequence of spherical symmetry and remains valid in general relativity (Birkhoff theorem).

For $a>R$, when integrating over φ, we get a factor 2π and the total force is directed toward the centre of the sphere (hence a factor $2\cos\theta$) with an amplitude

$dF= 2\, G\, \mu\, \rho\, dR\, 2\pi \cos\theta \sin\theta\, d\theta / \cos\Phi$.

The triangular relation $a/\sin\Phi = R/\sin\theta$ gives:

$½ d\sin^2\theta = \sin\theta \cos\theta\, d\theta = ½(R^2/a^2) \sin\Phi \cos\Phi\, d\Phi$, namely:

$dF=4\pi G\mu\rho dR(R^2/a^2) \sin\Phi d\Phi$.

As Φ varies between 0 and $\pi/2$, the total force is simply $G\, \mu\, M/ a^2$, where M is the mass of the shell, $4\pi\, R^2\, dR\, \rho$.

The force is the same as if the mass of the shell were concentrated at its centre. This important result can be extended to the case of a homogeneous sphere (instead of a shell) by integrating over the shell radius R. We then find, calling now R the radius of the sphere,

$F= G\, \mu\, M/ a^2$ for $a>R$ and $F= G\, \mu\, M(a/R)^3/ a^2$ for $a<R$ (2.1.1)

(in the latter case only the mass $M(a/R)^3$ of the inner sphere of radius a contributes).

2.2 Escape velocity

Consider a homogeneous sphere of radius R_0, density ρ and mass $M=4\pi/3\rho R_0^3$. A mass μ at its surface is given an outward radial velocity V_0. Energy momentum conservation gives at any $r>R$ where the mass has velocity $V(r)$:

$½\mu V^2(r) =G\mu M/r +½\mu V_0^2 -G\mu M/R_0$. Define $K= 2GM/R_0 - V_0^2$, then

$V^2(r) =2GM/r -K$ and $dt=dr/\sqrt{(2GM/r -K)}$ (2.2.1)

For $K<0$ we can calculate V up to $r=\infty$ where $V=\sqrt{-K}$. The mass μ escapes the gravity of the sphere. For $K>0$ there is a distance $r_{max}= 2GM/K$ where $V=0$ and beyond which the mass falls back onto the sphere. In between, for $K=0$, the mass just escapes and, on the surface of the sphere, $½\mu V_0^2 =G\mu M/R_0$, giving for the escape velocity the relation

$V_{esc}=\sqrt{(2GM/R_0)}$. Note that $K= V_{esc}^2-V_0^2$ (2.2.2)

Black holes correspond to the situation where the escape velocity exceeds the light velocity, namely where the radius of the sphere is smaller than $R_{Schwarzschild}=2GM/c^2$, the Schwarzschild radius. (2.2.3)

This is just an order of magnitude estimate, a slightly different expression could be expected, but it so happens that this is the right one and we would not learn anything more from the details of the relativistic calculation. The Schwarzschild radius of the sun is 3 km, that of the earth is 4.4 mm.

The equation $dt=dr/\sqrt{(2GM/r-K)}$ can be integrated analytically, giving:

$$Kt=-\{r(2GM-Kr)\}^{1/2}+2GM|K|^{-1/2}F(r)$$

(2.2.4)

with $\tan^{-2}(F(r))=|Kr|/(2GM-Kr)$ for $K>0$,

$\tanh^{-2}(F(r))=|Kr|/(2GM-Kr)$ for $K<0$.

For $K=0$, $dt=r^{1/2}/\sqrt{(2GM)}$ and $t=\frac{2}{3}r^{3/2}\sqrt{(2GM)}$, namely:

$$r=\{3/2\sqrt{(2GM)}\}t^{2/3}. \qquad (2.2.5)$$

2.3 Dynamics of a homogeneous and isotropic medium

We have good reasons to believe that the universe was essentially homogeneous and isotropic at the beginning of its evolution, as testified by the CMB, and that it still is on very large scales (>100 Mpc), I shall comment on this in the next section. It is therefore important to understand the dynamics of a homogeneous medium under the gravity force alone (all other forces can be neglected on such a scale).

In order to do so we start by repeating the argument developed in the previous section and remarking that the movement of the mass μ is independent from R_0 as long as $r>R_0$. In particular, the whole mass of the sphere may be concentrated in its centre. Let us imagine a very different scenario and set $R_0=r$, namely imagine that the sphere expands or contracts in such a way that the mass μ remains always just above its surface. This does not modify in any way the movement of the mass μ. But, now, the mass is at rest with respect to the surface of the sphere. The sphere radius and the sphere density now depend on time, the density like $1/R_0^3$.

Before pursuing, let us comment on the expansion, or contraction, of the sphere. It must preserve its homogeneity. It is convenient to think of a reference frame attached to the sphere and having a unit length, on each of the three axes, that expands or contracts as the sphere does. Such a reference frame is said to be comoving, the coordinates of a point in this frame are called comoving coordinates. Let the unit length, also called the expansion scale, be $a(t)$. A point of the sphere having fixed comoving coordinates (x,y,z) has therefore ordinary coordinates $r=(x\ a(t),\ y\ a(t),\ z\ a(t))$ and a velocity $V=(x\ da/dt,\ y\ da/dt,\ z\ da/dt)$, that is to say

$$V=H(t)r \text{ with } H(t)=da/adt \qquad (2.3.1)$$

We shall refer to Relation (2.3.1) as the Hubble condition and to $H(t)$ as the Hubble constant (constant with respect to r, but t dependent!). It states the proportionality of the distance between any two points and its time derivative. It preserves the homogeneity of the expanding, or contracting, medium.

We now evaluate the forces acting on a mass μ located inside the sphere, at a distance λr from its centre (remember r depends on t as discussed earlier and R_0 is kept equal to it), $\lambda<1$ constant with time. The potential energy is $G\,\mu M\lambda^3/(\lambda r)$, the kinetic energy is $\frac{1}{2}\mu(\lambda V)^2$; their difference is therefore $\frac{1}{2}\mu\lambda^2(2GM/r-V^2)$ which does not depend on time. Indeed, the quantity in parentheses is what we called K before. Thus, for any point inside the sphere, any increase of its gravitational potential is exactly compensated by a decrease of its kinetic energy, it is therefore at rest in the comoving frame. Its K value is simply $\lambda^2 K$. This implies that we do not need to keep the sphere together by some magic additional force, all points of the sphere move precisely as they should to satisfy Newton dynamics. We have therefore found the movement of a spherical homogeneous medium under its own gravity.

The last step is to let the sphere radius become very large compared to the size of the volume of the medium we wish to consider (in its interior). We do not wish any longer to talk about R_0 or M, but only about the time dependent (but space independent!) density $\rho(t)$. We then replace M by $(4/3)\pi r^3\rho$ and write:
$(8/3)\pi G\rho r^2-V^2=K$ with $V=Hr$. This is obeyed for any choice of r, it is sufficient to write it down for our chosen scale, $a(t)$, that describes completely the movement of the medium. Also replace da/adt by $H(t)$ and divide by a^2 : $(8/3)\pi G\rho-H^2=K/a^2$. This relation, which we rewrite as

$$H^2=(8/3)\pi G\rho-K/a^2, \tag{2.3.2}$$

together with the Hubble relation

$$H=da/adt, \tag{2.3.3}$$

summarize completely the movement of an homogeneous medium under the action of its own gravity. Here a, ρ and H are time dependent but space independent, while K is time independent but depends on the choice of scale. Indeed, there is arbitrariness in choosing the scale: we may choose it as we wish, K will simply scale as a^2. Do not be misled by this assertion: of course, K, which is time independent, does not scale as a as a function of time! What it means is that choosing a scale $a'=\lambda a$, with λ constant in time, implies having $K'= \lambda\, K$. Depending on its sign, the medium will either infinitely expand (if $K<0$) or contract after having somewhat expanded (if $K>0$). The limiting case, ($K=0$), corresponds to what is called the critical density, $\rho_{crit}=3H^2/8\pi G$. A medium having a density greater than the critical density will stop expanding and contract at some point (ending up in what is called a "big crunch") while a medium having a density inferior or equal to the critical density will expand for ever. We may rewrite $H^2 = (8\pi G/3)\rho - K/a^2$ as $K/a^2=H^2\,\rho/\rho_{crit}-H^2$

$$K/a^2= H^2(\Omega-1) \text{ with } \Omega= \rho/\rho_{crit} \text{ and } \rho_{crit}=3H^2/8\pi G \tag{2.3.4}$$

Finally we may write the so called "force equation" stating that the acceleration $dV/dt=-(4/3)\,G\,\pi r^3\rho/r^2$, namely

$$d^2a/dt^2 = -(4/3)\pi G\rho a \qquad (2.3.5)$$

There is no new information in this force equation. Indeed it is simply obtained by differentiating Relation (2.2.1) with respect to time: $d\{(8/3)\pi G\rho a^2 - (da/dt)^2\}/dt = 0$
$2\, da/dt\ d^2a/dt^2 = (8/3)\pi G d(\rho a^2)/dt$ and, as $d(\rho a^3)/dt = 0,\ a d(\rho a^2)/dt = -\rho a^2 da/dt$ namely $d^2a/dt^2 = -(4/3)\pi G\rho a$.

Relations 2.3.2, 2.3.4 and 2.3.5 are different expressions of the solution of our problem.

2.4 The cosmological "principle" and the Robertson-Walker metric.

The Hubble relation, $V(r)=Hr$, is in fact a vector relation: we have for any two points $\boldsymbol{r_1}$ and $\boldsymbol{r_2}$ the same proportionality relation between distance and relative velocity, $\boldsymbol{V}(\boldsymbol{r_1})\text{-}\boldsymbol{V}(\boldsymbol{r_2}) = H(\boldsymbol{r_1}\text{-}\boldsymbol{r_2})$. No point is playing a particular role, as expected in a homogeneous medium.

Of course, we do not know whether the universe is really homogeneous and isotropic. In fact, we know that we can explore only a tiny fraction of it. Moreover, we also know very well that today, at small scales, the universe is obviously not homogeneous. Here "small" means up to distances of 100 Mpc or so (not that small: nearly 1% of the horizon!). Indeed structures have been observed up to such a scale and simulations have been able to mimic them (Figure 2.2). Note that the Hubble diagram, relating velocities to distances, starts showing evidence for expansion at redshifts ~ 0.1.

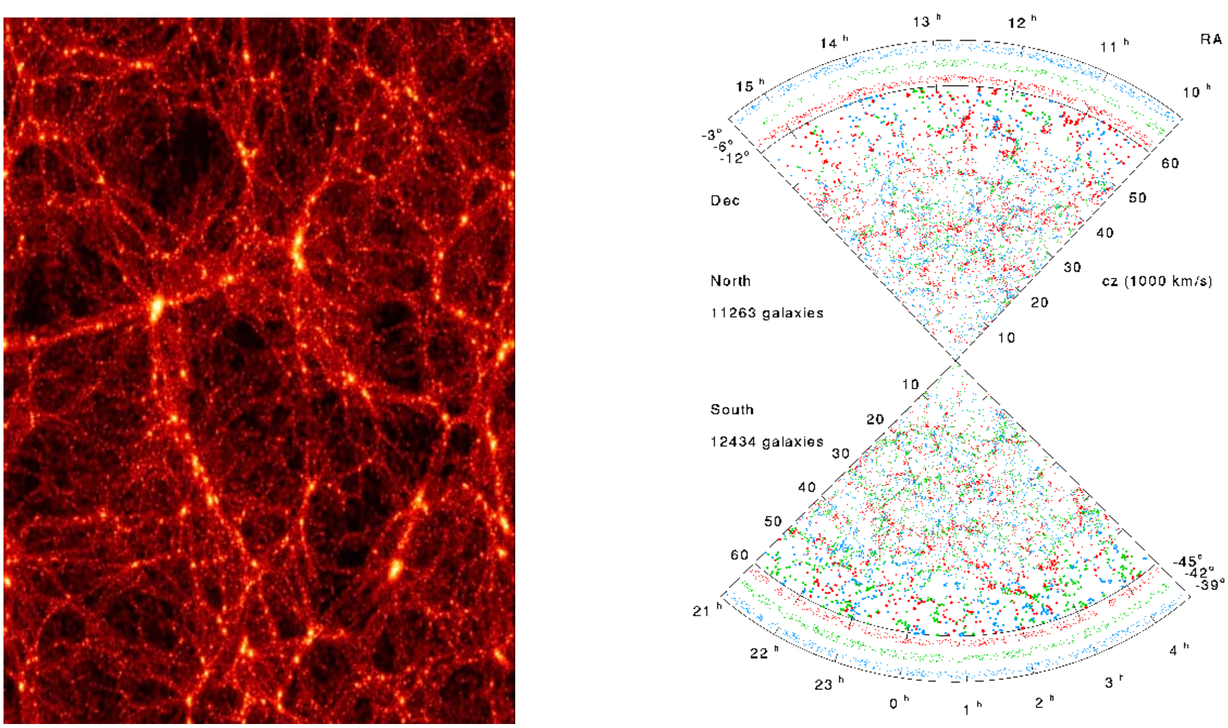


Figure 2.2 :At very large scales of several Mpc the matter is still unequally distributed across the universe. It shows voids separated by walls that are clearly evidenced when considering relatively narrow slices of the universe as is done in these figures. The figure on the right is the result of observation while that on the left is a simulation.

We do not know what happens close to the big bang where general relativity and quantum theory are no longer compatible. Consider a wave packet having an energy spread ΔE and dimension $R=c\Delta t=\hbar c/\Delta E$ (such that they satisfy Heisenberg uncertainty relations). The maximum value that the gravitational energy, $\sim GM^2/R$, can take is ΔE, in which case $GM^2=\hbar c$. For masses higher than the Planck mass,

$$M_{Planck}=\sqrt{(\hbar c/G)} \sim 10^{19}\ GeV \qquad (2.4.1)$$

the gravitational energy exceeds the energy spread of the wave packet and our theories must therefore be revisited. Today, the most popular candidate for a new theory is the so-called M-theory based on superstrings but it is still far from being mastered. The Planck mass is less than three orders of magnitudes higher than the grand unification (GUT) mass, $M_{GUT}>10^{16}\ GeV$, where we claim to have clear enough ideas of the physics to devise scenarii, such as inflation, that are supposed to describe the first 10^{-33} or so seconds of the life of the universe! This illustrates how fragile are the statements that we dare to make in this region of extreme energies.

It is nevertheless sensible, and a reasonable approximation in the universe that we can explore, to take as a ***working hypothesis*** that the universe was homogeneous and isotropic (apart from quantum fluctuations) at the "beginning" and that it remained so in its evolution (apart from the structures that evolved later on from quantum fluctuations). This working hypothesis is called the "cosmological principle", a bad name indeed, physics does not work on principles: if it turns out to be wrong, we will give it up!

We know that general relativity (see appendix) describes gravitation as a distortion of space-time, relating locally its curvature to its energy content. In a homogeneous universe, the curvature must of course be the same in each point. It is then convenient to work with a metric, the so-called Robertson-Walker metric, that embeds the expansion in a time-dependent scale, $a(t)$. Its line element in space time, ds, is simply given by the relation

$$ds^2=dt^2-a^2(t)\{dl^2+kdr^2/(1-kr^2)\} \qquad (2.4.2)$$

where dl is the Euclidian space element,

$$dl^2=dx^2+dy^2+dz^2=dr^2+r^2sin^2\theta d\varphi^2+r^2d\theta^2.$$

Here, l (*i.e.* x,y,z or r,θ,φ) and t are the comoving coordinates and $a(t)$ defines the expansion scale.

For $k=0$ we have a flat universe with an expansion rate described by $a(t)$.

For $k>0$ we have a closed universe (we may see its space part as the three-dimensional surface of a four-dimensional hyper-sphere of radius $a(t)/\sqrt{k}$), and for $k<0$ we have an open universe. Such a metric takes care automatically of the expansion studied in the previous section as we shall see and the meaning of k, as compared to K in the previous section, will soon become clear.

Note that we may rewrite ds^2 as

$$ds^2= dt^2-a^2(t)\{dr^2/(1-kr^2) +r^2sin^2\theta d\varphi^2+r^2d\theta^2\} \qquad (2.4.3)$$

There is some arbitrariness in the choice of a and k. If we define $r'=r\sqrt{|k|}$
$ds^2= dt^2-(a(t)/\sqrt{|k|})^2 \{dr'^2/(1-sgn(k)r'^2) +r'^2\sin^2\theta d\varphi^2+r'^2d\theta^2\}$.
Namely we can take $k=-1,\ 0$ or $+1$ without loss of generality. This choice corresponds to a radius of curvature equal to $a(t)$ when $k=1$, we extent this definition to the case $k=-1$ (that is to say that we define $R_{curv}=sgn(k)\ a(t)/\sqrt{|k|}$ in the general case).
For $k=0$, the Robertson-Walker metric differs from the standard Lorentz metric of special relativity by the only presence of the expansion scale $a(t)$. In that case the choice of scale is completely arbitrary. We may compare the content of this section with that of the preceding section: we had noted that we could choose λ as we please and take a scale $\lambda a(t)$ as long as we use λ^2K instead of K, which implied that it was $a(t)/\sqrt{|K|}$ that had an absolute meaning. We see now that indeed this is essentially the radius of curvature of the universe, K and k are perfectly equivalent, and in the future we shall only consider $k=-1,\ 0$ or $+1$. Note that K has the dimension of a velocity squared and a has the dimension of a length, what we call the "radius of the universe" has therefore the dimension of a time. But this is all right because we have implicitly taken $c=1$ in writing Relation 2.4.1, time and length have the same dimension.

2.5 Friedmann equations.

For a matter dominated homogeneous universe, it is remarkable that the relativistic treatment gives essentially the same result as what has been so simply found in Section 2.3. The relative complexity of the Einstein equations (see appendix) disappears when we restrict them to the case of a homogeneous medium. However, it is no longer the mass density but the energy density that needs to be considered. I will not demonstrate it but I hope that the arguments below will give the student reasonable confidence that it is a sensible result.

Special relativity tells us that a Lorentz transformation reads (with $c=1$)

$z'= z\cosh\alpha+ t\sinh\alpha, \qquad t'= z\sinh\alpha+ t\cosh\alpha,$

where $\tanh\alpha= \beta$ is the velocity of the moving frame and

$\gamma= \cosh\alpha=(1- \beta^2)^{-1/2}\ ,\ \ \sinh\alpha= \beta\gamma.$

This implies that velocities transform as $v'=(v+\beta)/(1+\beta v)$, namely that the light velocity can not be exceeded, and that $\Delta t^2-\Delta z^2$ is an invariant. Special relativity also tells us that energy and momentum form a four-vector if we redefine the energy of a massive particle at rest (a massless particle can not be brought to rest) as its mass m_0. Indeed, bringing it to velocity v with a Lorentz transformation, as $E=m_0$ and $p=0$, we find:

$E'= \gamma m_0 = m_0 +½\, m_0 \beta^2+...\ \ and\ p'= \beta\gamma m_0= m_0\beta +...$

where we see the classical expressions of kinetic energy and momentum appear as the leading terms of the development in powers of β.

Moreover, let us show that it is energy that weighs, not rest mass. Think of a constant vertical gravitational field having g as acceleration of gravity. Take two points A and B on top of each other, A above B at a distance h of it. Send a photon of energy E from A to B. In B the photon has an energy E'. To evaluate it consider the event in the non-inertial system where the gravitational field vanishes. This system starts at zero velocity from A and reaches a velocity gt in B, with $t=h$ being the time it took for the photon to go from A to B. A photon being massless has $E=p$ and the Lorentz transformation reads $E'=\gamma E+\beta\gamma p=E+Egh$ to first order in $\beta=gh$. Namely E has acquired an excess energy Egh in the gravitational field, corresponding to the usual m_0gh term in classical mechanics (m_0 being the rest mass) and it is indeed E and not m_0 which we must consider.

Accepting this result, and redefining ρ as the energy density, the equations that govern the dynamics of a matter dominated homogeneous universe read therefore:

$H(t)=da/a(t)dt,\ H^2(t)=(8/3)\pi G\rho(t)-k/a^2(t),$

$\rho_{crit}(t)=3H^2(t)/(8\pi G),\ \Omega(t)=\rho(t)/\rho_{crit}(t),$

$k/a^2(t)=H^2(t)\{\Omega(t)-1\},\ d^2a/a(t)dt^2=-(4\pi G/3)\ \rho(t).$

The second, fourth and fifth of these equations are different forms of Friedmann equations for a matter-dominated universe. They take exactly the same form as in the Newtonian case studied earlier.

However, the fifth one (that is essentially the derivative of the second one) takes a slightly different form in the case of a matter dominated universe. To see this let us consider a universe made of photons. The wavelength of a given photon expands as $a(t)$, namely its frequency, or equivalently energy, decreases like $1/a(t)$. This result is valid in the relativistic case and can be seen equivalently as the result of the Doppler "redshift" z defined as $1+z=\lambda(t_2)/\lambda(t_1)=a(t_2)/a(t_1)$, where t_2 and t_1 refer to the times of observation and of emission respectively. Note that the relativistic expression for the Doppler shift is $1+z=\sqrt{(1+\beta)/(1-\beta)}$. The energy density, instead of decreasing as $1/a^3(t)$ as in the matter-dominated case (describing the decrease of the number of photons per unit volume), decreases now as $1/a^4(t)$ as the energy per photon decreases as $1/a(t)$. Photons are of course not at rest in the comoving frame, in fact they fly at speed of light!

Consider a volume V containing an energy ρV, say a small cylinder of length l and cross-section S. Change l by dl. The photons exert a pressure p on the cylinder, namely a force pS that makes a work $pSdl=-pdV$ and that must compensate the change in energy $dE=d(\rho V)$, namely $d(\rho V)+pdV=0$. Quite generally we have therefore $d(\rho a^3)/da=-pd(a^3)/da=-3pa^2$, namely, for photons that have $d(\rho a^4)=0$, $ad(\rho a^3)/da=-\rho a^3=-3pa^3$. Photons exert therefore a pressure $p=\rho/3$ that needs to be taken into account in the force equation.

In general by differentiating the second equation with respect to time we get: $2VdV/dt=(8/3)\pi Gd(\rho a^2)/dt$, namely $d^2a/adt^2=(4/3)\pi G(d(\rho a^2)/dt)/(ada/dt)$
As $d(\rho a^3)/da=-3pa^2$, $ad(\rho a^2)/da+\rho a^2=-3pa^2$ and $d(\rho a^2)/dt=-(\rho+3p)(ada/dt)$.

Replacing in the former equation we get

$$d^2a/adt^2=(-4/3)\pi G(\rho+3p) \quad (2.5.1)$$

which is the relativistic form of the force equation. Here $p=0$ for a matter dominated universe and $p=\rho/3$ for a radiation dominated universe. In general the relation between p and ρ is called the equation of state.

The deceleration parameter is defined as

$$q = -d^2a/adt^2/H^2 = \Omega/2+3p/(2\rho_{crit}) \quad (2.5.2)$$

namely a flat, matter-dominated universe decelerates with $q=1/2$.

We may now rewrite Friedmann equations in their general form:

$$H^2(t)=(da/a(t)dt)^2=(8/3)\pi G\rho(t)-k/a^2(t), \qquad k/a^2(t)=H^2(t)\{\Omega(t)-1\},$$
$$d^2a/a(t)dt^2=-(4\pi G/3)\ (\rho(t)+3p(t)), \qquad q=\Omega(t)/2+3p(t)/(2\rho_{crit}(t)). \quad (2.5.3)$$

2.6 Time dependence of the parameters of the universe.

We know that today the universe is close to being flat (see later).

For a flat universe we have $da/dt=\{(8/3)\pi G\rho\}^{1/2}a$. As ρ is inversely proportional to a^3 for a matter dominated universe and to a^4 for a radiation dominated universe, da/dt is inversely proportional to $a^{1/2}$ and a respectively, implying that a varies as $t^{2/3}$ and $t^{1/2}$ respectively (Figure 2.3). Correspondingly $V(t)$ varies as $t^{-1/3}$ and $t^{-1/2}$ respectively.

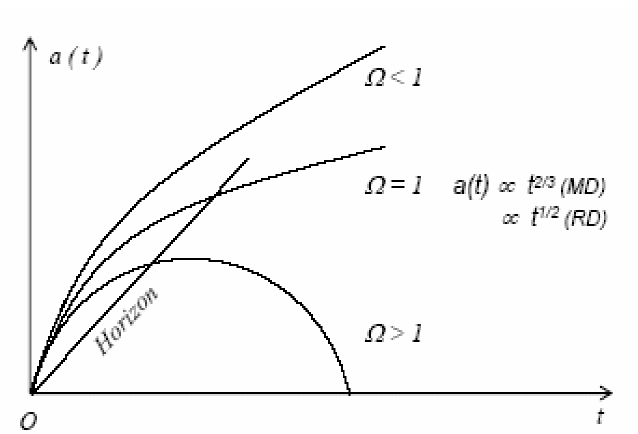


Figure 2.3 : Time evolution of the scale parameter a

From Friedmann equation, $k/a^2(t)=H^2(t)\{\Omega(t)-1\}$, or $k=V^2(t)\{\Omega(t)-1\}$, we see that $\Omega(t)-1$ must decrease as fast as $V^{-2}(t)$ when t decreases and therefore Ω must become closer and closer to 1. In fact, as we know that today Ω is very close to 1, we can infer that at GUT time Ω must equal 1 to an accuracy of 49 decimals! This is sometime referred to as the "flatness problem" because in models (claiming of course to describe the initial conditions at GUT times) that do not imply from first principles that Ω is exactly 1 it would be very difficult to fine tune it to such a precision.

As the universe is essentially flat, it is sufficient to quote the time dependence of the parameters of a flat universe, which is done in the table below.

The last entry, temperature, deserves some explanation. We know that the cosmic microwave background has a black body spectrum with an energy density $(E=h\nu)$

$d\rho/d\nu = 8\pi h\nu^3/(exp[h\nu/k_BT]-1)$, or, writing $x= h\nu/k_BT$,

$d\rho/dx = 8\pi hx^3(k_BT/h)^4/(e^x-1)$. As ρ is proportional to T^4 and to a^{-4}, the temperature is therefore inversely proportional to a, namely proportional to $t^{-1/2}$ and $t^{-2/3}$ in a radiation dominated and matter dominated universe respectively.

Parameter	Matter dominated $p=0$	Radiation dominated $p=\rho/3$
Energy density ρ	t^{-2}	t^{-2}
Expansion scale a	$t^{2/3}$	$t^{1/2}$
Expansion velocity V	$t^{-1/3}$	$t^{-1/2}$
Hubble constant H	t^{-1}	t^{-1}
Temperature	$t^{-2/3}$	$t^{-1/2}$

Knowing how the parameters of the (flat) universe evolve with time makes it possible to infer the age of the universe, t_0, from the knowledge of its present state, $H_0=H(t_0)$.

Neglecting the radiation dominated period we have:

$a(t)=a(t_0)(t/ t_0)^{2/3}$, $da/dt=(2/3)a(t_0)(t/ t_0)^{-1/3} /t_0 =(2/3)a(t)/t$

Namely, $H(t)=da/adt=2/3t$ and $t_0=2T_0/3 = 2/(3H_0)$ where we have introduced the so-called "Hubble age", $T_0=a(t_0)/(da(t_0)/dt)=1/H_0$.

Plugging in the currently accepted value of the Hubble constant, 71 km/s/Mpc, gives an age of only 9.2 billion years. In fact the best current estimate of t_{now} is 13.7±0.2 billion years (WMAP), as obtained from a more realistic account of what the universe is made of (including dark energy, see below). It is nearly equal to T_0.

2.7 The horizon and the causality problem

For two points to be causally connected at time t their distance $d(t)$ must not exceed the "horizon" defined as the distance spanned by light during that time, namely ct. But for some time $t'<t$ these two points must have been causally disconnected. In the matter dominated era it happened when

$d(t')=(t'/t)^{2/3}d(t)=ct'$, or $d(t)/ct=(t'/t)^{1/3}$, $t'=t(d(t)/ct)^3$

Similarly in the radiation dominated era we get $t'=t(d(t)/ct)^2$

This implies that the sky that we see today was in fact causally disconnected at some earlier time. Then how can it be that it is so homogeneous? Even if two causally disconnected regions can always be linked by a chain of causally connected regions, one should be bothered by this situation that is usually referred to as the "causality problem" or "horizon problem". Rather than thinking of causally disconnected regions, one should rather think of the relative gradient describing the space variation of some quantity, say the temperature T of a plasma in thermal equilibrium: dT/Tdl. It is this quantity that is expected to be constrained to be $\geq 1/\Lambda$, Λ being the horizon.

3 Inflation

3.1 Potential problems of the standard big bang model

In addition to the causality and flatness problems already mentioned, the standard big bang model must face also two more difficulties that are potential problems: the absence of monopoles in today's universe and the very large value of the product ρa^4 during the early radiation dominated era.

The first problem stems from the belief that at GUT times there must have been stable magnetic monopoles in such abundance that some should have been observed today in the very sensitive searches that have been performed.

The second problem is that ρa^4 is a constant during the radiation dominated era that should be obtained from first principles. But in natural units ($\hbar=c=1$), it is a pure number and one would therefore expect its value to be commensurate with unity. However a lower limit of this quantity is obtained by considering photons only (the density of which is well known), yielding the result $\rho a^4 > 10^{115}$. This is far from unity and may discourage theorists to devise a sensible model...

It is true that none of these problems may sound dramatic to the skeptical physicist who remembers that anyhow theory must fail at the Planck mass ($\sim 10^{19}$ GeV) that is only 2 to 3 orders of magnitude above the GUT mass ($>10^{16}$ GeV) and that we do not really know precisely what we are talking about in this region. Yet, there exists a model, the inflation scenario, which disposes simply of all these problems and has therefore become popular, indeed gaining credibility with time as its predictions were better and better verified by observation. It has emerged as a very sensible working hypothesis and is now included in the standard model of modern big bang cosmology (one talks of the "concordance model"). Yet, given its very conjectural nature, a hand waving presentation of its main features will be sufficient in the present introduction.

3.2 Negative pressure and false vacuum

Let us assume that at GUT times the universe has been for a while in a meta-stable state, similar to that produced by the potential invoked to describe the Higgs mechanism. In such a state a volume V embeds an energy $\rho_{meta}V$, where ρ_{meta} is the constant energy density associated with that state. We neglect any other form of energy density, matter or radiation. Increasing V by dV simply increases that energy by $dE=\rho_{meta}dV$, the real vacuum being taken as having zero energy density. As $dE=-pdV$

this implies a uniform negative pressure, $p=-\rho_{meta}$. From Friedmann equations (the force equation) $d^2a/adt^2=-4\pi G/3\ (\rho+3p)$ we get $d^2a/adt^2= 8/3\ \pi G\ \rho_{meta}$ which can be easily integrated as $a=exp(Ht)$ with a really constant (both in time and in space) Hubble constant, $H=\sqrt{(8/3\ \pi G\ \rho_{meta}\ )}$.

As ρ_{meta} , in natural units, has dimension of $(mass)^4$ and as the only scale at our disposal is the GUT scale $(>10^{16}\ GeV)$ we expect ρ_{meta} to be of order $10^{64}\ GeV^4$. Putting numbers in gives a value $H\sim10^{36}s^{-1}$ for the Hubble constant. The inflation scenario assumes that the universe was in such a regime from the big bang up to GUT times.

3.3 Inflation confronting observations

We therefore assume that there was exponential expansion due to inflation between a time $<<t_{GUT}$ and t_{GUT}, where $t_{GUT}\sim10^{-33}s$, $Ht_{GUT}\sim 10^3$.

During inflation ρa^4 blew up by a factor $exp(4Ht_{GUT})$ that we should like to be of the order of 10^{115} for the density to join smoothly to its expression in the subsequent radiation dominated era. This means that we should like Ht_{GUT} to be of the order of $0.25\times115\times ln10 = 66$. This is easily achieved by choosing $\rho_{meta}{}^{1/4}=\sqrt{.066}\ 10^{16}\ GeV= 1/4\ 10^{16}\ GeV$. The problem of the large ρa^4 value is therefore solved.

The flatness problem is similarly trivially solved as whatever was the value of Ω before inflation it has been driven very rapidly to 1 during inflation where $H^2(\Omega-1)$, and therefore $(\Omega-1)$ itself, have decreased by a factor $exp(-2Ht_{GUT})$ of order 10^{58}.

The monopole problem also is solved because any primordial particle has been diluted to such an extent during inflation that the probability to detect it today is negligibly small.

Last, let us consider the causality problem. What happened during inflation is that a small causally connected region of the universe has been suddenly blown up to our presently observable universe, solving the causality problem. The universe that we see today may be but a very small fraction indeed of the whole universe (whatever that means). Let us take this opportunity to correct an error that newcomers to the field sometimes make: they often tend to see the universe just after the big bang as a small cloud of very dense matter and radiation expanding in empty space. While the image of an expanding cloud (seen from inside) gives a good idea of the expansion of the universe, that of a finite cloud in empty space is erroneous. It is only the density of the universe that becomes infinite at the time of the big bang, but, all along, the universe keeps filling the whole space. This is surely a much better picture to have in mind than that of a finite cloud but it is still incorrect. In general relativity such a picture makes no sense, space-time exists only as generated by the presence of energy, in particular there is no more meaning to talk of “before the big bang” than to talk of “outside the universe”.

The inflation scenario has the further attraction of explaining the inhomogeneities observed today, galaxies, clusters, etc... as resulting from quantum fluctuations of the inflation field that have been blown up by a factor 10^{29} during inflation, leading to the density fluctuations that acted as seeds for the gravitational condensation of matter into stars, galaxies, clusters, etc...

4 The content of the universe

4.1 Ordinary matter, nucleosynthesis

We have been saying all along that the universe is nearly flat without giving any justification for this affirmation. If it were true we should have $\Omega=1$. It is therefore important to evaluate this quantity from observation. According to present knowledge the visible universe contributes only 1% to Ω. The most precise evaluation of ordinary matter, namely atomic nuclei (atomic electrons obviously contribute a negligible amount), often called for this reason baryonic matter, is obtained from a study of the formation of nuclei during the first minutes of the life of the universe (nucleogenesis or nucleosynthesis). Nuclear physicists can calculate what should be the relative amount of the various isotopes as a function of density. The result of this calculation is illustrated in Figure 4.1. The relative abundance of deuterium is a particularly sensitive indicator. It has been measured with good accuracy, in particular in high redshift clouds of gas (observing the absorption of light from quasars behind them). It gives a baryonic density corresponding to $\Omega_B \sim 4.4\%$. This particular value agrees equally well with the measured abundances of the other light isotopes, He, ^{3}He and ^{7}Li. This success is considered as a major achievement of the standard big bang model. Other, less direct, evaluations of Ω_B give results that are consistent with the nucleogenesis number.

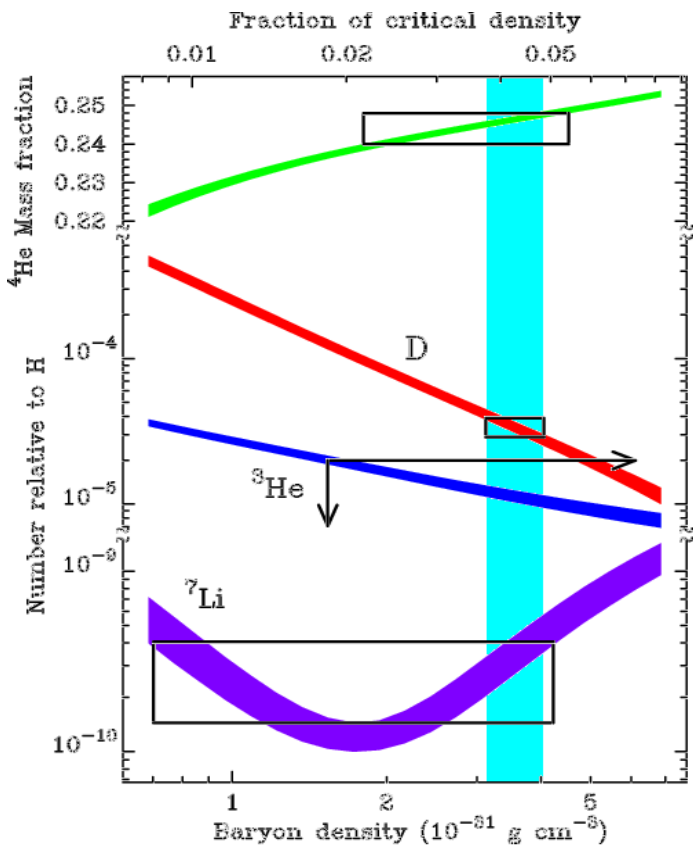


Figure 4.1 : Nucleosynthesis in the early universe. Elemental abundances relative to hydrogen are shown as a function of the baryon density. They show excellent agreement between predictions (bands) and observations (rectangular boxes).

One should remember that most of the heavier nuclei observed today were not produced in the early universe but much later in the interior of stars and in supernovae explosions (see later).

As visible matter accounts for only ¼ of ordinary matter, the remaining ¾ must be in the form of planets, brown dwarfs, dust clouds, black holes and, most importantly, hot gas surrounding galaxies. Indeed one knows of some clusters of galaxies where the amount of matter contained in such gas exceeds that contained in stars by a large factor.

4.2 Dark matter

The gravitational field of galaxies can be studied directly from the motion of stars orbiting in their halos or from gravitational lensing. The result of these observations is that galaxies contain nearly 7 times as much matter than ordinary matter, namely that there must exist a dark matter component $\Omega_{DM} \sim 6\ \Omega_B$. The movement of stars gravitationally bound around a galaxy, far enough from its centre to be outside the luminous region (the luminosity is observed to fall exponentially beyond some radius), provides a measurement of the radial distribution of the mass of the galaxy (Figure 4.2). For an orbit of radius r, larger than the hub radius, one expects each star to have a velocity $v(r)$ such that $rv^2(r)$ is a constant. However it is instead $v(r)$ that is seen to be approximately constant rather than inversely proportional to $r^{1/2}$, implying that the mass contained inside the orbit increases linearly with r, namely that the visible galaxy is embedded in a spherical halo having a density that decreases as $1/r^2$. Comforting this conclusion, one finds from computer simulations that the classical spiral galaxy structure is unstable without such a halo but becomes very stable in its presence.

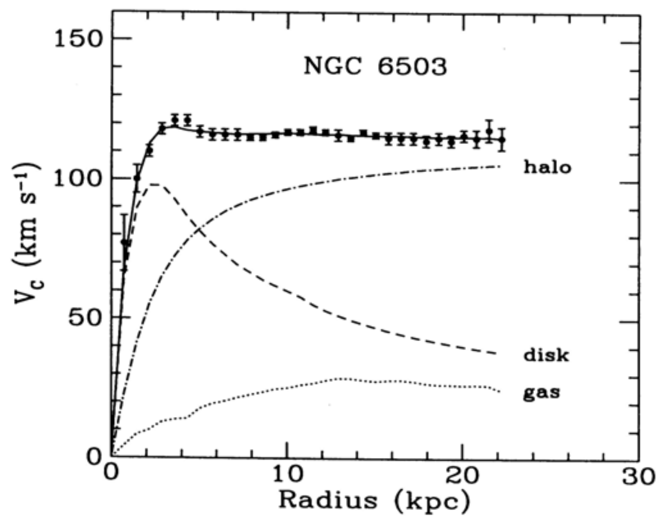


Figure 4.2 : The dependence of the radial velocity of stars as a function of their distance to the centre of the host galaxy provides one of the strongest evidence for the presence of dark matter extending well beyond the visible one.

Clusters of galaxies have long been known to require much more matter than they are seen to contain in order to be gravitationally bound. This was in fact the first indication, as pointed out by Zwicky, in the favor of dark matter. Moreover it is possible to study the gravitational lensing effect of dark matter on galaxies in the more distant background, resulting in the distortion of their images. The bending of the light rays around a mass interposed on the line of sight increases the solid angle over which the source is seen and therefore its apparent luminosity (Figure 4.3). The technique has been applied to observations at different wavelengths (radio, optical and x-ray) and gives cluster masses about 8 times larger than their baryonic masses.

The result, Ω_{DM} ~23%, is consistent with all observations and seems therefore inescapable. The question then arises to understand the nature of this non-baryonic dark matter.

Cosmologists are confident that the bulk of it is "cold", meaning that it is made of non relativistic objects. They talk of "cold dark matter", CDM. On the basis of simulations, they claim that if it had been "hot", namely made of particles moving at the speed of light or close to it, it would have smoothed out the small density fluctuations in the early universe and would have prevented the formation of galaxies. Instead large structures (super-clusters) would have formed first, followed by clusters and later by galaxies. But we are now confident that the contrary happened. While there is evidence for the formation of galactic objects (very large, elliptic galaxies, some of which are likely to include a very massive black hole in their centre) at redshifts reaching 10, typical galaxies (spiral) formed a few billion years after the big bang at redshifts between 1 and 3 as shown by the Deep Field survey of the Hubble Space Telescope, clusters formed later at redshifts of the order of unity (there is evidence for it from x-ray data and from the Sloan Digital Sky survey) and finally super-clusters are still in the process of being formed (Figure 4.4).

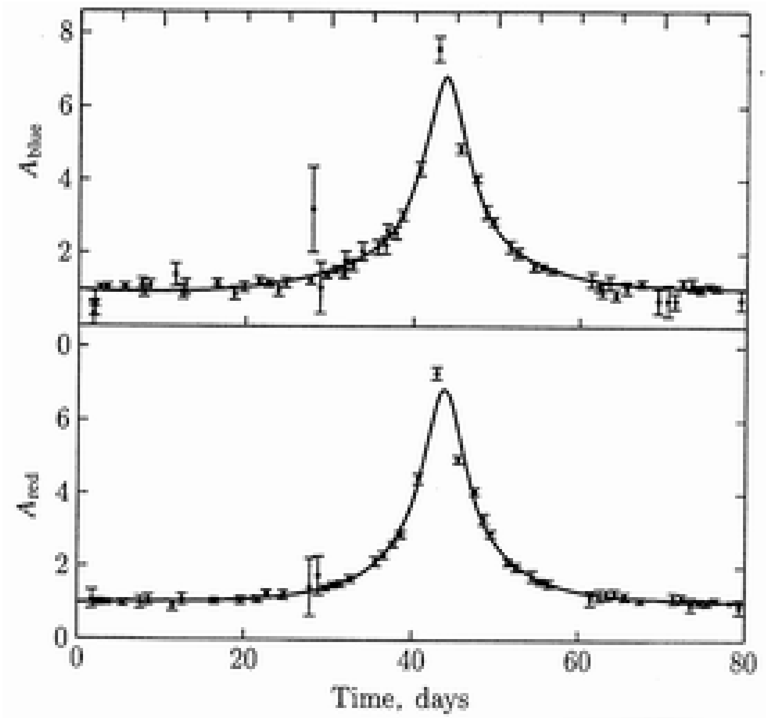
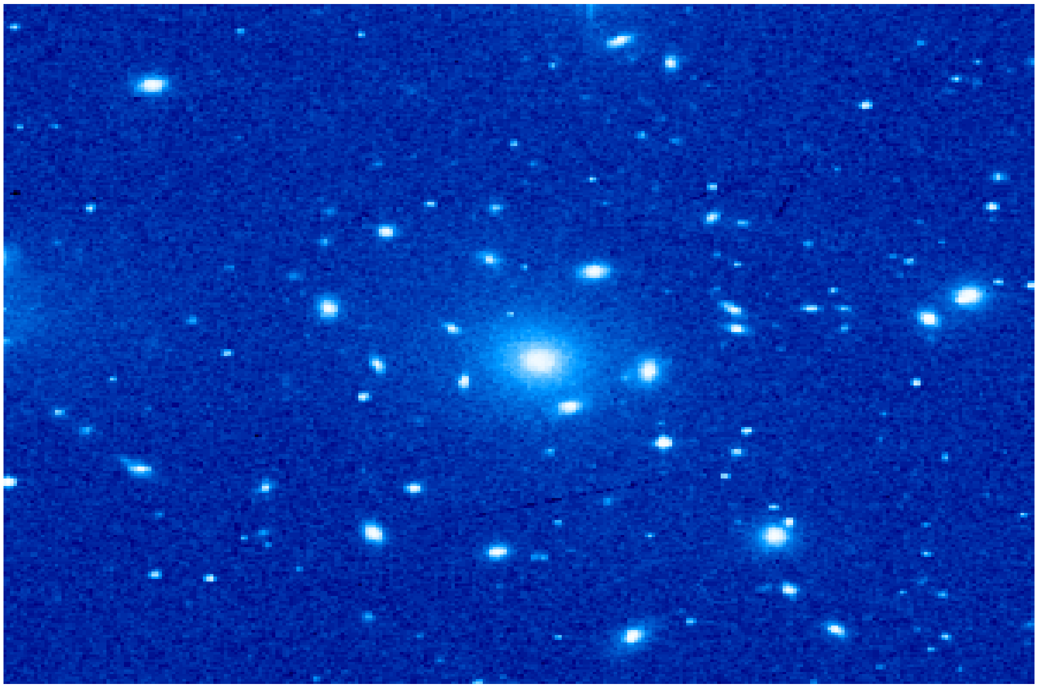


Figure 4.3 : An illustration of gravitational lensing. What is shown here is the increase of luminosity of a LMC (Large MagellanicCloud) star in the background resulting from the passage in front of it of an obscure object in the halo of our galaxy. Observation is made both in the blue (top) and in the red (bottom).

The cold dark matter is therefore expected to consist of weakly interacting slow particles. Much effort has been devoted to the search for such particles (often referred to



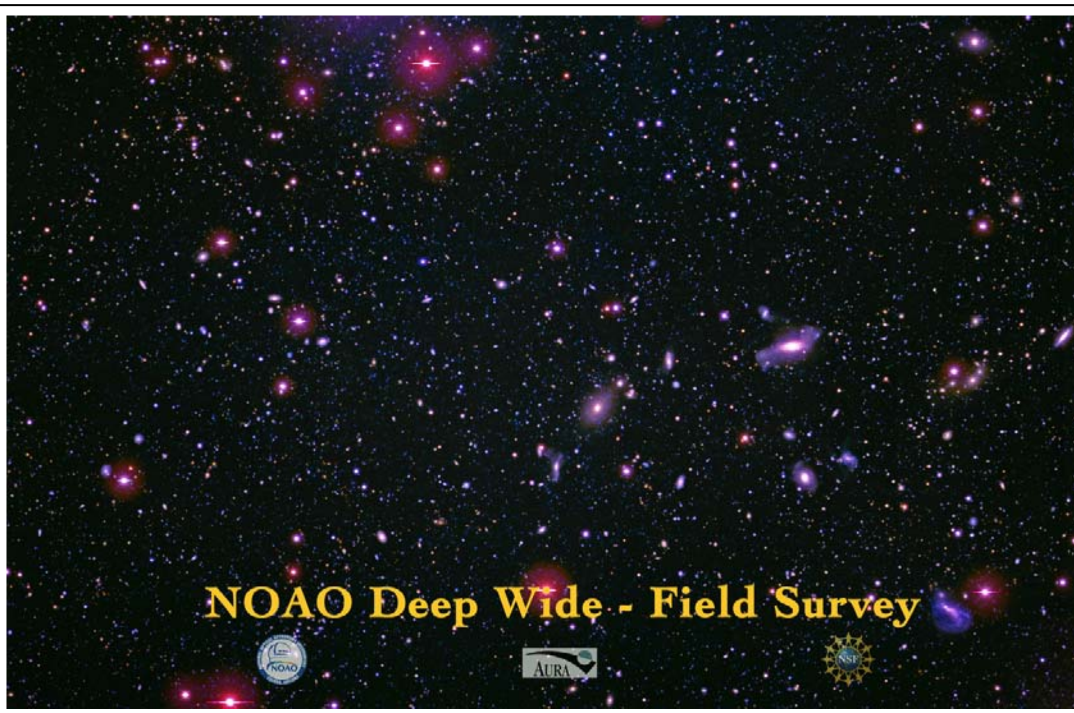


Figure 4.4 : Deep field surveys of the sky give evidence for very old galaxies, with redshifts in excess of 10.

as WIMP's, weakly interacting massive particles) with no success so far. The most popular candidates are the supersymmetric neutralino and the axion.

The former is particularly attractive as the case for supersymmetry in nature is quite strong and the lightest of the supersymmetric partners of ordinary particles should indeed be massive, stable and weakly interacting. Quantitative estimates of its abundance are consistent with what is required to account for dark matter. When the CERN accelerator LHC will start operation it will hopefully reveal such particles if they exist.

Axions are particles postulated to get around the fact that the strong force is CP conserving, which is not a necessary requirement from first principles. While axions in their most natural form (Peccei-Quinn) have been excluded by experiment, refinements of the theory have made it possible to keep the idea alive.

4.3 The cosmic microwave background (CMB)

Recent neutrino oscillation experiments have shown that neutrinos are massive. From the upper limit placed on their masses and from the density of neutrinos expected from the big bang phenomenology one can infer that neutrinos cannot contribute more than a percent or so to the energy density of the universe.

Similarly the contribution of photons from the cosmic microwave background (that dominates the electro-magnetic radiation content of the universe) accounts for a very small fraction, of the order of only 100 ppm.

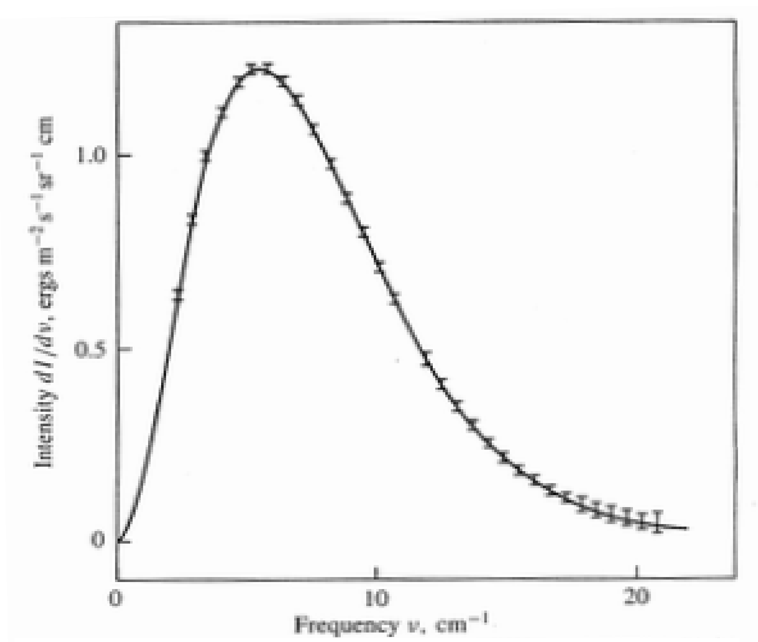


Figure 4.5 : The power spectrum of the CMB has been measured with great accuracy and is observed to be perfectly described by a black body distribution

However the very precise knowledge that we have recently gained of the CMB is a major element in our present understanding of what the universe is made of and deserves to be briefly summarized. The mapping of the sky temperature has reached unprecedented accuracy and space resolution during the recent years thanks, in particular, to the advent of subKelvin bolometers equipped with very low noise electronics. The power spectrum (Figure 4.5) was found to be very well described by a black body distribution with an average temperature of 2.725 ± 0.001 K and, most importantly, very tiny fluctuations at the 10 ppm level were accurately measured. The data (Figure 4.6) revealed an excellent global isotropy (after removal of a spurious anisotropy due to the movement of the earth in the universe). An important consequence of the Planck shape of the spectrum and of its measured average temperature is the confirmation that the universe became essentially transparent to photons at an early age (*~400 000 y, z~1000*) before which photons and matter had been in thermal equilibrium since long, and that no new source of photons existed in the universe during the dark ages until the formation of galaxies at $z<10$. The fact that the temperature of the CMB could be correctly predicted is held for a major

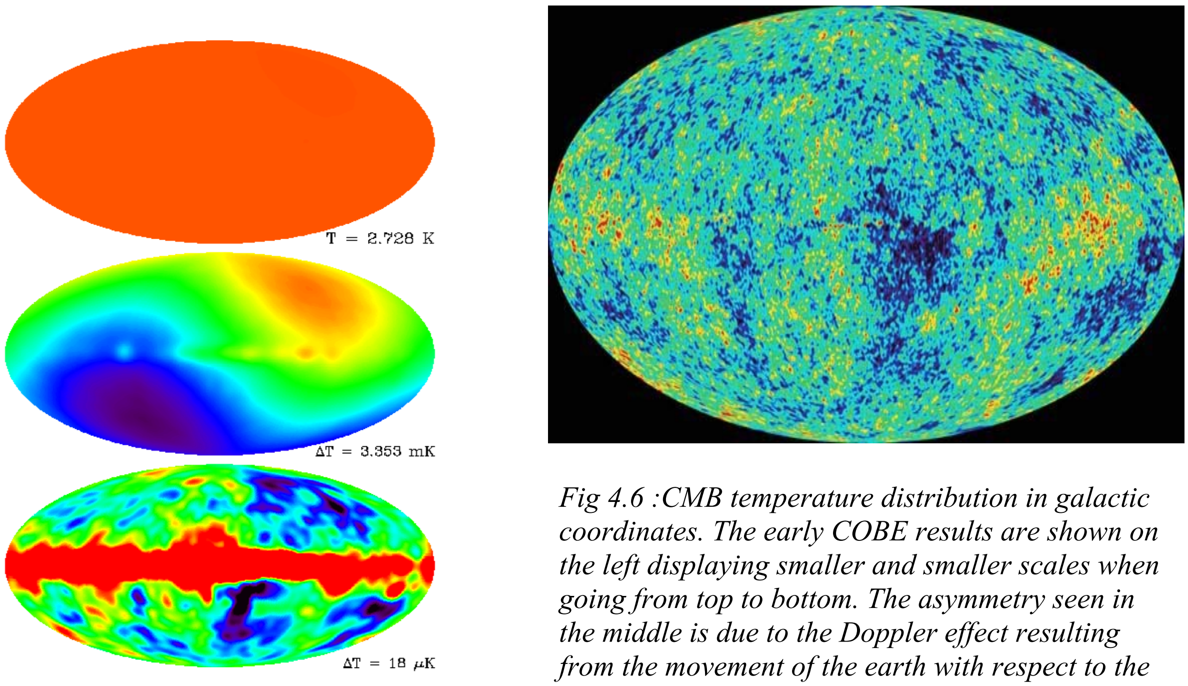


Fig 4.6 :CMB temperature distribution in galactic coordinates. The early COBE results are shown on the left displaying smaller and smaller scales when going from top to bottom. The asymmetry seen in the middle is due to the Doppler effect resulting from the movement of the earth with respect to the universe, it is subtracted on the lower scale views. The later and finer results from WMAP are shown on the right.

success of the standard big bang model (only very few assumptions are necessary to get the right answer).

The tiny anisotropies (Figure 4.6) that have been measured describe the state of the density fluctuations in the universe at the time when photons decoupled. What is observed today is merely a snapshot of the universe at recombination time, t_{rec} , all what happened to the photons after having interacted with matter for the last time ("the last scattering surface") has been to be redshifted by *~1000* units as they subsequently expanded with the universe. The general idea is that the fluctuations that we observe today have been essentially unaffected since their emission at the time of recombination. Moreover since matter and radiation were in thermal equilibrium before recombination, these observed fluctuations reproduce the matter fluctuations at the time of recombination. Finally, as this thermal equilibrium lasted for a long time, of the order of 400 000 years after the end of inflation (the quantum fluctuations having their origin in whatever scalar field generated inflation), we believe that we are able to describe accurately the evolution of the density fluctuations over this long period as long as we know what they were at the end of inflation and as we know the cosmological parameters that describe expansion during that period. The CMB snapshot can indeed be compared with the present fluctuations revealed by the existence of galaxies observed today, with a typical correlation length of the order of 1Mpc, corresponding to a correlation angle of the order of one degree.

When we measure today (time t_0) the temperature difference between two directions (Figure 4.7) separated by a small angle θ_0 we look at two sources that are spatially separated by a distance $(t_0 - t_{rec})\theta_0 \sim t_0\,\theta_0$. At recombination time they were therefore separated by $t_0\,\theta_0/1000$. The horizon was then precisely equal to t_{rec}. This implies that for angular separations larger than $1000\,t_{rec}/t_0 \sim 1.6°$ we are looking at two sources that were not causally connected at the time of emission (of course they had been causally connected earlier at the beginning of inflation but quickly moved out of the horizon as inflation proceeded). When looking for structures modulating the CMB map we would therefore expect to find essentially nothing beyond such an angular separation. The temperature fluctuations of the CMB are described in terms of the moments of their expansion in spherical harmonics, $\Sigma\, a^l_m Y^l_m(\theta,\varphi)$, where the mean values of the squares of the moments, $< a^l_m>^2 = C_l$, depend only on the spherical wave number l in the case of an isotropic distribution (we may choose the z axis as we wish). C_l is called the angular power spectrum and is related to the two-point correlation function, $\xi(\theta)$, by the relation

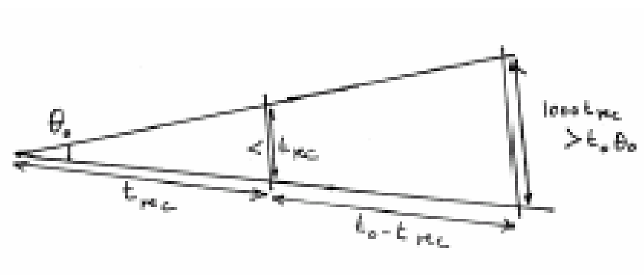


Figure 4.7 : The CMB cut-off at low l values.

$$\xi(\theta)=\Sigma(2l+1)/4\pi\; C_l\, P_l(\cos\theta).$$

Here $1+\xi(\theta)$ is simply the ratio between the density probability at an angular separation θ and what it would be in the absence of correlation. The P_l are Legendre polynomials.

The angular power spectrum (Figure 4.8) is expected to display a series of peaks called "acoustic peaks", starting at a value l_0 of l for which P_l has its last maximum (near $\cos\theta = 1$) in the middle of an aperture $\Delta\theta \sim 1000\; t_{rec}/t_0$ namely at $\theta \sim 0.8°$. The location l_0 of the first acoustic peak is therefore expected to be ~220.

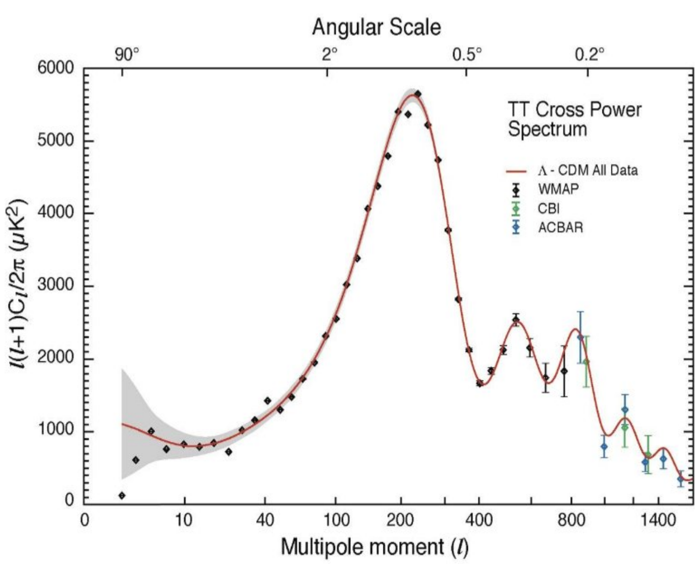


Figure 4.8 : The CMB angular power spectrum. The amplitudes of the components of its decomposition in spherical harmonics are shown as a function of the wave number l.

Any larger scale modulation, having l smaller than this value, would correlate regions that were not causally connected at the time of recombination. Here we have implicitly assumed that the universe was flat. If it were not, the modulations would evolve differently between t_{rec} and t_0 with the result that the measured value of l_0 provides a direct measurement of the curvature of the universe: $\Omega_0 \sim (l_0/220)^2$. The current result is $\Omega_0=1.03\pm0.03$, considered as the strongest evidence in favor of a flat universe.

The overall dependence of the angular power spectrum over l can be approximately described by a power spectrum of the form l^{-n} where n is predicted to be 1 by inflation. The current result, $n=1.05\pm0.09$ is considered as a success of the inflation scenario.

Other detailed features of the angular power spectrum depend on the cosmological parameters and the observed spectrum can therefore be used to place constraints on them.

4.4 Dark energy

Taking seriously the evaluation $\Omega_0=1$ of the curvature of the universe resulting from the study of the angular power spectrum of the CMB would imply that something like 73% of the energy density of the universe is not accounted for by matter and radiation, whether visible or dark, baryonic or else. This surprising conclusion has recently received support from other observations, in particular from the fact that very distant galaxies appear fainter than they should, thereby suggesting that they are farther away than we estimate, namely that the expansion of the universe is currently accelerating rather than decelerating. Indeed the "deceleration" parameter measured in this way takes the value $q_0=-0.67\pm0.25$. This deserves some explanation.

The determination of the Hubble constant implies a simultaneous measurement of the radial velocities and distances of a broad range of stars or galaxies. While radial velocities are relatively easily obtained from redshift measurements (but require corrections for the fact that the photons of a galaxy observed today tell us how it was at the time when they were emitted and not how it is today, for proper motion velocities, etc...) distances are much more difficult to evaluate. They usually rely on the comparison of the measured apparent luminosity with the absolute luminosity that we pretend to know well in some cases. Such a case is that of Cepheid's, stars that have a periodic luminosity dependence and are observed in both the Milky Way and other galaxies. For those in the Milky Way and in the nearby Magellanic Clouds the distance can be directly measured from the seasonal variation of their angular coordinates due to the parallax effect resulting from the movement of the earth around the sun. The luminosity of these stars is observed to be a well defined function of their period. The understanding of the physics causing the luminosity modulation (oscillations in the stars atmospheres inducing spectral as well as luminosity changes) makes us reasonably confident that all Cepheid's obey the same relation between luminosity and period.

Measuring their period gives therefore their absolute luminosity which, in turn, when compared to their apparent luminosity, gives their distance.

Another case in which we claim to understand the physics well enough is that of Type Ia supernovae. Type Ia supernovae are believed to be white dwarfs having a large companion from which they accrete matter. When the accretion is sufficient for their mass to exceed the Chandrasekhar limit (of the order of 1.4 solar masses, see later), they collapse into a neutron star. A gigantic explosion follows, accompanied by the violent expulsion of their superficial layers in the form of very hot matter. At that point of their evolution they can be expected to be in a standard state (the typical density of the accreting white dwarf when the mass reaches the Chandrasekhar limit having a well defined value if one neglects effects related with chemical composition) and therefore to have a well defined absolute luminosity.

As can be seen from Figure 4.9, the naive extrapolation of the Hubble diagram measured for low z galaxies does not fit well the data obtained from high z supernovae. A good fit requires a deceleration parameter of the order of –0.7 as mentioned above, namely an accelerating expansion. Recently a similar result has been obtained from large redshift gamma ray bursters.

An accelerating universe requires a negative pressure in the Friedmann "force" equation. The ratio w between that pressure and the energy density (one talks of the dark

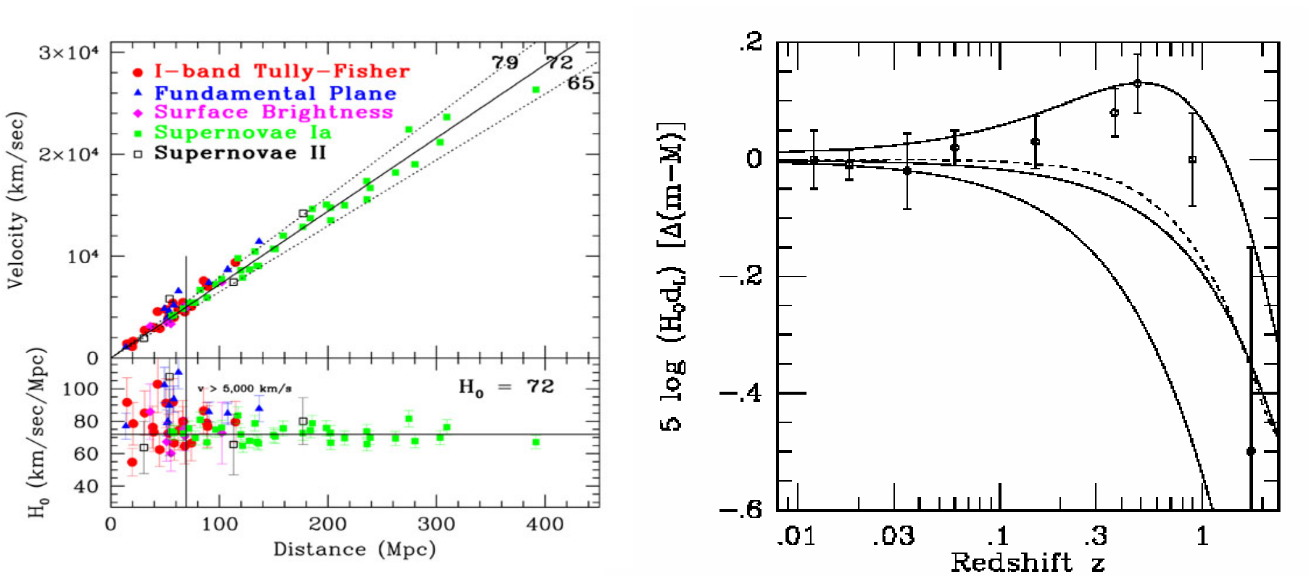


Fig 4.9:The Hubble diagram for galaxies at distances of a few hundred Mpc (mostly SNIa above 100 Mpc) is shown on the left. The figure on the right presents evidence for deviations from its extrapolation when looking at very large red shift galaxies. The dotted line would be a non-accelerating flat universe; the top line is for 30% of matter and 70% of dark energy.

energy "equation of state") is found to be -1.0 ± 0.2, the value that would obtain in the case of a cosmological constant Λ.

Such a cosmological constant would describe a repulsive force that would increase linearly with distance as Λr. It was first introduced by Einstein in order to allow for a static solution of his equations applied to a homogeneous universe, which is not possible otherwise (at that time one did not know about the expansion of the universe, it was before Hubble's work, and before Friedmann's work). Nothing is known of this so-called "dark energy". While its existence is not inconsistent with any other observation – in fact it is not only consistent but even favored by some other observations – it must be taken as nothing more than a convenient working hypothesis that may very well hide our ignorance or misconceptions. Yet, inflation and dark energy are the two new ingredients that have been recently added to the standard big bang model to constitute the current standard model of cosmology, the so-called "concordance model", that is consensually used by most cosmologists and astrophysicists on the road to a better understanding of their nature (Figure 4.10).

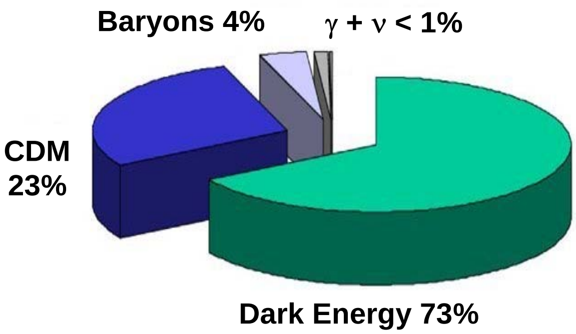


Figure 4.10 :The content of the universe according to the concordance model.

5 Birth and life of stars

5.1 General considerations

While it is generally accepted that the main mechanism of the birth of stars is the condensation under gravity of dense clouds of matter, many details are not yet understood. If there were no other force than gravitation any set of matter having a density in excess to critical would collapse. However, in the real world, as the density of the collapsing body increases, other forces, electromagnetic, strong and weak, come into play and resist the gravitational collapse by opposing to it a pressure. It is useful in this context to start with some very general considerations.

The gravitational energy contained in a cloud of dimension R, density ρ, pressure p and mass M is $\sim GM^2/R$ that must exceed pR^3 for the condensation to take place, namely $M>(p/G)^{1/2}R^2$ or, replacing $R=(M/\rho)^{1/3}$, $M>M_{Jeans}$ with the Jeans mass defined as $M_{Jeans}=(p/G)^{3/2}/\rho^2$. When the mass of the cloud is smaller than the Jeans mass, the cloud does not condense. For an equation of state $p=w\rho$ the Jeans mass becomes $(w/G)^{3/2}\rho^{-1/2}$. Equivalently we have a critical density $\rho_{Jeans}=w/GR^2$. Just before recombination the density was 10^{-21}g/cm^3 and $w=1/3$, namely $M_{Jeans}=5.10^{18}$ solar masses. After recombination the pressure had dropped by a factor 10^9 and the Jeans mass had become $10^{27/2}$ times smaller, namely 1.6 10^5 solar masses.

In practice, the pressure to be taken into account depends on the process of relevance and on temperature.

Below 0.01 solar masses there are no stars any longer, namely the temperature in the core is not sufficient to ignite the nuclear reactions that generate the inner pressure that in turn prevents the hot gas of which the star is made to collapse gravitationally. In this mass range we have brown dwarfs and planets; the limit between the two is at ~0.001 solar masses. In their cores there may be radioactive elements that generate heat from their decays, as is the case for our earth, but this has of course nothing to do with the nuclear reactions going on in the cores of stars.

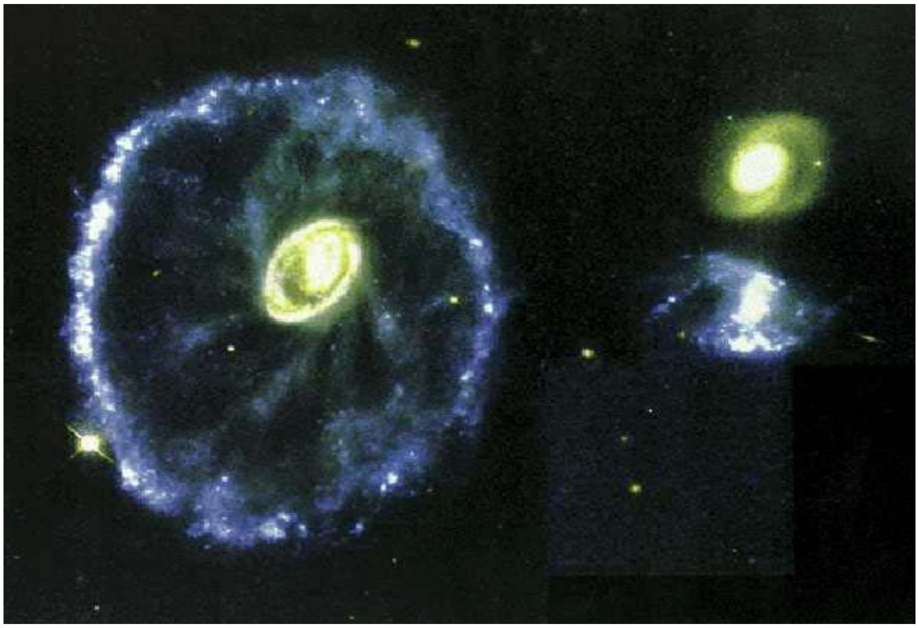
In any gravitationally condensed spherical object the density and the temperature vary as a function of radius, resulting often in phase transitions that make one distinguish between different layers. Moreover it is the seat of a complex ensemble of oscillations. The very detailed studies that have been made on our sun reveal the richness and the complexity of the physics phenomena taking place in the interior of stars.

Binary stars and single stars are observed in nearly equal abundances in the Milky Way while triple stars are exceedingly rare. This results from the fact that the formers (binaries and singles) are very stable systems at variance with the latter that evolve easily into a chaotic unstable system. This should bring our attention on the complexity of the description of the evolution of a gravitationally condensing cloud, starting from a very irregular shape to ultimately end up into a set of spheres: small inhomogeneities and anisotropies are apt at becoming dramatically amplified in the condensation process.

Above 100 solar masses or so we do not find typical stars any longer, spending most of their lifetime burning hydrogen into helium, but less stable aggregates that may be the seat of very violent events. In general the heavier a star, the shorter its lifetime: the latter is inversely proportional to the 2nd power of the star mass in the case of the more massive stars, to its 4th power in the case of the lighter ones. Indeed the gigantic and very massive clouds that condensed just after the dark ages could not form stars. Instead they formed globular clusters that became the seat of violent events with successive local collapses and collisions and into which many stars have later on be formed.

The condensation of matter under gravity in a dense cloud will be strongly helped if some external event creates local density increases. Such events may be nearby explosions of other stars that completed their normal lifetime and collapsed in a supernova explosion into a denser core. They may also be collisions between dense regions (and/or galaxies) or density waves of an external origin (as in the arms of spiral galaxies). A broad spectrum of such events has been observed and studied (Figure 5.1).

One gets this way the crude picture of first condensations taking place in large massive clouds being the seat of many collisions and of a succession of violent explosions of short lived objects, ultimately evolving to a quieter environment where stars having a long lifetime and a small collision probability can survive. Such stars will then spend most of their lifetime burning their hydrogen into helium and shortly after become red giants, leaving a white dwarf in their centre, or eventually collapse in a supernova explosion with a neutron star (possibly observed as a pulsar from the earth) or even a black hole in their centre. Supernovae explosions eject matter in space, enriched into heavy elements that have been synthesized when the star was nearing the end of its



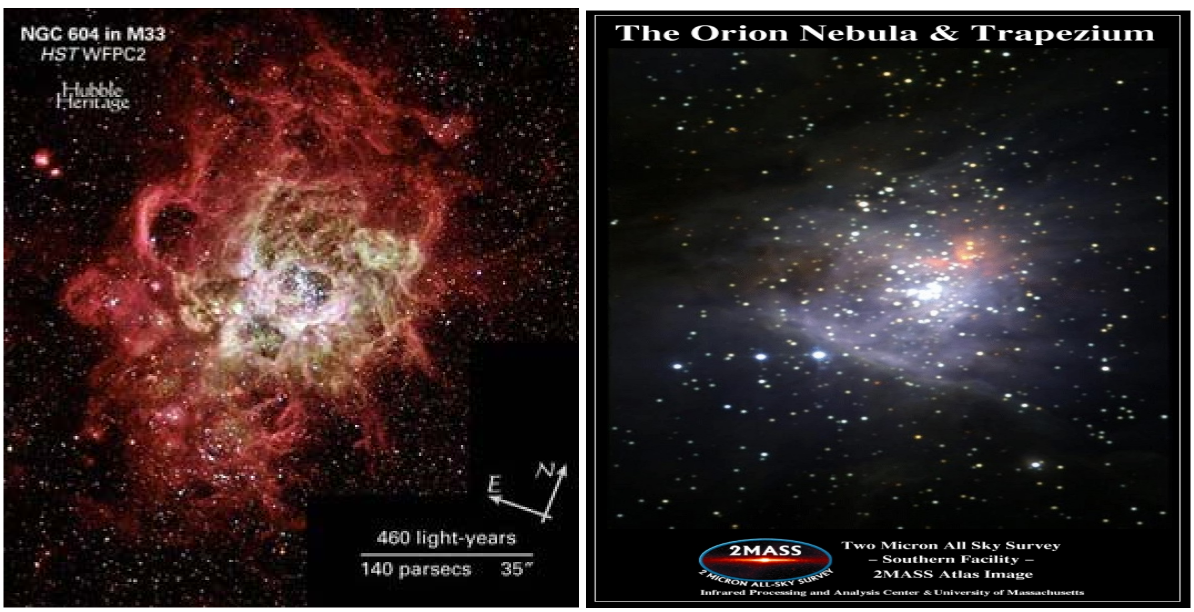


Figure 5.1 :The top photograph is of the so-called Chariot Wheel galaxy that is believed to have experienced a collision with of one of the smaller galaxies on the right.

The lower left HST picture is of one of the largest star forming regions known to us in a nearby spiral galaxy(NGC604 in the Triangulum, 2.7Mly away from us). The lower right picture is of a star forming region in our own galaxy, the Trapezium in the Orion nebula.

life or that had been inherited from former supernovae explosions that had polluted the material from which the star was formed. Such clouds of matter may ultimately merge with others and possibly condense into new stars. The dense cores, white dwarfs, neutron stars or black holes, left in the place of the dead star may also enjoy a second life when they have a companion from which they can accrete matter. In this way several generations of stars may follow each other. Stars of the new generation will usually contain more heavy elements than their parents had when they were born and witness less violent events than their parents did during their lifetime.

5.2 The HR diagram, main sequence stars

We can resolve and study stars not only in the Milky Way but also in nearby galaxies such as the Magellanic Clouds and Andromeda. The study of their spectral lines, emission and absorption, tells us about their temperature (or rather that of their outer layers) and their apparent luminosity. Astronomers often use a color index instead of temperature and magnitude instead of luminosity. The color index is the ratio of the apparent luminosities measured through two different filters, say blue and yellow, while the magnitude is simply the logarithm of the luminosity raised to the power –2.5 (up to a constant term, the absolute magnitude of the sun is ~5), a high luminosity meaning a low magnitude. The Hertzsprung-Russell (HR) diagram (Figure 5.2) shows the location of stars in the two-dimensional plane spanned by these parameters (temperature or color index and absolute luminosity or magnitude). Most stars, including the sun, are seen to populate a narrow band, high luminosities being associated with high temperatures and conversely low luminosities with low temperatures. These stars make up the main sequence (MS) and, like our sun, are mostly made of hydrogen and helium gas at the beginning of their evolution (in the OB region). Two additional populations correspond to the red giants (at high luminosities and low temperatures) and the white dwarfs (at low luminosities) respectively. In the MS there is a correlation between the temperature (or color) and the spectral type that refers to the dominant lines in the spectrum. From high to low temperatures the spectral types are called O, B, A, F, G, K and M. There is an additional

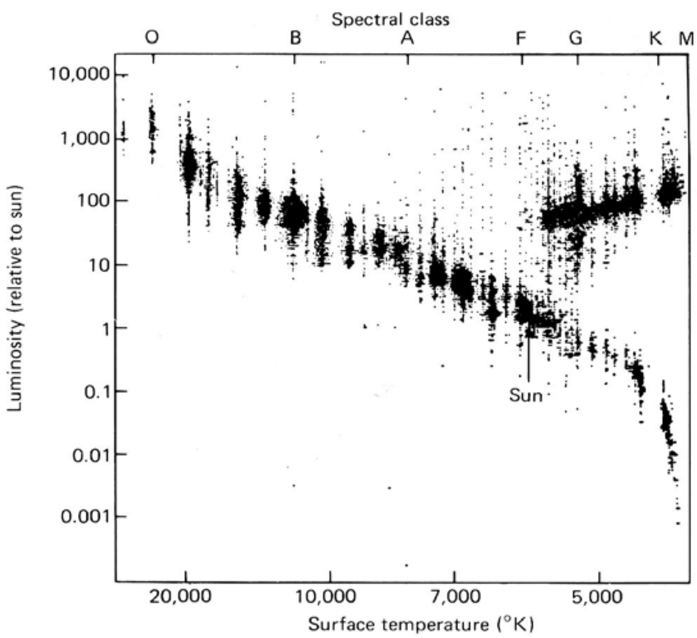


Figure 5.2 : The Hertzsprung-Russell diagram. The ordinate is the star luminosity relative to the Sun, the abscissa is the surface temperature in K. The MS and the giants are clearly seen, white dwarfs populate the region below the MS.

class called W, that of Wolf-Rayet stars, and two classes C(carbon) and S(zirconium oxide) above M. Wolf-Rayet stars are very massive hot stars with a very short lifetime.

MS star masses can be measured directly in the case of double stars and are all found to be between 0.1 and 20 solar masses. Their masses are observed to increase with luminosity. Indeed, a black body of a given temperature radiates a well defined power per unit area: one expects its luminosity to scale with the square of its radius.

Star distances are measured from their parallax, the apparent movement of the star with respect to "fixed" far away stars over a year. This method can only be applied to nearby stars, mostly from the Milky Way or the Magellanic Clouds. In the case of double stars of known masses, the measurement of the period of revolution and of the apparent diameter of the relative orbit gives another evaluation of the distance. But for more remote stars one must rely on less direct methods using stars of known absolute luminosities such as the Cepheids (calibrated using parallax). These have periodic (1 to 50 days) radial oscillations of their outer layer that induce changes in their luminosity (0.5 to 2 units of magnitude) as well as in their spectral type, period and luminosity being strongly correlated.

Star diameters can only be measured in particular cases, such as occultation by the moon, from the corresponding luminosity profile. However, in cases for which no direct method is available but for which the spectral lines show unambiguously that one is dealing with a MS star, one can use the HR diagram to get an estimate of the absolute luminosity from the known color index: assuming a black body radiator, one can then evaluate its area and therefore its diameter. MS stars have diameters that vary between 0.1 and 10 solar diameters, decreasing when moving from O to K. Red super-giants have diameters several hundred times larger than the sun and giants twenty to thirty times. White dwarfs are much smaller, with diameters similar to that of the earth and luminosities two to five orders of magnitude lower than the sun. They have masses of the order of one solar mass, implying enormous densities.

MS stars, such as the sun, are known to be the seat of nuclear reactions, the main ones being the *pp* and the *CNO* (or carbon) cycles.

The *pp* cycle is $p\ p \to d\ e\ \nu,\ d\ p \to {}^{3}He\ \gamma,\ {}^{3}He\ {}^{3}He \to {}^{4}He\ p\ p$
and the *CNO* cycle $p\ {}^{12}C \to {}^{13}N\ \gamma,\ {}^{13}N \to {}^{13}C\ e\ \nu,\ {}^{13}C\ p \to {}^{14}N\ \gamma,$
${}^{14}N\ p \to {}^{15}O\ \gamma,\ {}^{15}O \to {}^{15}N\ e\ \nu,\ {}^{15}N\ p \to {}^{12}C\ {}^{4}He\ \gamma$,
namely $4p \to {}^{4}He + 2e\nu + 2\gamma$ for *pp* and $4p \to {}^{4}He + 2e\nu + 4\gamma$ for *CNO*.

Other nuclear reactions can also take place, depending on temperature, with the production of heavier elements.

During their evolution stars free gravitational energy of which one half is radiated and the other half stored as internal energy. This is the result of the so-called virial theorem that states that the negative of the total gravitational potential energy of a star (therefore a positive quantity) is equal to twice its total internal kinetic energy (including both the thermal and collective movements).

MS stars with a mass of the order of the solar mass evolve rather slowly, spending typically 10^{10} years burning their hydrogen into helium via the *pp* cycle (more for stars lighter than the sun, less for stars heavier than the sun, a time inversely proportional to the 4th power of the star mass). More massive stars use the *CNO* cycle and the time they spend on the MS is shorter, inversely proportional to the 2nd power of the star mass.

At the end of the hydrogen burning period, MS stars enter very complex regimes, helium fusion into carbon ($3\ ^4He \to {}^{12}C$) becoming an important reaction. The more massive stars become red giants with their outer shells expanding while their core produces heavier and heavier elements, up to Fe and Ni, and eventually collapses. The lighter stars also become red giants but following a different path in the HR diagram with possible catastrophic accelerations of the helium fusion reaction ("helium flashes") and periods of instabilities yielding oscillations of the RR Lyrae type, ultimately ending as a white dwarf core and a planetary nebula envelope.

Nucleogenesis, namely the study of how nuclei are being formed in the universe, is a very active and complex branch of nuclear astrophysics. It is beyond the scope of these lecture notes to go into any detail of the various processes at play. The table below gives a brief description of the most important of them. It indicates the typical temperatures (T) and densities (δ) characteristic of each process, together with the sites in which they occur. Iron is the most stable of all nuclei and is associated with a peak in the abundance distribution of elements. In the formation of lighter nuclei, α particles play a dominant role and even-even nuclei dominate the scene. Most of the helium in the universe is of primordial origin, just after the big bang. Above iron, nuclei are formed by successive neutron captures in competition with β-decays, the so-called s-process when there is no time for multiple captures to take place because of the short decay lifetimes, or r-process, allowing for multiple sequential neutron captures, otherwise. Very light nuclei, Be, B and Li, are believed to be mostly produced from cosmic ray interactions.

Process	Description	T (K)	δ	Preferred site
H-burning	pp and CNO cycles (see text) essentially $4p \to {}^4He + \gamma\text{'s}/\nu\text{'s}$	$1–5\ 10^7$	–	Big bang MS stars
He-burning	$3\alpha \to {}^{12}C + \gamma$ ${}^AZ_N + \alpha \to {}^{A+4}Z{+}2_{N+2} + \gamma$ up to A=24 Also (α,γ) reactions starting from ^{14}N from CNO cycle up to Mg	$2\ 10^8$	$10^2–10^4$	Red giant cores, He white dwarfs
α-process C-/O-burning	${}^{12}C + {}^{12}C \to {}^{20}Ne + \alpha$, ${}^{23}Na + p$, ${}^{23}Mg + n$ ${}^{16}O + {}^{16}O \to {}^{28}Si + \alpha$, ${}^{31}P + p$, ${}^{31}S + n$	$1–2\ 10^9$	$10^5–10^6$	Massive stars, pre SN stage
Equilibrium process, Si-burning	Complex set of nuclear reactions populating up to the Fe peak	$4\ 10^9$	$3\ 10^6$	Final stage before SN explosion
Neutron captures possibly followed by β-decays	Above Fe peak, step-by-step n-captures. Short β-lifetimes: single capture, s-process Long β-lifetimes: multiple captures, r-process	$>3\ 10^8$ $10^{9–11}$	– $10^{9–11}$	Red giants, s-process SN, r-process
p-process (p,γ), (p,n), (γ,n)	Populates p-rich side of stability valley that cannot be reached by s-process	$1–2\ 10^9$	$10^2–10^4$	SN shock +H-rich outer envelope
x process	Cosmic-ray interactions producing Li, Be, B	–	–	Various

5.3 The birth of stars in spiral galaxies: OB associations and HII clouds.

It is only recently, with the advent of space astronomy and the observation of remote structures over a broad frequency spectrum, that we have started to learn about star forming regions (SFR). Too much is still unknown for a well organized presentation to be given but many recent observations have contributed to give us some idea of what is going on. It must also be mentioned that intensive and successful efforts are currently devoted to the computer simulation of gravitational condensations in different regimes, namely assuming different initial conditions.

The brightest and hottest MS stars are O and B stars. The discs of spiral galaxies (Figure 5.3), and in particular of the Milky Way, contain accumulations of such stars. They correspond to a population of stars defined by Baade as including in addition Cepheids, type II supernovae and large ISM clouds (population number I). Spiral galaxies are made of a central bulge (or hub, or bulb), a disc and a CDM rich halo. The arms correspond to density waves that excite the matter in the disc as they pass by. Baade's population number II includes stars in a more advanced state of their evolution, type I supernovae, RR Lyrae, etc… They populate elliptical galaxies as well as the hub and the halo of spiral galaxies. As far as O and B stars are concerned, they will not all follow quietly the Main Sequence. Many of these, very massive (up to 120 solar masses!), will have a very short lifetime (say 4My for a mass of 60 solar masses) and explode more or less at the same place where they were born. O and B stars are observed to cluster in "associations". But these associations are not gravitationally bound, they diverge and get diluted in 10 to 20 My, namely in a much shorter time that the lifetime of the less massive stars that will evolve on the MS, but in a longer time than the lifetime of the most massive stars that will explode when they are young. All stars in a same association do not have the same age: the explosion of the most massive stars, taking place before the dilution of the association, trigger the birth of new stars and a same association contains stars of different generations. Such OB associations extend over 100 pc or so and contain something like 100 stars. One believes today that most of the stars in our galaxy, say of the order of 70%, were born in such associations.

OB associations do not contain only stars: they are embedded in bright nebulae locally obscured by clouds and globular formations. In their neighborhood one finds very massive ($\sim10^4$ solar masses) molecular clouds (H_2) extending over some 20 to 100 pc. Around the association, surrounding the stars and pointing toward the H_2 cloud, one finds a region of very hot ($\sim10^4$ K) ionized hydrogen (H_{II}). The more and the hotter the OB stars, the larger the H_{II} cloud, some reach up to 50 pc. These H_{II} regions are the seat of very violent events, including supernovae explosions, generating density waves that trigger the condensation of the nearby H_2 cloud that gets this way progressively "eaten". One understands this way the simultaneous presence of several generations of stars in a same association.

Figure 5.3 : The top photographs are of Andromeda, a spiral galaxy in our local cluster, very similar to our own Milky Way and visible with the naked eye. That on the left is in the visible while that on the right is in the UV and shows preferentially the star forming regions.

The lower photograph is of M81, another spiral galaxy. The photograph is taken in the infrared and shows particularly well the dust rich arms of this galaxy.

The first generation is thought to have been triggered by the density wave associated with the spiral arms of the galaxy, a phenomenon that is relatively well understood and correctly modeled and simulated.

6 The death of stars, gravitational collapses

6.1 White dwarfs

When a star has completed its thermonuclear cycle, gravity dominates and the star tends to collapse under its own weight. What happens then depends strongly on the mass of the dying star. In the Milky Way, as a thumb rule, one estimates that there are $\sim 10^{11}$ stars, $\sim 10^{10}$ white dwarfs, $\sim 10^{9}$ neutron stars and $\sim 10^{8}$ black holes.

Stars having a mass of the order of a solar mass usually end up into a white dwarf. When ~12% of the hydrogen has been burned the core condenses, more hydrogen gets burned around it and more helium drifts to the core. For masses < 2 solar masses the helium core density increases and electrons in the core condense in a degenerate (or Fermi) gas. The phase transition to a degenerate gas of electrons occurs when the number of electrons per cubic meter exceeds $1.44\ 10^{10}\ T^{3/2}$ (where T is the temperature). This means that the electrons are no longer confined on atomic orbitals around the nuclei but have become an electron gas, as free electrons do in a metal. They occupy the Fermi sphere in momentum space, two electrons (spin up and spin down) in each cell, with a Fermi momentum proportional to the cubic root of the density. This configuration makes it possible for the nuclei to get much closer to each other than atoms can do, resulting in very large densities between 10^{8} and 10^{12} kg/m^{3}. Such a degenerate electron gas has a very high thermal conductivity with the result that the core is at a same temperature over its whole volume. Inside the Fermi sphere electrons are so to say "frozen", they cannot take part in interactions, only those close to the Fermi surface can (at the price of somewhat smearing the surface of the sphere). Helium fusion in the core will not take place before the core mass has reached about 0.5 solar masses. Meanwhile the outer envelope expands considerably, its external temperature decreases, the core contracts further, the luminosity of the star increases rapidly (as a result of its growing so much), it becomes a red giant. At the same time the core keeps contracting and helium fusion into carbon ignites (around 10^{8} K). The pressure of a degenerate gas is nearly temperature independent, nuclear energy can be produced without any mechanical reaction of the electrons that would regulate the process: the nuclear reaction may diverge (one talks of "flashes") up to the point where the electrons switch back to a thermal state and regulate again the system. After what, they condense again in a Fermi gas and a new phase may start (and eventually yield further flashes). The core gets enriched in C and O but the temperature is too low for them to burn into heavier nuclei. The star will burn its helium for some 10^{8} years. During this phase hydrogen keeps burning into helium in the outer layers that surround the core and instabilities, such as observed in RR Lyrae, are often observed. The hydrogen fusion into helium takes place closer and closer from the outer envelope that may get blown out into what is called a planetary nebula. Ultimately, the core will slowly cool down and become a white dwarf. Heavier elements drift to the centre while hydrogen floats on the surface, the spectra show only hydrogen lines, no heavier elements. White dwarfs have a small luminosity (10^{-2} to 10^{-5} of the sun luminosity) but are relatively hot in their early life. Masses are close to one solar mass, radii are close to the earth radius.

6.2 Neutron stars, pulsars

When the density exceeds 10^9 kg/m^3 or so, the electrons in the degenerate gas become relativistic (more correctly the Fermi velocity approaches that of light) and for a core mass of 1.4 solar masses, the so-called Chandrasekhar limit, they are no longer able to sustain the gravitational attraction. In such a case the preceding scenario does not hold any longer and the final state of the core will be a neutron star in which protons and electrons have merged into neutrons via inverse β decay. In the case of these more massive stars the collapse proceeds along different steps than in the white dwarf case. The helium core starts to fuse before electrons condense into a degenerate gas. But, as in the preceding case, the envelope grows into a red giant, while the core keeps burning heavier and heavier elements. It becomes structured in layers, the central one being ultimately made of iron and nickel. The mass of the iron-nickel core is close to the Chandrasekhar limit, its temperature and density are of the order of 10^{10} K and 10^{13} kg/m^3 respectively, its electrons form a degenerate gas. Then the core collapses very rapidly, in a fraction of a second, protons transforming into neutrons via inverse β decay and a lot of energy being emitted in the form of neutrinos (Figure 6.1). We have a supernova explosion leaving in its centre a neutron star or, if the core mass exceeds 2 to 3 solar masses, a black hole. The outer envelope is blown up in the form of a

Figure 6.1 : SN1987a exploded in February 1987. Before it exploded, it was seen as a normal star in the sky (see arrow on the left). After its explosion (right) it stayed very bright for several months.

Figure 6.2 : The remnant of the Crab nebula, a supernova that exploded in 1054. This object hosts a pulsar that shines very brightly in the X ray sky.

nebula. As an example, the Crab nebula (Figure 6.2), the remnant of the supernova explosion that occurred in 1054 and left a pulsar in its centre, consists of a filamentary distribution of mass that radiates thermally and expands with a velocity of order 1000 km/s. The present size of the gaseous envelope expelled from the original star is about 3 ly, its mass is a few solar masses. Neutron stars have diameters much smaller than those of white dwarfs, at a scale of a few kilometers. In the condensation process the angular momentum and magnetic flux of the original core are conserved with the result that white dwarfs, and even more spectacularly neutron stars, may have very large angular momenta and magnetic dipole moments (most MS stars are in rotation and have magnetic fields that may be revealed by the Zeeman splitting of their lines and reach a few Tesla in extreme cases. Our sun is particularly well studied in this respect and the recombination of field lines on its surface is known to produce large flares). When the magnetic and rotation axes do not coincide, the light-house effect that results makes such neutron stars appear as pulsars.

Pulsars were first discovered from their radio emission, with very short periods, between a millisecond and a few seconds. The magnetic fields vary between one billion Tesla in the case of the younger pulsars and ten thousand Tesla in the case of the older ones. This creates gigantic electric fields, exceeding thousand billions volts, which take their maximum values at the magnetic poles of the pulsar. Atoms are ionized over the whole surface of the star and are expelled into its atmosphere, or more correctly its magnetosphere. Electrons and ions are ejected from the magnetic poles with enormous energies and radiate photons by synchrotron emission in the magnetic field gradient. The result is two relatively narrow jets of relativistic particles (~10° aperture), emerging along the magnetic axis of the star, which are powerful radio-emitters. If the radio beams happen to cross the earth, we have a pulsar. The rotation periods are observed to slowly increase, revealing a progressive slowing down. While the core of a neutron star is essentially made of neutrons, typically at nuclear densities (in a superfluid state if the density exceeds 10^{14} g/cm^3) one may find a transition to a quark-gluon state in the very central region when the density exceeds 10^{16} g/cm^3. On the contrary, the outer surface is a crust of solid iron (and nickel). When small changes in the radial structure of the star occur, either locally or globally, the iron crust opposes some resistance until it breaks down, producing a kind of “starquake”. Whatever the exact mechanism, such events can explain the occasional sudden changes of regime observed in the rotation of young pulsars. The very large rotation velocities, say 1000 turns per second, imply that no matter can follow at radii exceeding ~50 km: its velocity would exceed that of light. This has important consequences on the dynamics of the “magnetosphere” of neutron stars. Typically, a pulsar appears in a supernova explosion with a period of 10 ms or so and slows down during a lifetime of ~10 million years to a period of the order of 1 s or so. Then the dynamo is not strong enough anymore and the pulsar switches off. However it may belong to a binary system and be able to switch on again and start a second very long life by accreting matter from its companion. This is not a rare process and such rapid (ms) “recycled” pulsars are often quite old. Their rotation is remarkably regular and can be used as a laboratory for general relativity. In particular, one hopes to be able to detect this way small irregularities resulting from a possible background of primordial gravitational waves.

6.3 Black holes.

Above 2 to 3 solar masses the neutron star is no longer able to sustain gravity: it collapses further into a black hole. Seen from the earth, the transition from a neutron star to a black hole is a relatively smooth one. Let us forget about angular momentum and magnetic field for the moment and imagine a neutron star far from any other matter and having a radius just slightly larger than its Schwarzschild radius. (In practice, however, the neutron stars and black holes that we know about and observe do have angular momentum and magnetic field; moreover they are usually accompanied by supernova remnants and are very often the seat of intense activity due to the accretion of neighboring matter). It is very unlikely for matter and radiation to escape from such a neutron star and nothing particular happens when the star radius reaches and overtakes the Schwarzschild limit. Indeed, all what happens is that anything occurring inside the Schwarzschild surface (one calls it the horizon of the black hole, not to be confused with the horizon introduced in cosmology) is confined inside it. All that we can "see", or better feel, from outside such a neutron star, or from outside the horizon of such a black hole, is its gravity. From far enough such a neutron star is simply unnoticed, exactly as a black hole is. In particular, if a new phase transition were permitted, yielding to a lower energy density, the freed energy could not escape and there would be nothing to be seen from outside (at variance with the Chandrasekhar transition where the energy evacuated by neutrinos can escape... and there is a lot to be seen). It is usually said and written that black holes are singularities and that their cores contract infinitely until their radius exactly vanishes. This is of course an abuse of mathematical language that simply hides our ignorance of the physics of relevance close to the Planck mass. Singularities and infinities are mathematical concepts that are not good friends of the physicists. A black hole has a well defined mass, a well defined charge, a well defined angular momentum and its horizon has a well defined size. Beyond the horizon the star contracts surely to much smaller dimensions than the horizon but this is not a good enough reason to claim that the core of a black hole is simply a mathematical point. To the physicist who wishes to clarify the mysteries of the current incompatibility between quantum theory and general relativity, in particular to the superstring theorist, black holes are obviously a privileged laboratory (even if very secret, how could one tell what is going on inside?). But to the student of modern astrophysics who has more modest ambitions, as is supposed to be the case of the reader of the present notes, it is sufficient to think of a black hole as something that can "swallow" whatever falls beyond its horizon (the horizon and the mass increasing accordingly) but that looks simply from outside as a very compact source of gravity.

There exist however major differences between black holes and neutron stars: the most important difference, by far, is that the mass of the former is unlimited, while that of the latter cannot increase beyond 2 to 3 solar masses (above this limit it becomes a black hole). Indeed black holes span a gigantic mass range, from a few to billions of solar masses. One usually distinguishes between stellar black holes and galactic black holes. The former have been produced typically in a supernova explosion as the end product of the life of a massive star. There are many of these in the Milky Way and similar galaxies. They may be accreting matter from the surrounding environment but even then their masses remain at the stellar scale, a few hundred or at most thousand solar masses. The latter are usually located at the centre of a galaxy and their masses may reach billions of solar masses. The most massive of these trigger an intense activity in the host galaxy that becomes the seat of extremely violent events. Another important difference between a black hole and a neutron star is the inability of the former to contain magnetic field lines as a neutron star does, the field is expelled from the horizon a bit like in the Meissner effect of superconductivity. However, in practical cases, accreting black holes spin rapidly and trap gigantic magnetic fields in their accreting disks. But, contrary to pulsars, their magnetic dipole moment is always aligned along their axis of rotation.

A common and most important feature of compact objects such as white dwarfs, neutron stars and black holes is that they can accrete material from a companion or from the environment. Such events are not rare and often result in particularly violent and spectacular phenomena. This is the subject of the next section.

7 Violent events, accretion

7.1 Active galactic nuclei (AGN)

We already mentioned white dwarfs accreting matter from a companion and exploding when they reach the Chandrasekhar limit (type Ia supernovae). It may also happen that while one of the members of a binary system grows up into a red giant its outer envelope gets accreted by the less massive star, leaving a helium rich white dwarf in the place of the red giant and producing thermonuclear reactions on the surface of the smaller star (one then speaks of a nova). Similarly many examples are known of neutron stars accreting material from a companion. In such cases the accreted matter is ionized and channeled by the magnetic field

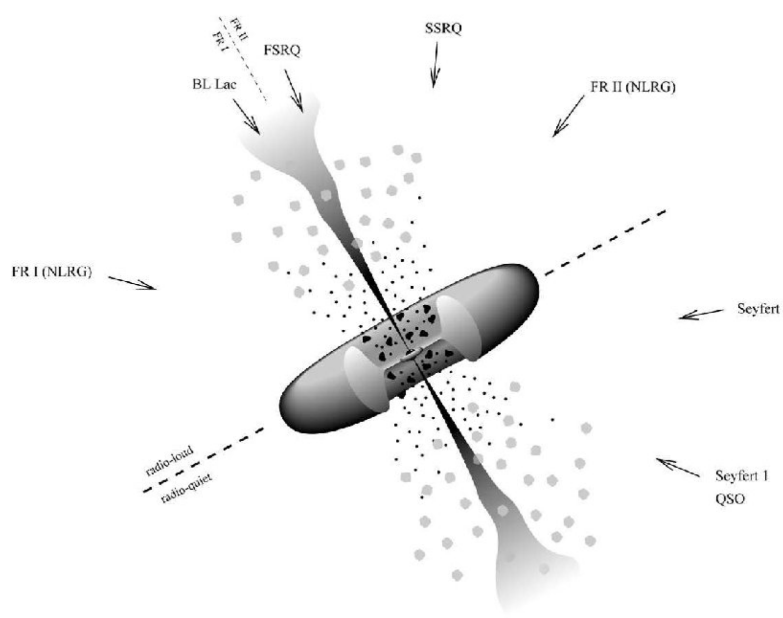


Figure 7.1 : An artist view of an accreting black hole showing the accretion disk, the torus of dust and the two jets ending as radio lobes.

of the neutron star, making its magnetic poles become x-ray emitters. It is itself heated up to high temperatures and becomes a strong thermal source of x-rays. Depending on their orientation, such binaries may show eclipses. Generally, depending on the orientation of the system with respect to the earth, on the strength of the magnetic field and on the angle between the axis of rotation and that of the field, different configurations may be observed. Many x-ray sources in the Milky Way are believed to be of this nature.

Accretion from a black hole can become very spectacular in the case of very massive black holes surrounded by a dense medium. One believes today that many of the most spectacular events in the visible universe are indeed the results of such events. The general picture (Figure 7.1) is that of a black hole in the centre of a galaxy (possibly spiral but preferably elliptical in the case of the most active nuclei).

The black hole is very massive, say 10 to 100 million solar masses (typical galaxies have masses of the order of 10^{11} solar masses). The gas and the stars in its neighborhood "fall" on it. The gas condenses in a disk around the black hole and ionizes, spinning faster and faster as it comes closer to it, exactly as predicted and very well reproduced by simulations. While speeding up the gas gets hot and radiates at x-ray frequencies. The (inward) flux of accreted matter is limited by the (outward) resistance offered by the photons radiated in the process. This kind of self-regulation implies an upper limit for the luminosity L of an accreting black hole of a given mass M. Its value, $L(erg/s)\sim 10^{38}M(solar\ masses)$, is called the Eddington limit. From the observed luminosity of the brightest quasars one infers this way enormous mass values for the central accreting black hole, up to 10 billion solar masses! Around the accreting disk one finds a torus of dust that is opaque to visible light. The strong magnetic field produced by the rotating plasma creates two jets propelling electrons and ions at very high energies along the disk axis (while no magnetic field may penetrate beyond the horizon, the field lines are strongly anchored in the rotating plasma that is itself locked, near the horizon, to the angular momentum of the black hole). Synchrotron radiation in the jets gives a strong photon emission over a very broad frequency range. At variance with the pulsar case, the magnetic field and rotation axes are exactly the same. Moreover, the rotating black hole behaves as a gyroscope and the direction of the jets remains absolutely invariant as testified by their absence of curvature (despite their very long extensions). Such systems are generically called Active Galactic Nuclei (AGN) and cover many different observed events that are briefly reviewed in the next section.

7.2 Quasars, Seyfert galaxies, BL Lac and blazars

i) Quasars. Quasars, standing for quasi-stellar radio sources, were discovered as being intense radio-sources (the radio-lobes at the ends of the jets) coinciding with very distant (high z), compact (no apparent diameter) and bright star-like optical objects (the accretion disc). Rapid (on the scale of several days) and erratic variations of the optical luminosities are observed, confirming that the star-like object is very compact indeed.

Quasars are observed to emit over a broad range of photon frequencies. In most cases two radio lobes can be resolved on either side of the nucleus at typical distances of 100 kpc, sometimes linked to it by a narrow radio jet. Radio galaxies (Figure 7.2) are sometimes classified into two different families, FRI and FRII, where FR stands for Fanaroff and Riley who have proposed the classification.

The two families are also called "edge-darkened" and "edge-brightened" respectively, the criterion for deciding to which family a given radio galaxy belongs being to observe whether the radio emission decreases or increases from the central nucleus to the edge of the lobes. The difference between FRI and FRII is mostly the effect of a different angle between the jets and the line of sight, a smaller angle corresponding to FRI. Many elliptical galaxies have been observed to behave similarly to quasars while not being such strong radio emitters as quasars are.

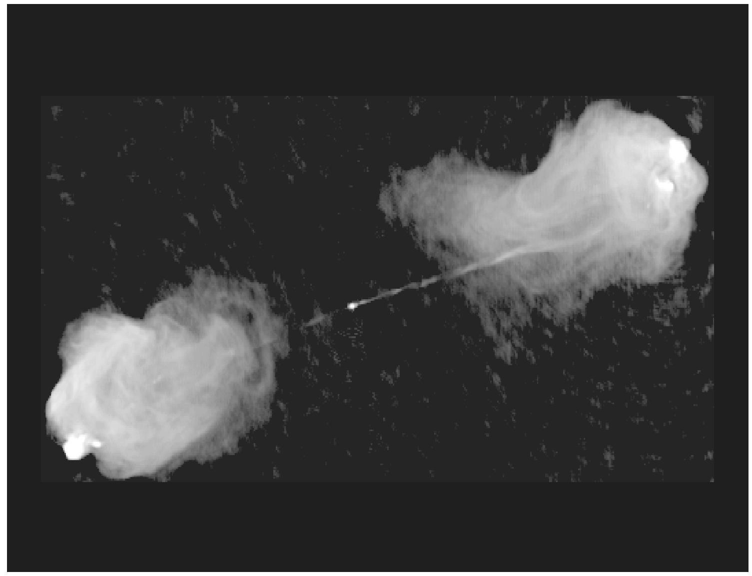


Figure 7.2 : A radio (VLA) image of the radiogalaxy Cygnus A, showing the accretion disk (the bright spot in the middle), the jets and the radio lobes.

Going backward in time, namely to larger and larger redshifts, one seems to observe relatively more and more quasars and less and less stars. This has obvious implications on our understanding of the evolution of the universe. It is currently premature to claim that each galaxy has a black hole in its centre, but if it were true the question of the origin of such black holes would become particularly important. Are black holes formed naturally at the centre of galaxies (apparently a reasonable and natural assumption, stars migrate preferentially to the center of the host galaxy as the result of their occasional collisions) or were there primordial black holes?

ii) Seyfert galaxies. These are galaxies having a bright and compact (variable) nucleus that are understood today as being quasars having a nucleus luminosity not quite as strong as quasars have: the accretion disc does not hide (blind) the surrounding galaxy (one speaks of type 1 Seyfert galaxies). Galaxies having similar spectral features but showing no central nucleus (type 2 Seyfert galaxies) are of the same kind but the nucleus is now hidden by the dust torus (as evidenced in the infrared). Depending whether a Seyfert galaxy is seen from a region close to its axis or close to its equatorial plane, it will appear as a type 1 or type 2 respectively. The observation of such galaxies may suggest that many galaxies may have a massive black hole in their centre, however not active enough to even be observed. Indeed, the study of the centers of the Milky Way and of Andromeda has unambiguously revealed the presence of black holes.

iii) BL Lacertae (or Lacertides, or BL Lac), blazars. First discovered as a variable “star”, BL Lacertae was later found to be a compact galactic nucleus associated with an intense radio-source. A remarkable feature is that the radio emission comes from two regions that are observed to recess from each other at very high velocities. Such configurations correspond to cases where one of the jets is approximately pointing toward the earth and the other away from it. For such AGN configurations one sometimes observes the emission of very high energy gamma rays: one then talks of blazars; they exhibit rapid flux variability in essentially all measured frequency bands, from radio to gamma rays.

7.3 Gamma ray bursts (GRB)

Gamma ray bursts are transient phenomena with a very short time structure (from milliseconds to a minute or two) and very high intensities. The burst itself is usually observed from space at a typical rate of one a day in an x-ray to γ-ray energy range between 10 keV and 10 MeV. In a very few cases the burst has also been observed in the optical range. The burst duration distribution shows two populations, about one third lasting 0.01 to 1 second (0.1 on average), and the remaining two thirds lasting 1 to 100 seconds (10 on average). The time profiles of the bursts show a great diversity, sometimes a single isolated pulse, but more often an irregular and apparently anarchic sequence of pulses having in general a rise time shorter than their fall time. The average photon energy is higher for short bursts than for long bursts and decreases with time during the burst. During a number of days following the burst, say up to 100 days or so, a counterpart may be detected at the location of the burst at various frequencies, x-ray, optical and radio. One refers to it as the “afterglow” or, in the optical case, “optical transient” (OT). Typically, the intensity of the afterglow decreases with time as an inverse power law, $t^{-\frac{1}{2} \text{ to } -2}$. The redshifts measured in such afterglows range typically between 0.2 and 5, demonstrating the extragalactic origin of GRB’s. This is also in agreement with their isotropic distribution in the sky. An amazing consequence is that the energy liberated during a burst must be as high as 10^{44} Joules, comparable to that liberated in a supernova explosion, but concentrated within a narrow band of frequencies. The host galaxies are usually faint galaxies with possible indications for a high star formation rate (SFR), and an excess of massive stars. The burst seems to be located within typically 10 kpc from the center of the host galaxy. A few GRB’s have been found to be associated with supernovae.

The current ideas about GRB’s imply a very violent event at cosmological distances, occurring in a small volume of space and producing ejecta moving at relativistic velocities. Compactness is implied by the short time structure of the bursts, relativistic speeds are necessary to boost the photons to sufficient energies (otherwise they would have cooled too much when the ejecta have become sufficiently transparent for them to escape). The energy will then evolve from a radiation dominated to a matter dominated phase while the system cools down.

A detailed model is therefore expected to explain three features: i) What is the initial violent event, sometimes called the “central engine” of the GRB? ii) What is the mechanism that allows for the radiation of energy (the burst) when the ejecta have moved away from the source sufficiently to become transparent? iii) What is the mechanism producing the afterglow when the ejecta have moved away even further from the source and are being significantly decelerated?

The first of these three features is still today the most mysterious. One class of models considers “mergers” as the most likely candidates. These are collapses of compact binary systems such as two neutron stars. Their rate of occurrence seems however too low to reproduce the GRB rate of about one per thousand supernovae and they would be expected to occur preferentially far away from the galactic center. Another class of models considers the collapse of a very massive star into a black hole, one speaks of “hypernovae” or “collapsars”.

The second feature is the prompt emission, the burst itself; it is often described in terms of the “internal shock model”. The idea is that the central engine produces ejecta with a spectrum of Lorentz γ-factors succeeding each other. The time scale is proportional to the mass of the central black hole, M_0. It is of the order of $6GM_0$, namely 30μs per solar mass. For some time the faster ejecta will overtake the slower ejecta until they get ordered, the faster first, the slower last. This sequence of overtaking events generates a series of shocks, the “internal shocks”, which leave hot radiating matter behind them. Radiation may come from synchrotron radiation of relativistic electrons or/and from inverse Compton scattering on UV photons of the same electrons.

The third feature, the afterglow, is usually described in terms of an external shock produced by the encounter of the mass of matter swept away by the ejecta (that are still relativistic but are being strongly decelerated) with the external medium. By “external shock” one means in fact a forward shock responsible for the afterglow, propagating in the external medium, and a reverse shock propagating back toward the source. Synchrotron radiation is the dominant radiation mechanism. Most likely the GRB energy is not isotropically emitted, but concentrated in jets which favor geometric beaming.

8. Cosmic rays

8.1 General features

The existence of cosmic rays has been known for nearly a century. Yet, we still do not understand very well where they come from and how they have been accelerated. The most surprising feature of cosmic rays is that there are so few of them and yet they carry so much energy: they reach up to 10^{20}eV, namely 16 Joules! But their flux at such high energies is only of the order of one per km^2 and per century. This implies detectors covering gigantic areas and using the atmosphere as revelator: when cosmic rays penetrate the atmosphere they produce highly collimated showers which may contain several billions of secondary particles. The light that results, either fluorescence of the nitrogen molecules or Cherenkov radiation in the air, can be used as a measure of their energies. Another method consists in sampling the shower on ground using an array of detectors, scintillators or water Cherenkov counters.

For most of the energy range cosmic rays are known to consist of ionized atomic nuclei, mostly protons. Their energy density in the universe is of the order of magnitude of the electronvolt per cubic centimeter, similar to those of magnetic fields, of visible light and of CMB photons. Their flux on the earth spans 32 orders of magnitude and decreases with energy as a power law $\sim E^{-2.7}$ on average (Figure 8.1).

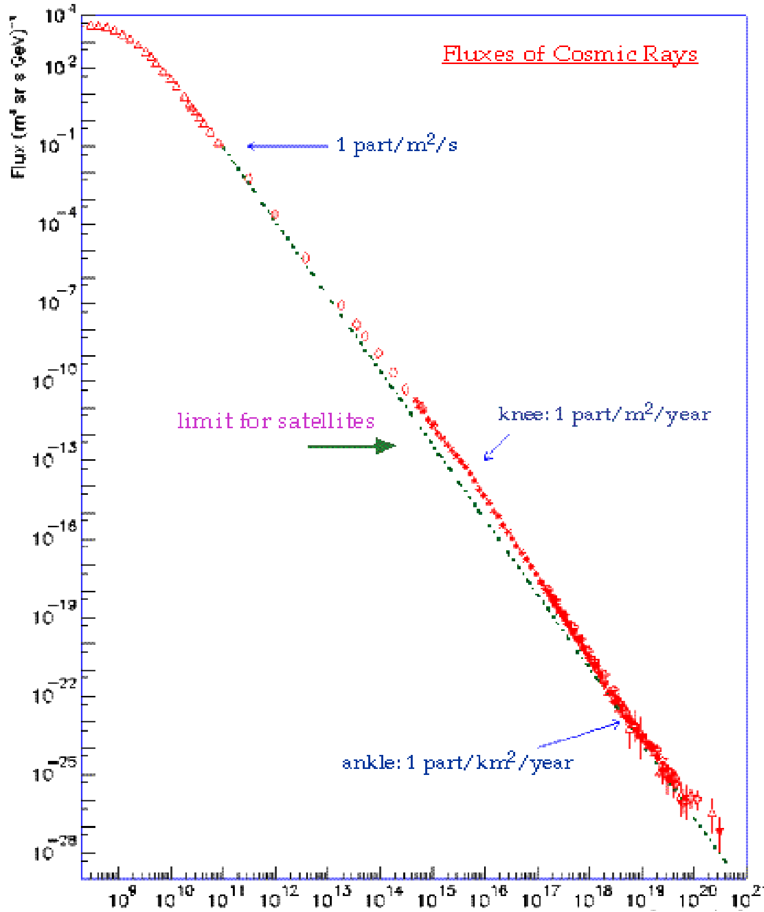


Figure 8.1 : The energy (eV) distribution of the primary cosmic ray flux reaching the earth

This law is in fact the convolution of the production at the source with the effect of propagation in the interstellar medium. Taking the latter in due account (it takes 10 My to a typical cosmic ray to reach the earth from its source) one finds a spectral index between 2 and 2.5 rather than 2.7. In fact small variations are observed; in particular two breaks, above 10^{15} and 10^{18} eV respectively, referred to as the "knee" and the "ankle", have not yet received satisfactory explanations. The abundance of elements present in cosmic rays follows closely that of elements in the universe (Figure 8.2), except for the rarer elements which are present in cosmic rays with a higher abundance. This is explained by the spallation reactions which occur when the cosmic rays interact with interstellar matter. On average they traverse some 7g/cm^2 of matter before reaching the earth.

At low energies they are strongly influenced by the magnetic fields of the earth and of the solar system in general, that have the effect of shielding the earth: one speaks of a rigidity cutoff. Similarly, most cosmic rays are trapped into the Milky Way by the magnetic field (~1µG) that pervades it. It is only above 10^{15}eV that one believes that cosmic rays are of mostly extragalactic origin.

The solar wind is known to be a source of low energy cosmic rays modulated by the 22 year solar cycle. Galactic cosmic rays are believed to be produced dominantly in supernova remnants. The galactic magnetic field prevents the identification of precise sources in the sky up to very high energies. But gamma rays are good tracers of high energy cosmic rays: when very high energy nuclei interact with matter present on their path they produce neutral pions (among other mesons) that decay very promptly into a pair of photons. Several sources have been identified this way; they generally consist of a supernova remnant (SNR). In the case of young SNR's they may be energized by a pulsar acting as the central engine; such systems are called "plerions".

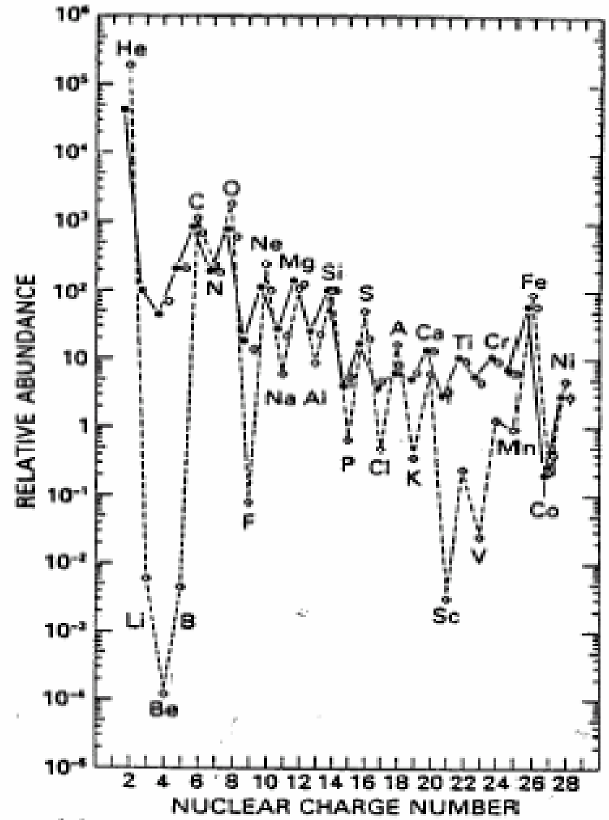


Figure 8.2 : The abundance of elements in cosmic rays (dotted line) and in the solar system (full line). The filling up of the valleys is the result of spallation reactions of the cosmic rays on the ISM

Extragalactic cosmic rays are more difficult to understand. Because of the extraordinary energy requirement, only violent events, such as AGN's or GRB's, are plausible candidates. Indeed several AGN's (blazars) have been observed to be sources of very high energy gammas. As far as GRB's are concerned one would expect the cosmic ray burst to be delayed with respect to the gamma ray burst by such a long time (several years) that it is hopeless to detect any correlation: over such large distances the intergalactic magnetic fields should significantly lengthen the path of cosmic rays. The most popular models extend the phenomena that are believed to occur in SNR's to the cases of AGN's and GRB's. However, above an energy of the order of $0.5\ 10^{20}$ eV, the universe is expected to become opaque to cosmic rays over an attenuation length of the order of 30 Mpc; such energies are above the threshold for photoproduction of pions on the CMB photons, called the Greisen Zatsepin Kuzmin (GZK) threshold. It is a matter of current controversy whether cosmic rays have been or have not been observed beyond this threshold.

This very brief overview has already suggested that the interstellar medium (matter and fields) plays an essential role in both the acceleration and the propagation of cosmic rays: an introduction to the subject is given in the next section; some words about the acceleration mechanism and very high energy gamma rays follow.

8.2 Inter stellar medium(ISM)

We restrict the discussion to a brief presentation of the interstellar medium of the Milky Way, which we know best. Matter is continuously recycled from stars to ISM (via supernovae explosions or simply by strong solar winds) and back from ISM to stars by star formation. In the star phase it gets enriched in heavy elements. ISM accounts for 10% to 15% of the total mass of the galactic disk, where it concentrates, particularly along the spiral arms. Molecular hydrogen is concentrated around a ring at a radius of 5 kpc or so; it is in the region of this ring that the supernova explosion rate is maximal. Half of the ISM is condensed in clouds that fill up only 1% to 2% of the volume populated by ISM. There are dark molecular clouds (usually gravitationally bound) at ~10 to 20 K and diffuse transparent atomic clouds at ~ 100 K. The rest of the ISM spreads between the clouds and is shared between warm atomic, warm ionized and hot ionized as summarized in the table below:

Component	T (K)	Protons/cm^3	Probe
Molecular	10-20	10^2-10^6	CO 2.6mm emission
Cold atomic	50-100	20-50	21cm line of H
Warm atomic	6000-10000	0.2-0.5	21cm line of H
Warm ionized	~8000	0.2-0.5	6563Å H_α
Hot ionized	~10^6	A few 10^{-3}	UV and soft X

The abundance of elements in ISM is similar to that in the universe.

OB associations act on ISM by dissociating molecules at the surface of molecular clouds, creating compact H_{II} regions in their neighborhood by ionization and by ionizing the nearby ISM and heating it up to ~8000 K.

Stellar winds are present in all stars but particularly important in the early and late parts of their lives. More massive stars have stronger winds, in particular stars having a mass in excess of 30 solar masses or so and ending up in a Wolf-Rayet phase with very powerful winds. Winds and supernova explosions occurring in a diluted region end up in remnants that ultimately appear as interstellar clouds. On the contrary, winds and supernova explosions occurring in a dense region, such as an OB association, keep triggering new explosions and build up a so-called “super-bubble”. Typically, in a local ISM, the remnant of an isolated SN grows up to a radius of 50 pc for about 1.5 My. While an average super-bubble produced by 30 SN explosions grows up to 300 pc in some 15 My. Massive remnants will be gravitationally bound and become a possible site for new star formation.

An essential parameter of ISM is the magnetic field that pervades it, typically at the microgauss scale. SN explosions generate perturbations that amplify existing ISM magnetic fields and SN remnants are likely to retain a part of the original stellar field. The ISM magnetic field is revealed by stellar polarimetry (magnetic dust grains align with the field and preferentially attenuate one polarization of the light passing through them), by Zeeman splitting, by Faraday rotation (in ionized ISM) and by synchrotron radiation of fast electrons. Like in the sun, magnetic field lines wind up at different rates for different galactic radii and turbulences play a major role. But, at variance with the spherical sun one deals here with a flat disk and instead of the oscillatory regime of the sun (22 y period) one obtains a slow exponential dependence on time. As much of the ISM is ionized, the magnetic field plays a major role in the details of the dynamics. In regions of steep field gradient, magnetic reconnection may occur when field lines of opposite polarities approach each other (again like what happens in the sun). Such reconnections heat up and ionize the surrounding ISM.

8.3 Diffusive shock acceleration

When discussing GRB's we already got some idea about the possible importance of shock phenomena. The evidence presently accumulated (Figure 8.3) in favor of supernova remnants (SNR) acting as sources of galactic cosmic rays has made the astrophysics community confident that significant acceleration is indeed taking place at the shock and is responsible for both the radio-emission of SNR (synchrotron radiation) and the acceleration of cosmic rays.

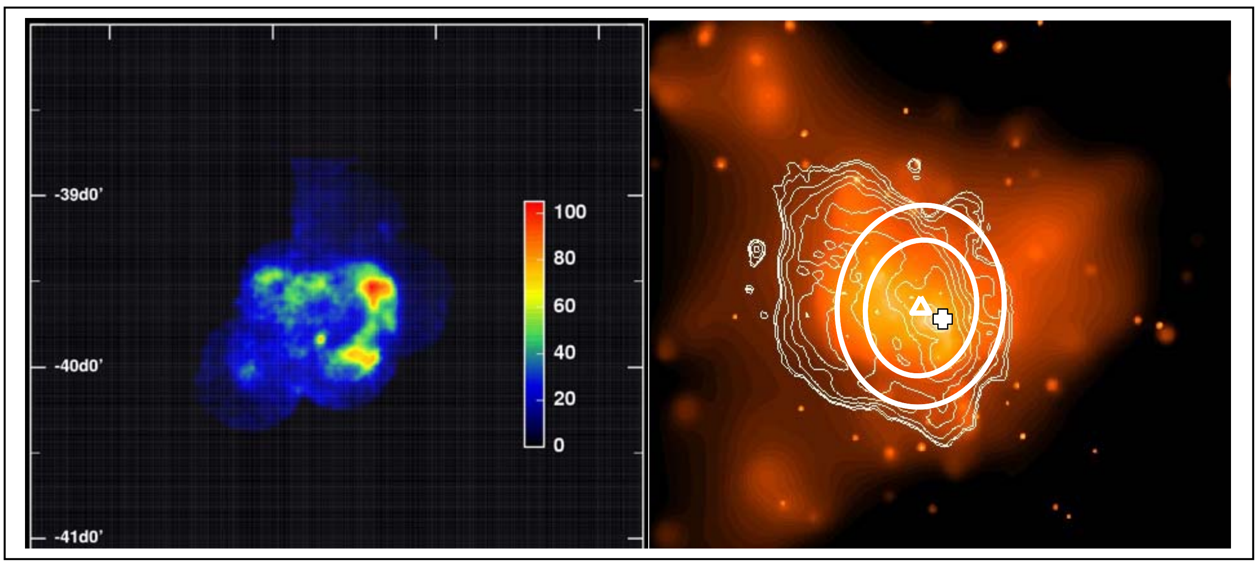


Figure 8.3 : A SN remnant (right) and its gamma counterpart (left).

The electromagnetic energy spectrum of SNR's has been studied at all wave lengths. In the radio range the frequency spectrum is typical of synchrotron emission by fast electrons in the region of the shock (at variance with plerions where emission takes place near the pulsar). This is the best evidence in favor of acceleration in the shock. The synchrotron spectrum (frequency ν) is proportional to the magnetic field B and to the square of the parent electron energy, E: $dN/d\nu \propto BE^2$, with GHz frequencies corresponding to GeV energies for realistic ISM μG fields. Most of the infrared emission is due to dust heated up by thermal electrons and most of the optical emission is due to secondary radiative shocks in interstellar clouds. While most of the x-ray emission is usually thermal, the synchrotron high energy tail is sometimes visible in this range. The same electrons that have enough energy to radiate x-rays can also produce γ-rays by inverse Compton scattering. In such cases stronger magnetic fields are believed to occur because of the amplification effect of turbulences.

The most popular mechanism invoked to describe the acceleration process is called diffusive shock acceleration (Figure 8.4). It is so generally accepted as a valid theory that a brief presentation must be given in the present lectures. Some features of it are understood, others are not. Consider a cloud, say a supernova remnant, expanding and moving (fast) into the surrounding ISM. The density of the cloud is supposed to have become so low that the particles mean free paths are large enough for collisions to be always safely ignored. Moreover we assume that both the cloud and the surrounding ISM are fully ionized, say H_{II} to make it simple, namely electrons and protons. The only relevant interaction between the particles is via magnetic fields. Both the ISM and the cloud have a history from which they have probably inherited some local collective motions sufficient for having trapped some fields. While expanding, the fields have decreased in amplitude to a microgauss scale or so. Original hydrodynamical turbulences have left their imprint in the form of field inhomogeneities. In the process of dilution some field line reconnection has taken place. We may therefore expect a very complex and erratic magnetic field structure in both the cloud and the ISM. It is essentially not understood and makes up the least convincing part of the story. We therefore proceed with the cleaner part first: it will clarify which are the properties that the magnetic field structures need to have for the acceleration mechanism to function properly. Only then will we return to the question. We next assume that nearly relativistic particles, say protons (or electrons), having energies very much higher than the respective energies of the thermal protons (or electrons) of the ISM and cloud (otherwise how could we tell them apart?), are present in the region. We call them "cosmic particles" to distinguish them from the thermal particles. They are the particles that will be accelerated further in the process. We assume that the interface between the cloud and the quiet ISM, called the "shock", is sufficiently thin to guarantee that cosmic protons entering it at not too large an angle of incidence will traverse it. We also assume that the magnetic field structures in the ISM and in the cloud are such that a cosmic

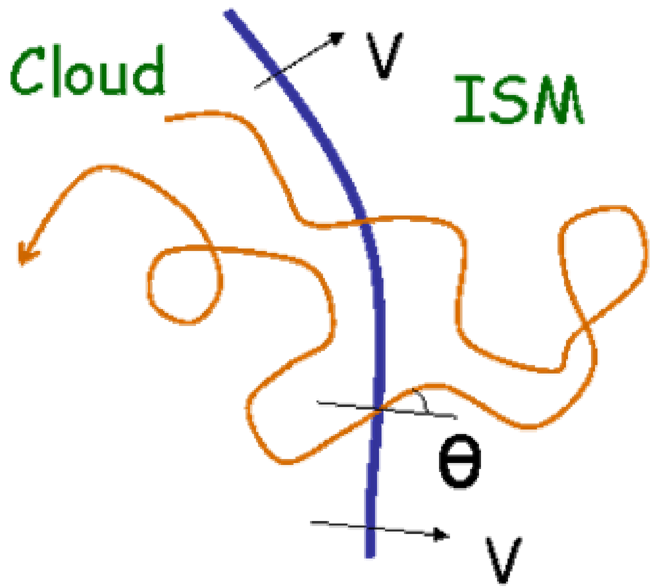


Figure 8.4 :Diffusive shock acceleration. The random walk of the accelerated cosmic particle is the result of magnetic fields, not of collisions on ISM particles.

particle coming from the shock will interact in a random walk with the magnetic field inhomogeneities and have a fair chance to return to the shock and traverse it once more. Two points are essential here. First, in its "interactions" with the magnetic fields, which are nothing more than magnetic bends, the cosmic particle changes momentum but it retains its energy. Second, we can talk of a "cloud rest frame" and of an "ISM rest frame" moving toward each other at velocity V. A cosmic particle, having energy E and momentum p in the rest frame of the medium in which it is, reenters the shock at an incidence θ_{inc}. It comes out in the other medium with energy $E'=\gamma\beta\, p\, cos\theta_{inc}+\gamma E$ with $\beta=V/c$ and $\gamma^2=1/(1-\beta^2)$. To first order in V/c, $E'= E+\beta\, p\, cos\theta_{inc}$, and, as the cosmic particle is relativistic, $E\sim p$, hence an energy gain (it is always a gain, never a loss, whatever the direction of the shock traversal) $\Delta E/E= \beta\, cos\theta_{inc}$. Depending on assumptions $< cos\theta_{inc} >$ may take different values; a reasonable model gives 2/3. Then, after n pairs of traversals, the energy has increased to $E_n=E_0\ (1+4\beta/3)^n$. However, at some stage, the cosmic particle may ultimately escape the system and acceleration will then stop. Writing P_{esc} the escape probability (assumed to be E-independent) the fraction of remaining cosmic protons after n cycles is simply $(1-P_{esc})^n$

Hence $dN/dE=-NP_{esc}/(4E\beta/3)$, namely $N=N_0\ (E/E_0)^{(-Pesc/(4\beta/3))}$. This is a power spectrum having an index $\nu= P_{esc}/(4\beta/3)$. As P_{esc} is more or less proportional to β, the spectral index may be expected to take similar values in different configurations. Putting numbers in for the size and age of the SNR one finds that this mechanism is in principle able to accelerate particles up to 100 TeV or so. Its extension to shocks occurring in AGN's might then be able to explain the highest cosmic ray energies, around 10^{20}eV.

The multitraversal aspect of the story is essential. The stochastic scattering on field inhomogeneities is claimed to be the cause. It implies that the Larmor radius of the cosmic particles (that of course keeps increasing during the acceleration phase) is large with respect to the shock thickness and at least of the same order of magnitude as the characteristic scale of the magnetic field inhomogeneities, this is necessary for making it meaningful to describe the Brownian motion as occurring in the rest frame of the medium

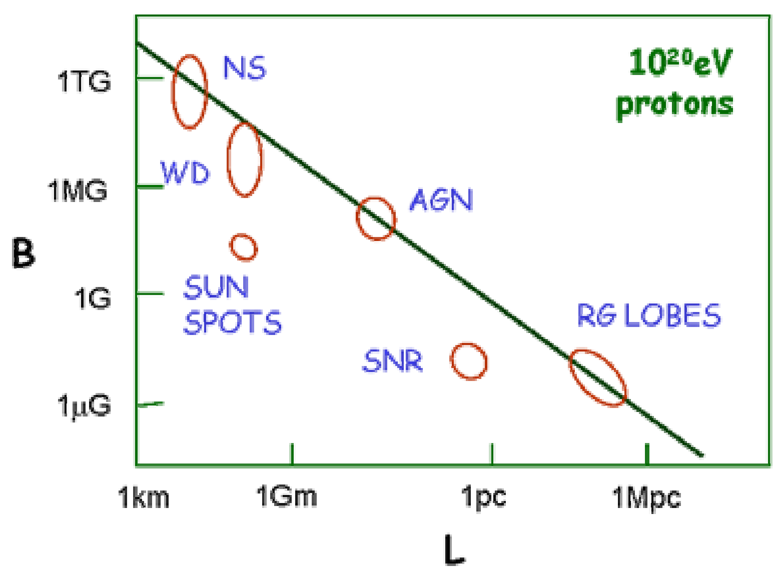


Figure 8.5 :Hillas plot showing the minimal value of the product BL (magnetic field × size) for accelerating a proton to 10^{20} eV. The limit is shown as a line in the log-log plot (note the scales!). A few candidate sites are indicated.

concerned. Note that the ISM and SNR plasmas in which the cosmic particles are moving are themselves subject to proper oscillation modes that may be resonantly excited in particular conditions. Finally one should mention that, like most other mechanisms that have been studied, diffusive shock acceleration requires that the accelerating site has a large size and is the seat of a large magnetic field. Quantitatively it is the product of these two quantities that defines the maximal attainable energy. This is illustrated in Figure 8.5, usually referred to as the "Hillas plot".

8.4 Gamma ray astronomy.

Gamma ray astronomy has recently made rapid progress in the highest energy domain thanks to arrays detecting the Cherenkov light produced by the electromagnetic shower upon entry of the gamma ray in the atmosphere. Below 20 GeV or so the rate is high enough to be detected by satellite experiments.

Solar flares exhibit a continuous component up to 100 MeV or so that is interpreted in terms of interactions with matter of charged particles that have been accelerated. Indeed, this is the explanation given of most of the continuous gamma emission wherever it comes from. Three processes are invoked: bremsstrahlung, that is emission of forward photons when the particle is bent in a magnetic field; Compton scattering, $e+\gamma \rightarrow e+\gamma$, called inverse Compton scattering by astronomers to insist on the fact that the projectile is an electron and the target a photon; and neutral pion decays, $\pi^0 \rightarrow \gamma\gamma$, a process that occurs above the pion (m=135 MeV) production threshold. On some occasions, lines are also detected: nuclear lines such as the 4.4 MeV carbon line and the 6.1 MeV oxygen line or the 0.511 MeV electron-positron annihilation line.

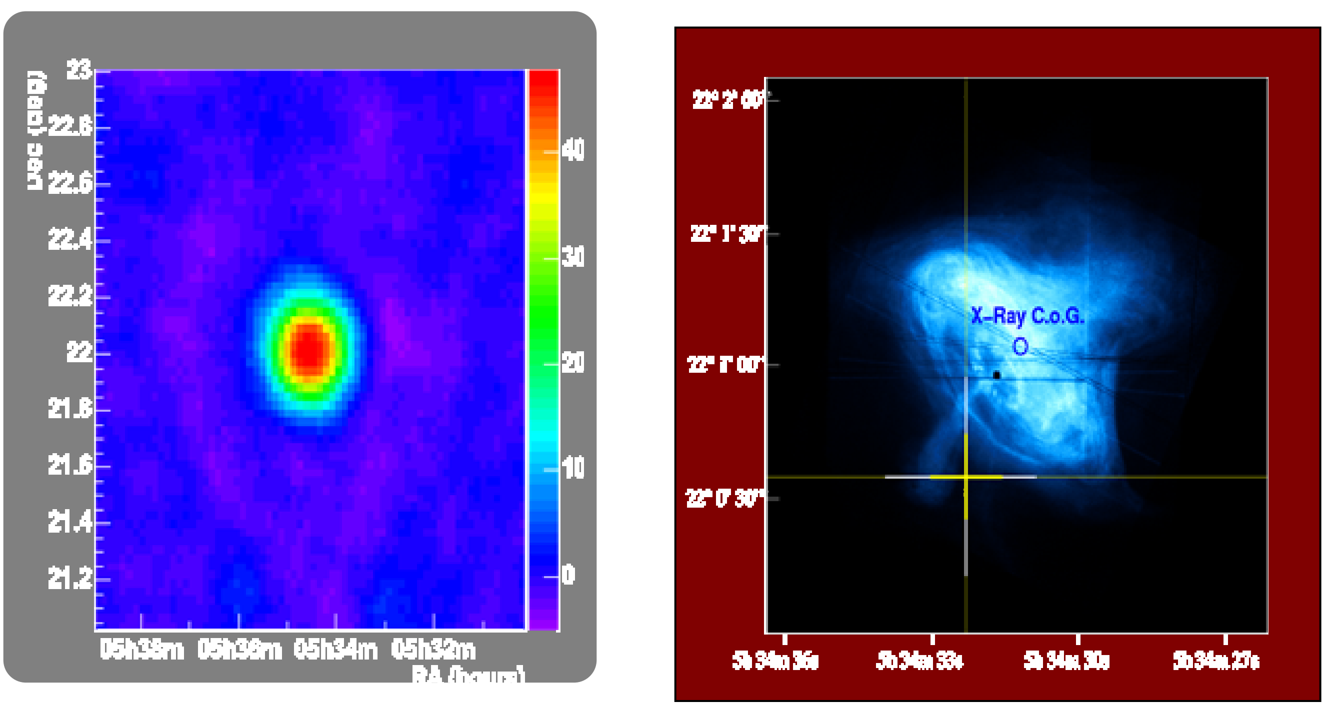


Figure 8.6 : The Crab (right) and its gamma counterpart (left). The cross on the right indicates the centroid of the gamma signal. Note that the central ellipse on the left is six times larger than the whole right figure.

Galactic gamma rays, coming from the disk region of the Milky Way, are mostly of the kind described above, produced by the interaction of charged particles with the ISM in the disk. However some room is left for discrete sources, such as pulsars and accreting black holes, of which some specimens have been clearly identified. In particular the Crab (Figure 8.6) has been detected very early as a strong gamma source and its spectrum measured up to 100 TeV or so.

Similarly, several AGN's have been identified as high energy sources, in the TeV region or above: high energy gamma rays appear to be very clean tracers of the highest energy charged cosmic rays. In AGN's one may distinguish between several possible acceleration sites:

i) the accretion disk, with ~mG fields and densities of the order of 5 10^8 protons/cm^3, may, under favorable conditions, accelerate particles to relativistic energies but they must be the seat of important radiative losses which make it unlikely that they could contribute to the high energy part of the spectrum;

ii) jets, where the magnetic field B decreases with distance R as BR~10 mG pc for a black hole of one million solar masses (BR is proportional to the square root of the mass of the black hole); at large scales there are 10^{-2} to 10^{-5} p/cm^3; at the pc scale very massive blazars may accelerate protons in this region that would subsequently generate showers, part of which would succeed to escape; at the kpc scale there is evidence for the presence of very high energy electrons that emit synchrotron radiation in the hard x range;

iii) hot spots, where the jets meet the ISM and where the magnetic field is between 0.1 and 1 mG and the density about 10^{-2} p/cm^3, may be the best candidates to produce ultra high energy protons by diffusive shock acceleration up to 10^{20} eV, but this remains to be proven.

9. Appendix

9.1 Gravitation

Newton's law reads $F=Gmm'/r^2$, where F is the force exerted between two masses m and m' at a distance r from each other. The gravitational energy of a hydrogen atom is (1GeV × ½MeV / 1.4 10^{38} GeV^2) /0.5 10^5 fm /200 MeVfm, namely 1.4 10^{-38} eV compared to 10 eV for its electromagnetic binding energy! Yet, in a star that contains 10^{58} such atoms (10 solar masses), it dominates over all other forces. This is the result of its cumulative properties: masses have all the same sign and are not shielded from each other. In contrast, as matter is neutral in terms of electric charge and of color, these forces are screened at very short distances. For what concerns the weak force it is intrinsically short range.

The two body problem is easily solved as shown below and its solution is expressed in terms of the three Kepler's laws.

$\boldsymbol{V}=d\boldsymbol{r}/dt$ and $\boldsymbol{L}=\boldsymbol{r}\times\boldsymbol{V}$ imply $d\boldsymbol{L}/dt=\boldsymbol{r}\times d\boldsymbol{V}/dt$. But, as $\boldsymbol{F}=md\boldsymbol{V}/dt$ (in the rest frame of m') is directed along $\boldsymbol{r}$, $d\boldsymbol{L}/dt=0$ and $\boldsymbol{L}=\boldsymbol{L_0}=cte$ (the angular momentum is conserved): this says that the area swept by $\boldsymbol{r}$ increases linearly with time, it is the first Kepler's law. It also implies that the orbit is planar (in the plane normal to $\boldsymbol{L}$). A trivial solution is that of a circular orbit with constant V and r, normal to each other. Calling α the angle between $\boldsymbol{V}$ and $\boldsymbol{r}$, we have $L_0=Vr\sin\alpha$ and, writing that the total energy is conserved (and negative for a bound system), $-E_0=\frac{1}{2}mV^2-Gmm'/r$ we find a relation between r and α that is the equation of the orbit: $r^2-2ar+b^2/\sin^2\alpha=0$, with $a=\frac{1}{2}Gmm'/E_0$ and $b^2=L_0^2/(2E_0/m)$: this is the equation of an ellipse having a as major axis, b as minor axis and m' at one of its foci. It is the second Kepler's law. Note that $b^2/a=L_0^2/(Gm')$. Finally, in one revolution of period T, $\boldsymbol{r}$ sweeps the area of the ellipse, πab. Hence $(L_0T)^2=(\pi ab)^2$ and $T^2/a^3=(\pi b)^2/aL_0^2=\pi^2/(Gm')=cte$. This is the third Kepler's law.

Special relativity introduces Lorentz transformations that mix space and time. A Lorentz transformation along a given space axis, say z, is very similar to a rotation around the same axis. Taking the velocity of light, c, as unit velocity, they read $z'=z\,ch\alpha+t\,sh\alpha$, $t'=z\,sh\alpha+t\,ch\alpha$ and $x'=x\cos\alpha+y\sin\alpha$ $y'=-x\sin\alpha+y\cos\alpha$ respectively. An elegant way to unify them is to introduce a metric $(1,-1,-1,-1)$ in space time (t,x,y,z), namely $g_{ij}=g_{ij}\delta_{ij}$, $g_{00}=1$, $g_{11}=g_{22}=g_{33}=-1$. Scalar products are now expressed as $U.V=\sum_{ij}U_ig_{ij}V_j$ and are conserved in any transformation of the Poincaré group, namely space time translation, space rotation or Lorentz transformation. The so-called "principle of relativity" states that the laws of nature are the same in any "inertial" frame, meaning any frame in which a freely moving body proceeds with constant velocity. The quantity α is the rotation angle in the case of space rotations. In the case of Lorentz transformations it is related to the velocity β of the final system with respect to the initial one, through the expression $\beta=th\alpha$. The equivalent of α, which increases by a constant quantity in a Lorentz transformation, is called rapidity.

One notes that $dz'/dt'=(dz\,ch\alpha+dt\,sh\alpha)/(dz\,sh\alpha+dt\,ch\alpha)=(dz/dt+\beta)/(1+\beta dz/dt)$, implying that dz/dt is confined to the interval $[-1,1]$: it is not possible to exceed the velocity of light. Relativistic dynamics is built up by extending the Lagrange formulation of classical mechanics to the new space-time metric. It starts by the assumption that for any system moving from a to b in space-time one can define a scalar action $S=\int_a^b dS$ where dS runs along the path followed by the system and such that the integral be minimal. The only scalars one can build in the case of a free particle are of the form $dS=-\lambda ds$, with $ds^2=dt^2-dx^2-dy^2-dz^2$. Namely a free particle follows a straight line in space time, at velocity β, $ds=(1-\beta^2)^{\frac{1}{2}}dt$. One defines the Lagrangian L from the identity $S=\int Ldt=\int dS$, where the first integral runs from t_a to t_b and the second from a to b. The free particle Lagrangian is therefore $-\lambda\partial s/\partial t=-\lambda(1-\beta^{2)\frac{1}{2}}$. It is of course unimportant to specify the value of λ but, in order to be consistent with classical mechanics, we set it equal to m: the free particle Lagrangian reads $L=-m(1-\beta^{2)\frac{1}{2}}$. We next define the momentum $\boldsymbol{p}$, a space vector having coordinates $p_x=\partial L/\partial\beta_x$, $p_y=\partial L/\partial\beta_y$, $p_z=\partial L/\partial\beta_z$. For a free particle $\boldsymbol{p}=m\boldsymbol{\beta}/(1-\beta^2)^{\frac{1}{2}}$. Then the force acting on the particle is simply $d\boldsymbol{p}/dt$. The energy of the particle is defined as $E=\boldsymbol{p.\beta}-L$, which becomes for a free particle $E=m(\beta^2+1-\beta^2)/(1-\beta^2)^{\frac{1}{2}}=m/(1-\beta^2)^{\frac{1}{2}}$. For $\beta=0$ one has $E=m$, the well known Einstein relation. We see that energy E and momentum $\boldsymbol{p}$ form a 4-vector and its norm, $m=(E^2-\boldsymbol{p}^2)^{\frac{1}{2}}$ is invariant. This formalism may be extended to any system consisting of several particles and electromagnetic fields. The action is then the sum of a particle term, a field term and a mixed term:

$S=-\sum mds-(16\pi)^{-1}\int F^2_{ik}\,dx^4+\sum\int eA.dx^4$ where e is the electric charge, F the electromagnetic field tensor and $A=(\boldsymbol{A},\varphi)$ the electromagnetic 4-potential defined from the electric and magnetic fields, $\boldsymbol{E}$ and $\boldsymbol{H}$, as: $\boldsymbol{H}=rot\boldsymbol{A}$, $\boldsymbol{E}=-\partial\boldsymbol{A}/\partial t-grad\varphi$. The electromagnetic field tensor is defined as $F_{ik}=\partial A_k\partial x_i-\partial A_i\partial x_k$, namely:

$$F=\begin{matrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & -H_z & H_y \\ -E_y & H_z & 0 & -H_x \\ -E_z & -H_y & H_x & 0 \end{matrix}$$

From there on one can derive the relativistic forms of the Hamilton-Jacobi and Lagrange equations, which govern the dynamics of any system, exactly as in the case of classical mechanics.

The extension to gravitation, referred to as general relativity, follows the same general path with two additional ideas: first one notes that locally, namely in a small region of space-time where the gravitational field is uniform and constant, there exists a non-inertial frame in which the system moves freely (at constant velocity), this is called the “principle of equivalence”; second, one notes that in any gravitational field all particles move in the same way, independently from their mass or any other attribute they may have. Hence the idea to describe the effect of the gravitational field by a modified metric, it being understood that a different metric will need to be used in different space-time points if the gravitational field varies with time or from one place to the next. With a modified metric that mimics exactly the effect of the field one can then retain the above formalism and find the movement of the system by minimizing the action, the world line which it follows being a geodesic. A first remark is that a diagonal metric, such as that which was used in special relativity, is no longer viable. It is obvious by looking at special transformations such as a space rotation or Lorentz transformation having α increasing in proportion to time, both of which generate a constant acceleration.

For example, the case of a space rotation, $\alpha=\omega t$, gives:

$$-ds'^2=(dx\ cos\alpha - xsin\alpha\ \omega dt + dy\ sin\alpha + ycos\alpha\ \omega dt)^2 +$$
$$(dy\ cos\alpha - ysin\alpha\ \omega dt - dx\ sin\alpha - xcos\alpha\ \omega dt)^2 + dz^2 - dt^2 =$$
$$= -ds^2 + 2\omega dt\ (ydx - xdy) + \omega^2 dt^2 (x^2+y^2)$$

The crossed terms imply the use of a non-diagonal metric g_{ik} (however obviously symmetric in the exchange of i and k). Spaces equipped with such a metric are well known in mathematics, they are associated with the concept of curvature. We are therefore describing the effect of gravitation by introducing a metric tensor at each point of space-time that corresponds to a given local curvature. Its dependence on space-time is conveniently expressed in the form of a curvature tensor R in terms of the Christoffel symbols, $\Gamma^i{}_{kl}=\frac{1}{2}g_{im}(\partial g_{mk}/\partial x_l+\partial g_{ml}/\partial x_k-\partial g_{kl}/\partial x_m)$:

$$R_{ik}=\partial\Gamma^l{}_{ik}\partial x_l-\partial\Gamma^l{}_{il}\partial x_k+\Gamma^l{}_{ml}\Gamma^m{}_{ik}-\Gamma^l{}_{mk}\Gamma^m{}_{il}\ .$$

One may then proceed with minimizing the action and describe the dynamics of the system. One notes that the motion of a particle in a gravitational field is simply governed by the equation:

$$d^2x^i/ds^2+\Gamma^i{}_{kl}\ dx^k/ds\ dx^l/ds=0.$$

Finally one needs to relate the energy content of space-time to curvature. The energy content rather than the mass content because energy is expected to "weigh": think of a constant vertical gravitational field having g as acceleration of gravity. Take two points A and B on top of each other, A above B at a distance h of it. Send an energy E from A to B in the form of radiation. B receives an energy E'. To evaluate it consider the event in the non-inertial system where the gravitational field vanishes. This system starts at zero velocity when you send E from A and reaches a velocity gt when you receive it in B, with $t=h$ being the time it took for the radiation to go from A to B. Therefore the energy $E'=E(1+gh)$, namely E has acquired an excess energy Egh in the gravitational field, corresponding to the usual mgh term in classical mechanics (m being the rest mass) and it is indeed E and not m which we must consider.

In fact, in order to relate energy to curvature, it is not sufficient to know how much energy is contained in a local volume element of space-time but also how it varies when moving away from it. This is done in terms of a set of numbers that measure the energy and space momentum densities and fluxes in the volume considered. They form a tensor, the energy-momentum tensor T_{ik}. In Galilean coordinates, T_{00} is the energy density, $T_{0\alpha}$, the components of the momentum densities, $T_{\alpha\beta}$ the components of the tensor of momentum flux densities with $\alpha,\beta=1,2,3$. The relation between the curvature tensor R and the energy-momentum tensor T is then given by:

$$R_{ik}-\frac{1}{2}\ g_{ik}\ g^{lm}R_{lm}=8\pi\kappa\ T_{ik},\quad \text{with } \kappa \text{ a constant.}$$

These are the fundamental equations of general relativity, often referred to as Einstein equations. They reduce to Friedmann equations in the case of a homogeneous and isotropic space (and ignoring a possible cosmological constant), with the metric reducing to the Robertson-Walker metric, and to Newton's equation in the non relativistic limit.

9.2 Particle physics

Particle physics gives a description of the world as being made of elementary particles that interact via the electromagnetic, weak and strong interactions. This description, that ignores gravitation, is usually referred to as the "standard model". The standard model is remarkably successful in accounting for the observed phenomena. Yet it is unsatisfactory in many respects and current efforts concentrate on finding a new theory, of which the standard model would be a low energy approximation, which would include gravitation. Presently, the best candidate for such a theory is called the M-theory, based on superstrings. It is premature to say more in the present introduction but one should be aware of this trend. Here we shall be satisfied with giving a brief introduction to the standard model.

There are 10^{80} or so electrons and nucleons in the universe, 10^{90} or so photons and neutrinos, but this makes only four different kinds of particles! Indeed present day's particle physics describes the world as being made of a very few number of different particles. The description starts from the idea that there exist symmetries in the universe, both space-time symmetries and exchange symmetries. Space-time symmetries include invariance under space-time translations, space rotations and Lorentz transformations as well as supersymmetry. Of the latter, we shall only say a word at the end but we shall ignore it for the time being. Space-time symmetries imply that particles are expected to have a well defined mass and spin, that they should come in pairs of particles and antiparticles, that their energy-momentum and their covariant spin should transform as four-vectors. Particles are expected to obey different statistics depending on whether their spin is half-integer or integer: one calls them fermions in the former case, bosons in the latter. Of the fermions we know essentially only two kinds: the lepton and the quark. Leptons include electrons and neutrinos, the latter being produced, for example, in β-decay. Quarks are spin ½ objects of which nucleons are made. Of the bosons we know only three: the photon, the gluon and the weak boson, all of which have spin 1 (one calls them vector bosons).

Exchange symmetries express the invariance of the laws of nature under the exchange of different kinds of particles. They are represented in Hilbert space by unitary transformation matrices belonging to the group SU(n), n being the number of kinds of particles being exchanged. This implies that we may label a given kind of particle with an index *i* within a broader set and express a given law in a form independent of *i*. In practice the known fermions may be labeled by a set of indices describing their classification under three such symmetries: SU(3) for the strong force, SU(2) for the weak force and U(1) for the electromagnetic force.

SU(3) describes the strong force. Leptons belong to a singlet representation, i.e. there is no exchange possible and they ignore the strong interaction. Quarks belong instead to a triplet representation. This means that a given quark may be labeled by an index i=1, 2 or 3 but that the strong force is independent of the value of that index. Rather than talking of 1, 2 and 3 one usually talks of quarks of three different "colors". A nucleon is made of three quarks arranged in such a way that the total color vanishes. There exist other three-quark states that correspond, loosely speaking, to excited states of the nucleon. All three quark states are colorless and are called baryons. There exist also colorless quark-antiquark states, such as the pions, their generic name is mesons.

SU(2) describes the weak force. Here right-handed particles belong to a singlet representation: they ignore the weak interaction. Left-handed particles, whether quarks or leptons, belong instead to a doublet representation: one talks of a weak isospin that may be up or down. Here left-handed or right-handed refers to the orientation of the spin with respect to the momentum in the high energy limit (or for massless particles). One gives different names to particles having their weak isospin up or down: one speaks of the up and down quarks and, in the case of leptons, of the neutrino and electron. Of course this applies strictly to left-handed particles only but one extends this vocabulary to right-handed particles.

Finally U(1) describes the electromagnetic force. In order to unify the description of quarks and leptons, left-handed and right-handed, one introduces the hypercharge Y rather than the usual electric charge Q. A U(1) transformation simply rotates a state vector in the Hilbert space by a phase factor proportional to the hypercharge of the particle. For right-handed particles Y=Q, while for left-handed particles with weak isospin T, $Y=Q-T_3$.

At this point we may therefore summarize the classification of fermions as shown in the table below. If that were all, we might say that there is only one kind of fermions, labeled by different indices.

	Color	Weak isospin	Hypercharge
Quark L	Triplet	Doublet	1/6
Quark R	Triplet	Singlet	u = 2/3, d = –1/3
Lepton L	Singlet	Doublet	–1/2
Lepton R	Singlet	Singlet	e = –1, ν = 0

However, each of the fermions listed in the table may exist in two additional different forms, one speaks of three fermion families or generations. The new particles differ from the former by nothing but their mass. In particular they behave exactly in the same way as far as the strong, weak and electromagnetic forces are concerned. It is then possible to add a new index to our list that will define the family; however we do not know how to deal with the corresponding symmetry: we know of no interaction that changes the members of one family into those of another family (at variance with the strong interaction that changes quark colors). Transitions between two families are only possible because the gauge bosons ignore families and couple equally well to one or the other. The names extending the up-down quark doublet are charm-strange and top-bottom; the new charged leptons are called μ and τ, neutrinos being simply called e-, μ- and τ-neutrinos.

The next step in the elaboration of the standard model is to introduce a new symmetry referred to as gauge invariance. It states that the phase of any fermion state may be chosen at wish at any point of space time without changing the physics. This is in general not possible: such phase changes modify the scalar invariants that one can construct from a free fermion state. However, this problem disappears if one introduces a new massless vector boson that couples to the fermion in such a way as to exactly compensate for the modification of the scalar invariants under gauge transformations. Such massless vector bosons are called gauge bosons. In the case of U(1), all what this means is the addition of a massless vector boson that couples to the charged fermion exactly as the photon does: requiring gauge invariance kind of requires the existence of the photon and completely defines its coupling to the fermion. Similarly, in the case of SU(2), one needs to add a massless vector boson, W, called the weak boson and behaving as a triplet under SU(2); in the case of SU(3), one needs to add a new massless vector boson, called the gluon, behaving as an octet under SU(3). Having introduced new vector bosons, one must also be able to describe them in the absence of the fermions to which they couple: the free boson equations are completely determined by space-time symmetries as are Maxwell equations in the U(1) case of the photon.

We are not through yet because, if the photon and the gluon are indeed massless, the weak boson is a very massive object: it has a mass of the order of 100 nucleon masses. Precisely, there exists a neutral weak boson, the Z, having a mass of 92 GeV, and two charged weak bosons, the $W^{\pm}$, having a mass of 80 GeV.

But before tackling this problem, let us comment further on the strong force. The description obtained in the case of SU(3) does not require any additional ingredient; it provides an exact theory of the strong interaction with only one free parameter: the coupling constant α_s of the gluon to the quark. However, while the photon is not charged, the weak boson carries weak isospin and the gluon carries color. This implies that, while two photons do not interact directly with each other, weak bosons do and gluons as well. This is a major difference compared to quantum electrodynamics (QED). In the SU(3) case, the resulting dynamics, quantum chromo-dynamics (QCD), describes a force that increases with distance rather than decreasing. This means that quarks and gluons at short distances are essentially free (one speaks of asymptotic freedom) while it is not possible to take them apart at large distances. There exist no free quarks, one speaks of confinement (inside baryons or mesons), and perturbative calculations are not possible in this confined regime. Yet, at very high densities or temperatures, as was the case in the universe a microsecond or so after the big bang, or as may be the case today in the core of some neutron stars, quarks and gluons are close enough from each other to move freely: one speaks of a quark-gluon plasma to describe this new phase of matter.

To complete the construction of the standard model, one needs to add something to it that will make it possible for the three-component weak boson to acquire a mass. It is done by assuming the existence of a new scalar (spin zero) particle having four components described as a weak isospin doublet of two complex fields. All what is required of this particle is that the ground state break the symmetry, one speaks of spontaneous symmetry breaking. What happens then is that three of the degrees of freedom of the new field, instead of materializing in Goldstone bosons, become longitudinal components of the three weak boson components: they acquire a mass. The fourth degree of freedom survives as a new scalar particle having well defined couplings to the rest of the world: the Higgs boson. Moreover, the gauge boson of U(1) and the neutral component, W^0, of the W of SU(2) mix with an angle θ_W, called the "weak" or "Weinberg" angle, that is a well defined function of the U(1) and SU(2) coupling constants, α and α'. The results of this mixing are the photon and the Z that are observed in nature. With only three parameters, α, α', and a constant v defining the symmetry-breaking scale, ~250 GeV, one has been able to calculate the W and Z masses, the exact form of their couplings to all fermions, and, at the same time, to calculate any QED quantity. However the Higgs mass remains undefined as a fourth free parameter of the theory. Worse, the same spontaneous symmetry breaking mechanism generates also fermion masses and mixing matrices but this host of new parameters are again

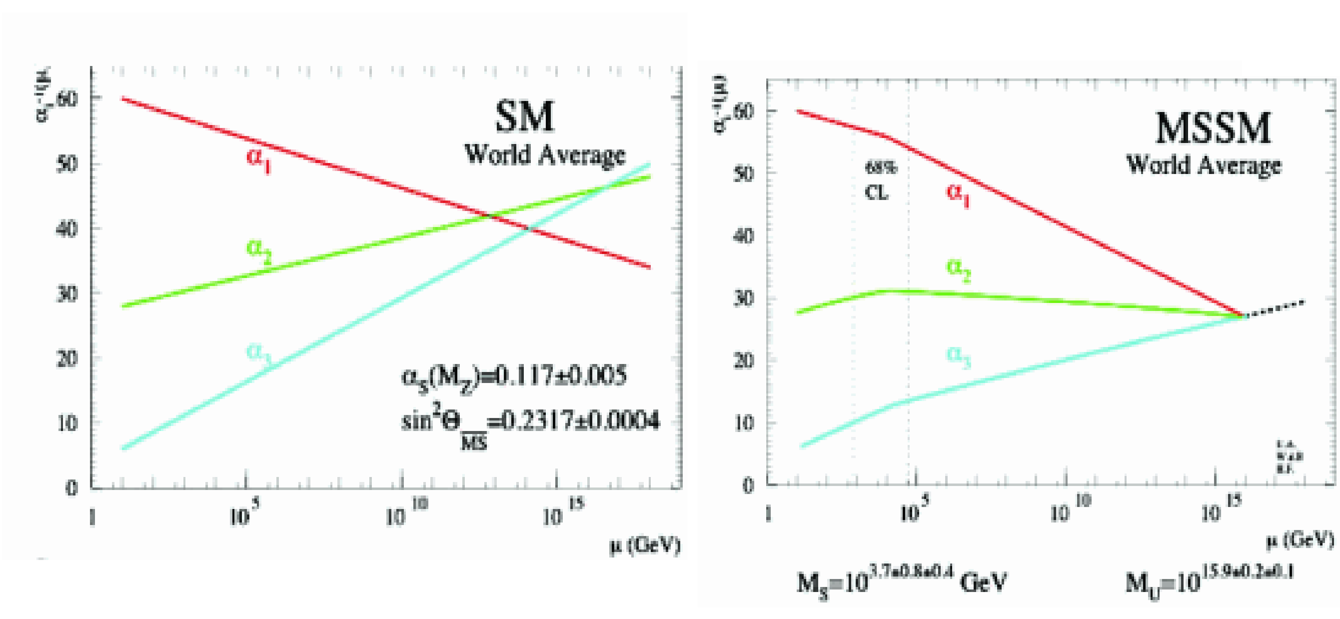


Figure 9.1 : The convergence of the strong, electromagnetic and weak couplings at a grand unification mass (GUT) of 10^{16} GeV. The quantity plotted is the reciprocal of the coupling constant as a function of the mass scale used in the renormalization group equation. The figure on the left is for the standard model while that on the right is for the minimal supersymmetric standard model (MSSM). The better convergence is considered as an argument in favour of SUSY.

unconstrained free parameters. For the time being no Higgs boson has ever been detected but there are strong indications that its mass is within reach of a new accelerator, the LHC, currently under construction in Europe.

Even if no experiment has yet been able to prove the standard model wrong, one can not be fully satisfied with it. In addition to having too many free parameters, it ignores gravitation; it does not explain why there must be three identical fermion families; it suggests an attractive solution to the mass generation problem but doesn't unveil which physics it is hiding; it hints at a higher symmetry –one speaks of grand unification theory, GUT– that would embed each of U(1), SU(2) and SU(3) (through the convergence of the running couplings, see Figure 9.1, at a mass of the order of $\pi\ 10^{16}$ GeV) but it does not tell us how to realize this grand unification; it does not explain why the spontaneous breaking of the electroweak symmetry into SU(2)×U(1) occurs at such a low mass (250 GeV compared to 10^{16} GeV for the GUT scale).

Supersymmetry appears as an excellent candidate to solve these weaknesses. It is a fundamental symmetry of space-time that had been overlooked for a long time because it acts on doublets made of a fermion and of a boson, with spins differing by one half, say, for example, a scalar Φ and a spinor Ψ. An infinitesimal transformation on such a doublet might read something like $\Phi'=\Phi+\langle\xi|\Psi\rangle$ and $\Psi'=\Psi+\Phi\xi$, requiring that the infinitesimal parameter ξ be a spinor and not a scalar as it is in other symmetries. Indeed such a symmetry exists (but the second relation contains a derivative) and has remarkable properties. First, the commutator of two such transformations is proportional to the particle's momentum: this provides a direct link to gravitation and, through the formalism of supergravity, a route toward an inclusion of gravity in the theory. Second, the theory predicts that all particles exist in doublets, a fermion and a boson. Today, none of the supersymmetric partners of the known particles has been found. Supersymmetry, if it exists in nature, must therefore be badly broken. But not so badly in fact: compared to the GUT scale all known masses are very small. The idea is then that all particles are massless to start with. Any spontaneous breaking mechanism would, in principle, tend to let particle masses increase to values commensurate with the GUT scale. However, fermions cannot acquire large masses because their mass term is of the form $\psi_L\psi_R$, coupling the right and left handed states that behave differently under SU(2)×U(1); and boson masses are locked to fermion masses via supersymmetry. Hence the abnormally low value of the electroweak symmetry breaking mass scale and, therefore, of the fermion and boson masses. Note that the lightest supersymmetric partner must be stable and weakly interacting: it makes an ideal candidate for dark matter.

Before closing this brief summary, let us say a word about Feynman graphs. The transition probability between an initial state i and a final state f is the square of a complex quantity called the transition amplitude. If the interaction is not too strong (namely if perturbative calculus can be used) the transition probability can be calculated as a perturbative series. Each term can be represented by a Feynman graph as shown in Figure 9.2. In such a graph each line represents the 4-momentum of a given particle. The open ended lines are associated with the initial and final state particles. The inner lines may be any particle as long as their couplings belong to the list permitted by the standard model. Contrary to the open lines, the 4-momenta associated with the inner

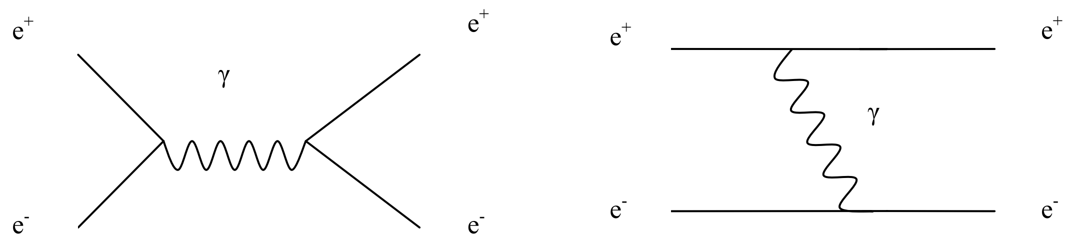


Figure 9.2 : Feynman graphs describing the leading order contributions to the electro-magnetic interaction between an electron and a positron. The diagram on the left dominates at low energies while that on the right dominates at low 4-momentum transfers. As each diagram corresponds in fact to a scattering amplitude, the cross-section, obtained by squaring their sum, includes an interference term between the two elementary processes (annihilation and exchange).

lines are not constrained to be "on mass shell", meaning that their norm $m=(E^2–p^2)^{½}$ is not constrained to be equal to the rest mass m_0 of the particle. One speaks of "virtual" particles. The higher their virtuality, i.e. the more off-mass shell they are, or the larger $(m–m_0)^2$, the smaller their contribution to the transition amplitude, typically by a factor $(m–m_0)^{–2}$. The simpler diagrams are called "leading" or "tree level". All other diagrams contributing to the same transition are obtained by adding lines to the leading diagrams. In most cases the series diverges, which can be taken care of be introducing an arbitrary cutoff and redefine the coupling constants as depending on this cutoff (one speaks of running coupling constants). Note that the series are on amplitudes, not on probabilities: hence possible interference terms when calculating the probabilities.

9.3 Nuclear Physics

Nuclear physics studies atomic nuclei in a regime where they can be considered as made of nucleons, protons and neutrons, interacting via potentials. Such nucleon-nucleon potentials have a radial dependence characterized by a strong repulsive hard core at a fraction of one fm, surrounded by a short range attractive well up to a radius approaching 2 fm. They also have dependence on spin and on isotopic spin. To a good approximation the A nucleons of a nucleus are simply packed against each other. Indeed the cross-section of a typical nucleus scales as $A^{2/3}$ fm^2 (1 fm^2 = 10 mb) and its radius as $A^{1/3}$ fm. Nuclear interactions can therefore be often qualitatively described in terms of an optical potential. The depth of the attractive well is measured in MeV. Absorption is phenomenologically described by allowing for the potential to take complex values. The repulsive Coulomb potential between protons becomes important for heavy nuclei and results in a repulsive barrier at large radius. In many nuclear interactions the optical nature of this potential description dominates the scene and, because of the strong absorptive power of nuclear matter, generates diffractive patterns modulated according to the value of the nuclear radius, Fourier transforms of the source radial distribution.

The calculation of the wave function of the ground and excited states of a given nucleus is a complicated many body problem. It can be solved approximately by searching for a self-consistent Hartree-Fock central potential. Each single particle state can be occupied by four particles corresponding to the two possible orientations of each of the spin and isotopic spin. Here "isotopic spin" corresponds to the symmetry of the strong nuclear force with respect to the exchange of a neutron with a proton; it is slightly broken by the Coulomb interaction. Such a description of nuclei results in what is called the shell-model, with closed shells corresponding to "magic nuclei". For example ^{40}Ca (20 p and 20 n) and ^{208}Pb (82 p and 126 n) are doubly magic nuclei with both their proton and neutron shells fully closed. One therefore expects, and indeed observes, a cyclic dependence of the nuclear binding energy on A with minima corresponding to magic nuclei and with a further quaternary modulation superimposed on it. The most stable nucleus is ^{56}Fe (26 p and 30 n). It may be seen as two closed shells (28 p and 28 n) modified by the exchange of two protons for two neutrons in order to cope with Coulomb repulsion. Most of the even-even nuclei are quite stable and can often be thought of as clusters of α-particles. A notable exception is ^{8}Be which, although very well described by a pair of α-particles, is almost stable but not quite. Had it been stable, the primordial nucleosynthesis would have proceeded differently and the universe would have been dramatically different from what it is today. Another prediction of the shell model is the existence of excited states resulting from the excitation of one particle to a higher energy level (single particle excitations). Indeed, such excited states are known to exist, typical excitation energies are measured in MeV's.

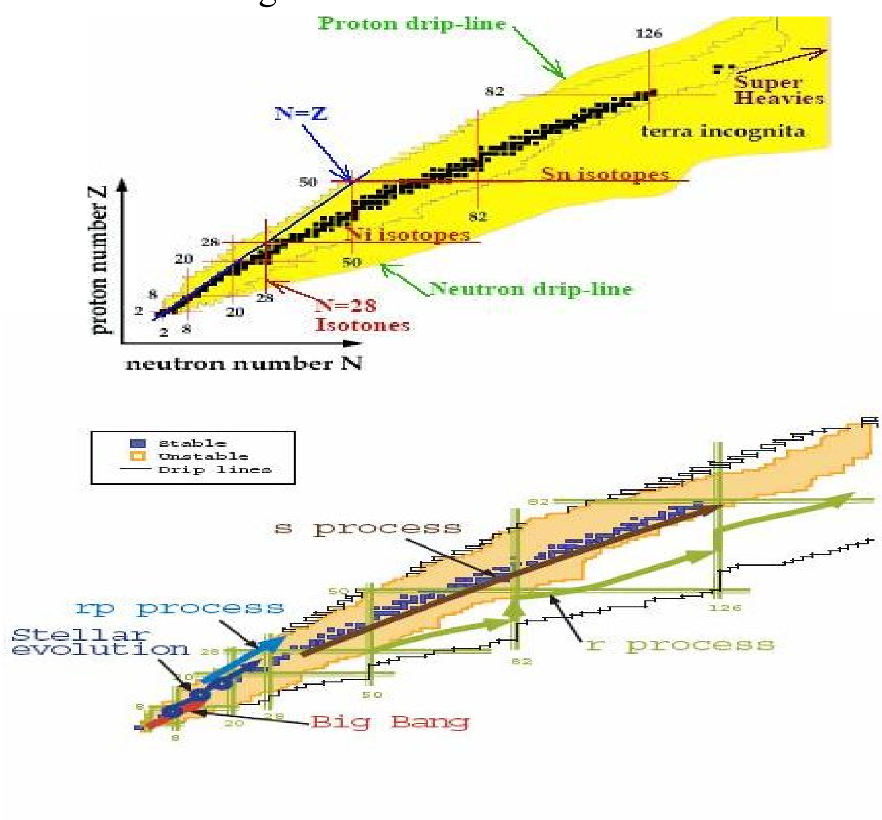


Figure 9.3 : The chart of nuclides showing the stability valley (top) and the main nucleosynthesis processes (bottom).

Because of Coulomb forces the ratio between the number of neutrons and the number of protons increases with A and the spherical symmetry may be broken in the excited states or even in the ground state: there exist oblate and prolate deformed nuclei. When rotating, they will generate a sequence of excited sates (rotation bands with energies growing as J(J+1), J being the total angular momentum quantum number). Nuclei may also be the seat of collective oscillations, again generating sequences of excited states.

Nuclei may decay spontaneously. Whenever changing a neutron into a proton or a proton into a neutron is energetically favorable, nuclei will β-decay. Heavy nuclei may spontaneously emit an α particle, i.e. a helium nucleus, or even fission into two pieces, one typically significantly larger than the other. Finally, excited states will usually cascade down to the ground state by photon emission (γ-radioactivity).

A chart of nuclides in the N (number of neutrons) versus Z (number of protons) plane is shown in Figure 9.3. On either sides of the region of stability, following more or less the line Z~0.7 N, one finds the bands of β^+ decaying nuclides on the proton-rich side and β^- decaying nuclides on the neutron-rich side where nuclei can be found up to the so-called neutron drip line that represents the limit of stability with respect to the addition of neutrons.

Nuclear reactions usually imply simply a redistribution of the participating nucleons among the nuclei. They are characterized by a cross-section, a measure of their probability of occurrence, and by an energy balance that may be positive or negative. Simple considerations of nuclear structure based on the rudiments given above are often sufficient to understand whether a given reaction will be exothermic or endothermic, will have a large or a small cross-section. When a nucleus is hit by a high energy proton or α particle, it may fragment in two or more pieces, one speaks of spallation reactions. The notation N(x,x')N' describes the reaction x+N→x'+N'.

9.4 Plasmas, MHD

Magnetohydrodynamics (MHD) studies the movements of conducting fluids submitted to electromagnetic fields. The coupling between the movement of matter (described by usual hydrodynamics) and the fields (described by Maxwell equations) gives its specificity to this field of physics. The strength of the coupling depends on a dimensionless quantity, the magnetic Reynolds number, R_m. MHD applies to a very broad range of phenomena, from mercury or liquid sodium to ultra rarefied media such as those present in ISM. It finds many applications in astrophysics where plasmas are omnipresent: solar spots and flares, solar winds, dynamo effects in stars, pulsars and accretion discs, magnetospheres, SNR's, GRB's, diffusive shock acceleration, etc.

The basic equations are trivial: defining E the electric field, B the magnetic induction, j the current, v the velocity of the fluid, ρ the density, p the pressure, σ the conductivity and η the viscosity, we have, for a liquid:

$\mathrm{rot}\, E = -\partial B/\partial t$, $\mathrm{rot}\, B = \mu_0\, j$, $\mathrm{div}\, B = 0$, $j=\sigma(E + v \wedge B)$,   and

$\mathrm{div}\, v = 0$, $\rho\, dv/dt = j \wedge B - \mathrm{grad}\, p + \eta \nabla^2 v$.

From these we derive $\partial B/\partial t = \mathrm{rot}(v\wedge B) + \lambda\, \nabla^2 B$, with $\lambda\mu_0\sigma = 1$. This equation makes the coupling between B and v explicit. The first term in the rhs is the convexion term, the second the diffusion term. For zero viscosity one finds:
$\partial B/\partial t = \mathrm{rot}(v\wedge B)$ and $E + v\wedge B = 0$: the field is "frozen in matter".
The relative importance of the convexion and diffusion terms is measured by the magnetic Reynolds number, $R_m = \mu_0\, \sigma\, v\, l$ where l is the characteristic length over which v and/or B vary significantly. There is a strong similarity between the above equation and that obeyed by the curl of the velocity field, $\omega = \mathrm{rot}\, v$:
$\partial\, \omega /\partial t = \mathrm{rot}(v\wedge\, \omega) + \eta/\rho\, \nabla^2\, \omega$ with the usual Reynolds number defined as $R=lv\rho/\eta$.
The ratio $P=R_m/R$ is called the Prandtl magnetic number.

From the above one sees that B exerts on the flowing matter a magnetic isotropic pressure, p_m , and a force directed along the field lines, t_m , that are both proportional to B^2. In particular, for a conducting fluid in a static and uniform magnetic field B_0, low frequency waves, so-called Alfvèn waves, can propagate parallel to B_0 with a velocity $v_A= (t_m/\rho)^{½} = B_0/(\mu_0\rho)^{½}$.

Of particular importance to astrophysics is the case of fully ionized gases, the so-called plasmas. Consider the case of a low density plasma embedded in a magnetic field B and such that collisions between particles can be neglected. The particles (of mass m) move with angular frequency $\Omega=v_t/R=B/m$ along helices of radius $R=mv_t/B$ and pitch angle $\alpha=\mathrm{arctg}\ v_t/v_l$ having their axes on the field lines. Electrons and ions turn in opposite directions at very different frequencies and their helices have very different radii. When an electric field E is added, both electrons and ions drift in addition with a same velocity $V_{drift} = B\wedge E/B^2$ in a same direction perpendicular to both B and E. The resulting trajectories project as cycloids on a plane normal to B. When B, instead of being uniform, varies slowly over distances large compared to R, it is still possible to describe the movement in terms of slowly deforming helices having parameters defined locally. What may happen then is that the pitch angle reaches $\pi/2$ and changes sign. This is the well known magnetic bottle effect responsible for the oscillation of particles in the earth magnetosphere from north to south and back with a slow east-west drift: the Van Allen belts. Note that nothing of what we said in this paragraph has much to do with MHD. Indeed it may seem abusive to use the language of MHD when dealing with such highly rarefied plasmas in cases where the field inhomogeneities are of the same order of magnitude as the mean field itself. Yet, it is possible to describe collective motions of such a plasma in macroscopic terms when external perturbations are applied and one may be able to extend such descriptions to cases where the perturbations are no longer external but self-consistently generated internally.

9.5 Spectroscopy

Spectroscopy is the study of the photon spectra resulting from radiative transitions between different states of a same system. One speaks of atomic and molecular spectroscopy depending whether the system is an atom (or ion) or a molecule. Nuclei also are potential photon sources (γ radioactivity), one talks of nuclear spectroscopy. The typical frequency scale is in the visible for atoms and from infrared to microwave for molecules. The x-ray range is associated with transitions between internal shells. Nuclei may emit photons in a broad range, keV to MeV. One distinguishes between emission and absorption spectra. The former are from the light emitted by a source, they give bright lines (characteristic of the source) on a dark background. The latter are from absorption of photons by some medium located in front of the source, they give dark lines (characteristic of the absorber) on a clear background. In all cases the frequency of an observed line is a direct measure of the difference between the energy of the initial state and that of the final state (emission lines correspond to transitions from a high energy to a lower energy state, absorption lines to the contrary). Depending on the spin and parity of the initial and final states, a single photon transition may be allowed or forbidden. One also distinguishes spectra according to their frequency domain: radio and microwaves, infrared, visible, UV, x-rays, and γ-rays.

Disregarding the line structure, a black body at temperature T radiates, per unit area, a power $P=\int dP/d\lambda\ d\lambda$; its dependence upon wave length is given by Planck's law: $dP/d\lambda=2hc^2\lambda^{-5}/(\exp[hc/kT\ \lambda]-1)$. Taking the derivative with respect to λ gives the Wien's laws: $\lambda_{max}\approx 2900\ \mu m.K/T$ and $T^{-5}dP/d\lambda\approx 8\pi\ \mu Wm^{-3}K^{-5}$ at maximum emission. Integrating instead over λ gives $T^{-4}P=5.7\ 10^{-12}\ Wcm^{-2}\ K^{-4}$, the Stefan-Boltzmann law.

The quality of a spectrometer is essentially defined by its resolving power, the minimal frequency difference above which two lines can be resolved, and by its luminosity, the minimal light intensity above which a line can be detected. In addition modern spectrometers are able to analyze many lines simultaneously: the solid angle that they can measure in a single go is an important figure of merit. In principle, from the various intensities of the lines observed, one can infer the relative abundance of the corresponding elements.

The Doppler effect is essential in astrophysics, not only for measuring the distances of remote galaxies as described in the lectures, but also to reveal double stars or even the rotation of a nearby object or of the accretion disc of a black hole.

In the case of a spherically symmetric system, its states are expected to be defined by an orbital quantum number l, a magnetic quantum number m and a set of other orbital numbers n. Moreover the $2l+1$ states having the same l and n but differing only by their m values are expected to be degenerated, namely to have the same energy. States having $l=0$, 1, 2, 3, 4, 5… are traditionally labeled as s (for sharp), p (for principal), d (for diffuse), f, g, h… Many effects lift the degeneracy, such as a spin-orbit coupling, the application of an external magnetic field (Zeeman effect), etc… In such cases, if the lifting is gentle, the line will appear as a multiplet split in its 2m+1 components.

The dependence over n of the Coulomb excitation energy is in $1/n^2$. Line frequencies obey therefore a law of the form $R(1/n^2-1/m^2)$ where R is the Rydberg constant, $R=11\mu m^{-1}$. For hydrogen and hydrogen-like atoms the series corresponding to n=1, 2, 3, 4 and 5 are named Lyman, Balmer, Paschen, Brackett and Pfund respectively. They range from infrared to ultraviolet. For ionized Helium the n=3 series is called after Rydberg and the n=4 series after Pickering.

In molecular spectroscopy one finds three families of lines corresponding to different frequency. The electronic lines correspond to the different electron energy levels and are similar to those of atomic spectra. They are mostly in the visible range. To each electronic line corresponds a band of vibration lines associated with oscillations of the atoms in the molecule. They are in the near to mid infrared. Finally, to each vibration line corresponds a band of rotation lines. They range from the far infrared to the microwave.

In the radio range spontaneous emission is very strongly depressed (probability inversely proportional to the cube of the wave length) and one sees only absorption lines (or induced emission lines). An important line is the 21 cm line of neutral hydrogen. Note that molecular hydrogen cannot be detected in the radio range; it is usually "traced" by the CO molecule, the abundance of which is proportional to that of H_I, and that can be detected.

Finally one should mention the importance of polarization measurements and several effects causing a broadening of the lines, in particular temperature.

9.6 Some useful numbers

All numbers below are approximate (sometimes grossly approximate) values; the idea is to encourage the student to remember the orders of magnitude, not to give precise values. Watch that ħ and/or c are sometimes set to 1; whenever they are, it should be obvious.

General

Light velocity: $3\ 10^8$m/s

Newton's gravitational constant: $7\ 10^{-11}$ m^3 kg^{-1} s^{-2}

Acceleration at the surface of the earth: 10 m s^{-2}

Planck mass: $1.2\ 10^{19}$ GeV

Grand unification mass: $\sim \pi\ 10^{16}$ GeV

Astronomical unit (AU)

= radius of the earth orbit around the sun

$=1.5\ 10^{11}$m

1 y = $\pi\ 10^7$ s

1 parsec (pc) = 1 AU/arcsec $\sim \pi\ 10^{16}$m

1 light year = 0.3 pc $\sim 10^{16}$m

Electron mass: 0.5 MeV

Proton mass: 1 GeV

Boltzmann constant: 0.1 meV/K

Planck constant (ħ=½h/π): ~200 MeV.fm ~ $6.6\ 10^{-22}$ MeVs

Rydberg constant: 11 μm^{-1}
Electron charge: 1.6 10^{-19} C; 1eV=1.6 10^{-19}J=1.6 10^{-12}erg
Avogadro number: 6 10^{23}
Magnetic bend: 0.3 GeV T^{-1} m^{-1} = 10^3 TeV μG^{-1} pc^{-1}
Luminosity (L) to magnitude (m) conversion:
$L_0 = \pi\ 10^{28}$W = zero magnitude luminosity
$L = L_0\ 10^{-0.4m}$
Sequence of multiples in steps of 10^3 from 10^{-15} to 10^{21}:
f(femto), p(pico), n(nano), μ(micro), m(milli), unity…
…k(kilo), M(mega), G(giga), T(tera), P, E, Z

At the scale of the solar system
Earth radius: 6.4 10^6m
Earth density: 5.7
Sun radius: 7 10^8 m =110 earth radii
Sun density: 1.4
Sun central temperature: 16 10^6K
Sun external temperature: 6 10^3K
Schwarzschild radius of the sun: 3km
Age of the solar system: 4.5Gy
Luminosity of the sun: 4 10^{26} W (~ 0.2 W/m^3)
Mass of Jupiter: 330 earth masses
Radius of Jupiter: 11 earth radii
Radius of the earth orbit: 1 AU= 8.3 light mn
Radius of Jupiter's orbit: 5.2 AU

At the scale of the Galaxy
Age of the Milky Way: 10 Gy
Diameter of the Milky Way:
Distance of the sun to the center of the Milky Way: 8 kpc
Distance to Proxima Centauri: 4.2 ly = 1.4 pc
Distance to Andromeda: 750 kpc
Mass of the Milky Way: 2 10^{11} solar masses
Velocity of the sun around the center of the Milky Way: 220 km/s
Local disk density: 5 GeV/cm^3
Local halo density: 0.5 GeV/cm^3
Galactic magnetic field: 1μG
Galactic ISM density: ≈ 1proton/cm^3 (see table in the text)
Typical energy liberated in a SN explosion:
Typical energy liberated in a GRB: 10^{51} erg
Power of AGN: 10^{41} to 10^{48} erg/s

Chandrasekhar limit: 1.4 solar masses
Eddington limit: 10^{38} ergs/s/solar mass

At the scale of the universe
Hubble constant: 71 (+4 –3) km/s/Mpc
Age of the universe: 13.7 (±0.2) Gy
Volume of the universe within the horizon: $\pi\ 10^{78}\ m^3$
Number of galaxies in the visible universe: 10^{11}
Deceleration parameter: –0.66 (±0.1)
Total energy density: 1.02 (±0.02) of critical
Visible density: 1% of critical
Normal (baryonic) matter density: 4.4% (±0.4%) of critical
= 3 protons/m^3 = 3 keV/cm^3
CMB density: 400 photons/cm^3 = 1.3 10^8 CMB photons/nucleon
CDM density: 23% (±4%) of critical = 0.1 eV/cm^3
Dark energy density: 73% (±4%) of critical
Equation of state of dark energy: $w = -1.0 \pm 0.2$
CMB temperature: 2.7 K
Velocity of the sun *wrt* CMB: 370 km/s
GZK cutoff: ½10^{20} eV
Scale above which the universe is homogeneous: 100Mpc

9.7 Some names, some dates

The first reliable information we have today on astronomy is from Egypt with a calendar (twelve months of thirty days + five days, based on the simultaneous rise of the sun and of Sirius). It is organized with 4228 BC as reference starting date. There are also evidences for significant astronomical knowledge in Babylonian, Sumerian and Chinese civilizations as early as the third millennium. In Greece, Thalès is known for having predicted (with what he had learned on the occasion of a journey in Egypt and with some luck) the solar eclipse of May 28th 585 BC. Apart from some dissidents, such as Heraclide and Aristarchos of Samos, who were preaching for a heliocentric system, the Greeks had the idea of stars moving on concentric spheres centered on the earth, the biggest one being that of the fixed stars.

Hipparcos made important and accurate astronomical observations while in Rhodes from 161 to 126 BC. He located some 800 stars precisely, giving each of these an estimated magnitude, he discovered the precession of the equinoxes and he contributed to the development of the theory of eccentrics and epicycles accounting for the observed movements of celestial bodies.

In Alexandria, following Hipparcos, Ptolemy, between 127 and 151 AD, refined and summarized the current knowledge in a monumental work, the Almagest. Up to the end of the Middle Age the Arabs have been the main contributors to the progress of astronomy, in particular with Alhazen who was born in Irak and lived in Cairo (965-1039).

With Nicolas Copernicus, born in Poland in 1473, astronomy enters a new age. As a monk, in Frauenburg, between 1509 and 1529, he makes observations which he compares with the predictions of Ptolemy's and heliocentric systems: finally, in 1543, he publishes "De Revolutionibus Orbium Caelestium" where he decides unambiguously in favor of the second. Following him, Tycho Brahe gave a new look to heliocentrism by expressing it in the earth rest frame (at least this is how we would say it today), thus elegantly overcoming the dogmatic oppositions to Copernicus' ideas. Born in Sweden in 1546, he was an excellent astronomer. He observed a supernova in 1572, established a new star catalogue, studied the moon, measured with unprecedented accuracy the movement of planets and created what may be considered as the first observatory ever built. His student in Prag, Johannes Kepler, born in Württemberg in 1571, made himself numerous observations and studied Tycho's papers, concentrating on mismatches between observation and theory in the case of Mars. This led him to publish in 1609 in his "Astronomia Nova" his first two laws, the third law following in 1619. He spent the rest of his life writing the Rudolphine Tables giving the movement of planets.

Galileo Galilei, born in Pisa in 1564, built the first observation eyeglass in the second half of year 1609 (magnification×16) with which, on January 7th 1610, he discovered Jupiter's satellites. Many other discoveries followed: the mounts of the moon, the phases of Venus, the resolution of the Milky Way into stars and that of the Cancer cluster. Convinced of the validity of Copernicus' ideas he devoted much energy to make them accepted. Following Galileo, the XVII th century witnessed a flow of new discoveries: in 1610 and 1612 the Orion nebula and Andromeda were observed for the first time; on November 7th 1631 Gassendi observed the transit of Mercury in front of the sun that had been predicted by Kepler; in 1656 Huyghens observed Titan, the first satellite of Saturn; Mira Ceti, the first variable star, was observed in 1638 and Mizar, the first double star, in 1650. The first telescopes are born in the middle of the century (Zucchi, Mersenne) and have been later improved by Gregory, Newton and Cassegrain. In 1672 and 1675 the Paris and Greenwich observatories have been established.

The end of the XVII th century is marked by the publication in 1687 by Isaac Newton of the Principia. Born in 1643, he studied in Cambridge in 1661 where he became professor in 1669. His contribution to physics is very broad. The successes of Newton's mechanics are numerous. One may quote a few among the most spectacular: In 1682, Halley understood that the comet he had been observing was in fact a re-apparition of that observed by Kepler in 1607 and by Apianus in 1531; In 1836, Le Verrier predicted the existence of Neptune, observed by Galle immediately after; In 1844, Bessel suggested that perturbations in the movement of Sirius were due to the presence of a companion, discovered in 1862 by Clark.

The XVIII th and XIX th centuries saw a large number of new observations. The 1st January 1801, Piazzi discovered Cérès, the first asteroid. The Messier catalogue had been published in Paris in 1784. Several new catalogues were established in the XVIII th and XIX th centuries (Flamsteed, Lacaille, Argelander, Gould, Auwers). In 1726 Bradley understands the stars “aberrations” as a correction for the time it takes for the light to reach us. A most notable figure of the century is William Herschel who, in 1781, discovered Uranus using a telescope he himself had built. He also discovered many double stars. He founded the technique of photometry. He separated the contribution of the peculiar movement of the sun from that of the other stars (that Halley had been first to observe in 1718). He was first to try counting stars and to draw the shape of the Milky Way and the location of the sun. Important observations continued into the XIX th century with Schwabe observing, between 1826 and 1851, the solar spots and their 22 years cycle. In 1838 Bessel measured the parallax of 61 Cygnus. During the XIX th century very detailed observation of the planets have been made.

Photography, photometry and spectrometry blossomed during the XIX th century giving observations a new boost and allowing for the establishment of new catalogues. Spectroscopy was born in 1802, with Wollaston noticing absorption lines in the solar spectrum; in 1811, Fraunhofer, in Munich, studied these lines and in 1859 Kirchhoff identified them with emission lines studied in the laboratory by Bunsen. From 1863 to 1868, Secchi, in Rome, studied and classified the spectra of 400 stars. He was followed in 1885 by Pickering who started spectrography in Harvard where, from 1918 to 1928, the spectra of 250 000 stars have been measured and the spectral classification (OBAFGKM) established. At the end of the XIX th century, the French astronomers Wolf and Rayet discovered very hot and massive stars in the O region. Between 1909 and 1920 the Danish and American astronomers Hertzsprung and Russell, independently, put order in the color-luminosity relation by replacing apparent luminosity by absolute luminosity and separating out from the main sequence the white dwarf and red giant populations.

With the XX th century, physics played the central role and astrophysics dominated astronomy. Between 1911 and 1913, Victor Hess, an Austrian physicist, intrigued by observations made by Wulf, a German Jesuit, discovered and studied cosmic rays. In 1938, Pierre Auger discovered the extensive air showers in the French Alps. In 1924, Sir Arthur Eddington wrote down the mass-luminosity relation that is named after him. In 1912 Miss Leavitt discovered the luminosity-period relation of Cepheids in the Small Magellanic Cloud.

Observations are no longer confined to the Milky Way, not to say the solar system. In the first twenty years of the century, Adams, Campbell, Charlier, Oort, Seares and Shapley measure the movement of nearby stars in the Milky Way or in the local cluster. Nebulae are now found to include spirals and in 1930, Hubble, using a spectrometer at Mount Wilson and having observed Cepheids in Andromeda, found the famous velocity-distance relation named after him. In the 1930’s Hans Bethe and Carl von Weizsäcker gave a nuclear physics description of the reactions taking place in the cores of stars. In the wake of their work, William Fowler gave in 1957 a detailed account of nucleosynthesis. In 1931 Subramanian Chandrasekhar realized that beyond 1.4 solar masses white dwarfs can no longer sustain gravity and must collapse. In 1934, Walter Baade and Fritz Zwicky, following a conjecture made by Landau two years earlier, suggested that such collapses should produce a neutron star accompanied by a supernova explosion. They searched for such events and found about a hundred. In 1942, the Swedish astronomer Hannes Alfvén, pioneer of plasma physics and of MHD, published his main results. In 1943, Carl Seyfert discovered the galaxies that have been named after him. At the same time, progress in technique, such as the Schmidt (1934) and Maxutov (1941) aberration correctors or progress in stellar interferometry, pioneered by Michelson Pease and Hale in the 1920's and 1930's, made optimal use of the very large telescopes, culminating in 1949 with the 200" at Mount Palomar.

The most revolutionary ideas occurred in cosmology. In 1916, Einstein extended special relativity to include gravitation in what he called general relativity. His equations were written for a static universe and included a cosmological constant term. As soon as they were written, Schwarzschild, a German astrophysicist, realized that very massive stars should collapse into black holes. In 1919, Eddington launched an expedition to observe a solar eclipse near the tropics and confirmed Einstein's prediction of the bending of light in the gravitational field of the sun. Between 1922 and 1924, a young Russian physicist, Alexander Friedmann, produced expanding solutions of Einstein's equations that did not require a cosmological constant. He died in 1935 of typhoid, aged 37. In 1927, Lemaître presented his ideas on the primeval atom, a day without a yesterday as he called it. The same year, Eddington identified redshifts. Twenty-five years later, Hoyle, defender of a static model together with Bondi and Gold, made fun of Lemaître ideas and talked of the "big bang", the name that has in fact survived. In 1948, George Gamov, a former Friedmann's student, published a physical model of the big bang taking nucleosynthesis in due account and predicted the existence of the CMB. The confirmation came in 1965 with the discovery by Arno Penzias and Robert Wilson of the CMB, detected as an excess noise in the horn antenna which they were setting up for Bell Labs in order to receive signals from Telstar (a TV satellite). In 1966, Sakharov explained the survival of some matter following the early annihilation of matter and antimatter. With the addition of inflation by Alan Guth in 1979, modern cosmology was born. In particular, the study of black holes has attracted the interest of numerous physicists and mathematicians such as Wheeler, Zeldovich, Ginzburg, Beckenstein, Hawking and Penrose.

To conclude this rapid and superficial survey, let us mention a few among the more recent events:

In 1967, Anthony Hewish and his student Jocelyn Bell discover the first pulsar, soon identified by Gold as being a rapidly rotating neutron star, while Goldreich, Julian and Pacini explain the formation of jets in the pulsar magnetosphere. Following his discovery of the first x-ray source in 1962, Riccardo Giacconi pioneers x-ray astronomy with the satellite Uhuru that reveals the existence of several accreting pulsars and black holes (Cyg X1). In 1974, two American astrophysicists, Joseph Taylor and Russell Hulse, discover the first binary pulsar, giving the first evidence for gravity waves.

In 1931 Jansky discovers the radio noise emitted by the center of the Milky Way, and in 1946 Hey, Parsons and Phillips discover the first radiosources in the Cygnus region. In 1949, radioastronomy was born and his founding fathers were Oort in Holland, Bolton and Stanley in Australia, Ryle and Smith in England who established the first catalogues. In 1954, in Leyde, Oort, Muller and Van de Hulst observed and studied the 21 cm line of atomic hydrogen that Van de Hulst had predicted. In 1951, the optical counterpart of Cyg A, that had been located meanwhile with improved precision by Ryle, is observed by Baade and, in 1953, Jennison and Das Gupta resolve its two radio lobes. In February 1963 Marteen Schmidt and Jesse Greenstein understand finally that the physical counterparts of such objects are highly redshifted: they have discovered quasars. Denis Sciama and Martin Rees made the observation, soon confirmed by Marteen Schmidt, that there are relatively more quasars at high redshift than there are today.

Proposed by John Mather in 1974, COBE was finally launched on 18 November 1990 and produced revolutionary measurements of the CMB. George Smoot, of the COBE team, studied the tiny CMB ripples using the Differential Microwave Radiometer.

It is also in the 1970's that the existence of dark matter, that had been previously claimed by Oort in 1932 and Zwicky in 1933, has been finally accepted by the community following the work of Vera Rubin on the velocities of stars in galaxy halos. In 1986, Bogdan Paczynski suggested the gravitational lensing effect.

In 1995 Margaret Geller and John Huchra started a deep field survey of galaxies.

For now more than thirty years, spatial astronomy and astrophysics have open new windows on the sky. A real revolution has resulted with all wave lengths having become within our reach. Successes such as those of WMAP on the CMB or the Hubble Space Telescope (HST) in the visible speak for themselves. X-ray and infrared detectors on board satellites have revealed a new and often unexpected picture of the sky. At the same time, on ground, stellar interferometry has made rapid progress, reaching resolutions of a few thousands of an arcsec in the visible. Giant arrays of radiotelescopes, gamma-ray Cherenkov (HESS, Cangaroo) or of cosmic ray detectors (Auger) have been or are being constructed. Figures 9.4 to 9.10 illustrate some of the most modern instruments used in the exploration of the universe.

Today, astrophysics is undoubtedly the field of physics that is progressing at the fastest pace. This will most probably last for many years to come.

Figure 9.4 : The VLA (Very Large Array) is an interferometer made of 27 antennas arranged in the shape of a Y located in the US (New Mexico). Each antenna has a diameter of 25 m. They cover a frequency range between 74MHz and 50GHz. At 43GHz the resolution is 0.04 seconds of arc.

Figure 9.5 : The Very Large Telescope (VLT) is a set of four telescopes, each equipped with a 8m diameter mirror, located in La Silla (Chile) and operated by ESO, the Eropean Southern Observatory. The telescopes can operate independently or in combined (interferometry) mode. The frequency covered ranges from near UV to IR (25μm.

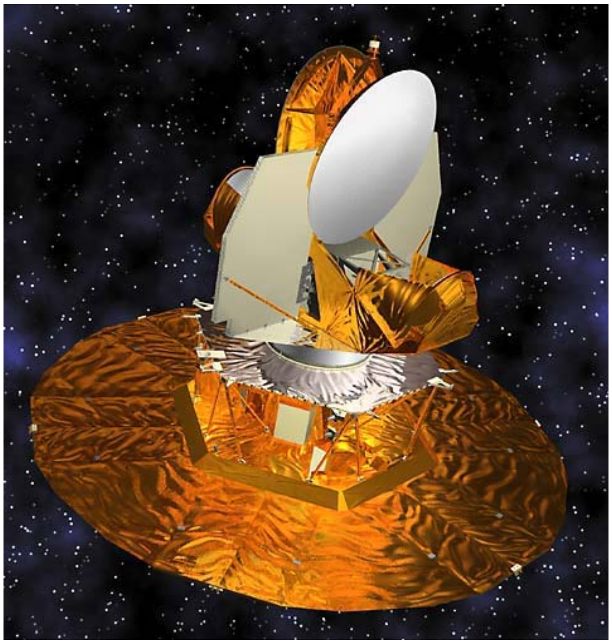


Figure 9.6 (top left) : WMAP (Wilkinson Microwave Anisotropy Probe) has a spatial resolution of 0.3º and a temperature sensitivity of 20μK over five frequencies between 22 and 90 GHz. It observes the CMB and is located at 500 000 km from earth.

Figure 9.7 (top right) : The Hubble Space Telescope (HST) operates in the visible with a two mirror arrangement of 2.4m aperture. Its effective magnification is 6000. It orbits at 600km above earth.

Figure 9.8 (bottom) ; The XMM-Newton observatory (Xray Multimirror Mission) is operated by the European Space Agency (ESA). It includes 3 grazing incidence mirrors and covers an energy range between 0.1 and 10 keV.

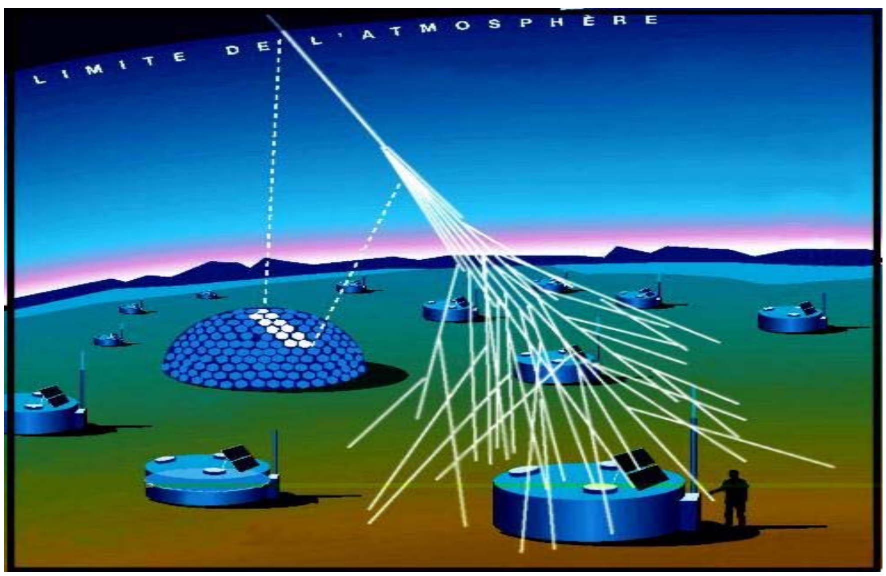


Figure 9.9 (top) : HESS (High Energy Stereoscopic System) is an array of four gamma ray telescopes in Namibia. The shower developing in the upper atmosphere produces Cherenkov light that is collected in the 382 mirrors(60cm diameter) of each telescope.
Figure 9.10 (bottom) : The Pierre Auger Observatory is an hybrid cosmic ray detector under completion in the Argentinean pampa. It includes four fluorescence eyes and 1600 water Cherenkov counters located at the nodes of an array having a mesh size of 1.5km..