

King's College London

UNIVERSITY OF LONDON

B.Sc. Third Year Optional Physics Course

CP/3630: General Relativity and Cosmology

Lecturer:

Prof. Nick E. Mavromatos
Department of Physics – Theoretical Physics Group

These notes contain a summary of the most important results and concepts covered in the course *CP/3630: General Relativity and Cosmology*, taught to the third year undergraduate students of King's College London Physics Department, during the second semester (January-March) of the academic years 2000/01 - 2007/08. They were amended in May 2008 by expanding the Cosmology section, based on lectures given to the first year graduate students (Doctorate Programme) of the department of Theoretical Physics of the University of Valencia (Spain). Only classical aspects of the theory will be covered in the lectures. It should be stressed that these notes are not meant to be substitutes of a book. The student is strongly advised to follow closely the books suggested during the course, on which the latter is based. These notes should not be distributed without the consent of the pertinent Physics Departments and/or the author.

updated: May 2008

2008 © King's College London
2008 © Dept. Física Teòrica, Univ. de València (Espanya)

1 Introduction

General Relativity is one of the most important theoretical developments of the 20th century. It is a theory about the structure and dynamics of space time itself, and its interaction with matter. Einstein's extraordinary intuition led him, in 1915, ten years after the development of Special Relativity, to suggest - something which was verified soon after by a number of important experimental measurements- that the gravitational 'force' as perceived by the Newtonian approach was incorrect, and that the correct approach was to assume that this 'force' was the result of non-zero curvature of space time, which itself was the consequence of a non-trivial mass distribution. This is the main idea behind Einstein's theory of Gravitation, the so-called General Relativity.

There are important differences from Newtonian Gravitation. For instance, a satellite orbiting around a massive body in Einstein's theory of General Relativity is floating freely, without the influence of any force, following a geodesic curve in the curved space time induced by the presence of the massive body. This is in sharp contrast to the Newtonian approach, where the inverse-square law of gravitational force characterizes the satellite motion. Moreover, General Relativity, being a 'relativistic theory, i.e. a natural extension of Special Relativity for non flat space times, shares all the novel ingredients of the latter, such as the lack of objective simultaneity of events, the existence of a limiting velocity, that of light in vacuo *etc*, which were absent in the Newtonian approach. Nevertheless, for consistency, there is a limit in which Einstein's theory reproduces partially some of the results of Newton's theory (e.g. for large distances away from the gravitational centres of attraction the orbits resemble those predicted by Newton).

Despite the 85 years that passed since Einstein put forward his famous equations, the theory of General Relativity remains a *classical theory* of the gravitational field, whose *quantum* version is still an elusive object of intense and exciting theoretical debate. This should be contrasted with the rest of the fundamental interactions in Nature (electromagnetic, weak and strong) whose quantum field theories are sufficiently developed, and confirmed by Experiment to a great extent. Nevertheless, the classical theory of Einstein's gravity has been verified by experiment to a point that no one doubts today about its validity, at least for sufficiently low energy scales that describe a big portion of the observable Universe to date. It should be noted, though, that there are still some predictions of this classical theory, namely gravitational waves, whose experimental confirmation is still lacking, and for this purpose important satellite and terrestrial experiments are currently under construction or design.

In these notes we give a summary of the most important results and concepts covered in the course *CP/3630: General Relativity and Cosmology*, taught to the third year undergraduate students of King's College London Physics Department, in the years 2001 - 2008. The notes were amended in May 2008 by expanding the Cosmology section, based on lectures given to the first year graduate students (Doctorate Programme) of the department of Theoretical Physics of the University of Valencia (Spain). Only classical aspects of the theory will be covered in the lectures. It should be stressed that these notes are not meant to be substitutes of a book. The student is strongly advised to follow closely the books suggested during the course, on which the latter is based. The notes, serve, however in guiding the student in his/her studies, and they also provide a number of exercises, covered during the tutorials, which are meant to sharpen up the critical understanding of the course material. The suggested books, which these notes follow closely are:

- (i) **B.F. Schutz**, *A first course in General Relativity* (Cambridge Univ. Press 1985)
- (ii) **E. Taylor** and **J.A. Wheeler**, *Exploring Black Holes* (Addison Wesley Longham 2000)

The more advanced student in General Relativity, such as a student in a doctorate Programme (advanced M.Sc. or first-year Ph.D.), with an interest in continuing research at a graduate and/or postdoctoral levels on General Relativity and Cosmology or Theoretical Particle Physics, might find the following books a great help:

- (iii) **C.W. Misner**, **K.S. Thorne** and **J.A. Wheeler**, *Gravitation* (Freeman 1973)
- (iv) **S. Weinberg**, *Gravitation and Cosmology* (Wiley, New York 1972)

- (v) **R. Wald**, *General Relativity* (Chicago Univ. Press 1984)
- (vi) **H. Stephani**, *General Relativity* (Cambridge Univ. Press 1985)
- (vii) **E.W. Kolb** and **M.S. Turner**, *The Early Universe* (Frontiers in Physics, Lecture Notes Series, Addison Wesley 1990)

The structure of the notes is as follows: We start with a brief description of Newtonian mechanics, especially orbits, which eventually we shall come back to, in order to understand important physical aspects and differences of the general relativistic theory of gravitation, as compared to that of Newton (for instance the precession of Mercury’s perihelion *etc.*).

Then we move onto our tour on curved space times by following a pedagogical approach. A covariant formulation of Special Relativity, including a discussion on fluid dynamics (in flat space times), is briefly presented, which will help us understand better the subsequent curved space time analysis.

Next we discuss the general principles underlying Einstein’s approach to Gravitation, namely the principle of equivalence, and we present generic physical arguments as to why Newtonian theory and Special Relativity are inadequate to provide a correct description. We also discuss ‘evidence’ for a non-trivial curvature of space time.

We then commence our study of General Relativity by first covering some formal definitions and concepts: curves in arbitrary metric space times, parallel transport along a curve, covariant derivative, geodesics, and Riemann curvature) in an attempt to quantify the most important concepts encountered in Einstein’s theory of gravitation.

In parallel, some notes on tensor calculus will be distributed, which also help the student assimilate the mathematics underlying the covariant formalism. This formalism facilitates considerably, and also is essential for a complete understanding of concepts and methods used in the analysis of curved space times.

Einstein’s equations for weak fields follow, where we discuss, as an important physical application, the generation and detection of gravitational waves, one of the most important predictions of General Relativity, which still lacks experimental confirmation.

We then proceed to discuss (rather briefly) an exact solution of Einstein’s equation, the Schwarzschild solution, which describes the space time in the exterior of spherical bodies of non zero mass. Such bodies include: Earth, Stars, and (non rotating) Black Holes to a good approximation. Detailed studies of orbits (geodesics) in such space times are given, with emphasis in the main differences from the Newtonian approach. In this context, we also discuss important physical predictions of General Relativity related to the behaviour of clocks in gravitational fields (e.g. different running of clocks depending on the altitude), gravitational redshift of photons, and bending of light near a massive body (e.g. deflection of light by the Sun and gravitational ‘lensing’). All these predictions have been verified by terrestrial and astrophysical experiments. Especially the Gravitational Lensing constitutes by now one of the most important techniques used in Astrophysics in order to receive information on distant celestial objects that could not be possible otherwise.

In the final part of the notes we discuss cosmological solutions to Einstein’s equations, covering the most important aspects of an expanding Universe (Friedmann–Robertson–Walker) solution. Some aspects of physical modern cosmology are also discussed, with emphasis on Astro-particle Physics issues (dark matter in the Universe and astrophysical tests of particle physics models, such as supersymmetry, recent ‘evidence’ for a non zero cosmological ‘constant’ or rather dark energy, with a detailed analysis of the pertinent astrophysical measurements used to uncover the Universe energy budget). In this last topic I pay particular emphasis on the strong underlying-theoretical-model dependence of the interpretation of the observations.

2 Newtonian Mechanics and Theory of Gravitation - A Brief Review

2.1 Newton's Laws

These can be summarized as follows:

1. Free particles move with constant velocity (i.e. constant speed along straight lines, and no acceleration).
2. The acceleration of a particle is proportional to the resultant force acting on it, with the constant of proportionality being the inverse of its mass:

$$\vec{F} = m\vec{a}. \quad (2.1)$$

Alternatively, since the momentum is $\vec{p} = m\vec{v}$, this law can be written

$$\vec{F} = \frac{d}{dt}\vec{p}. \quad (2.2)$$

3. The forces of action and reaction are equal in magnitude and opposite in direction.
4. LAW OF GRAVITATION: the gravitational force exerted by a body of mass m_B on a body of mass m_A when their separation is \vec{r}_{AB} is given by the celebrated inverse-square law:

$$\vec{F}_{AB}^{\text{grav}} = G_N \frac{m_A m_B}{r_{AB}^2} \hat{r}_{AB} = m_A \vec{g}, \quad (2.3)$$

where the $G_N = 6.673 \times 10^{-11} \text{m}^3 \text{kg}^{-1} \text{s}^{-2}$ is Newton's universal gravitational constant, $\hat{r}_{AB} = \vec{r}_{AB}/|\vec{r}_{AB}|$ is the unit vector along the direction between A and B; \vec{g} denotes the acceleration due to gravity.

2.2 Digression on Units

Throughout we shall use a special system of units. We take $G_N = c = 1$. The second of these implies that

$$1 = 3 \times 10^8 \text{ms}^{-1} \quad \Rightarrow \quad 1\text{s} \equiv 2.998 \times 10^8 \text{m}.$$

When $G_N = 1$ also, this implies that

$$1\text{kg} \equiv 7.424 \times 10^{-28} \text{m}. \quad (2.4)$$

Three reasons for using this system of units:

1. Elegance: equation (2.4) connects mass and geometry elegantly.
2. Convenience: a way to get rid of factors of G_N and c ; after all there seems no point in working in a system of units where the two fundamental constants have the ludicrous numerical values of $c = 3.0 \times 10^8 \text{ms}^{-1}$ and $G_N = 6.7 \times 10^{-11} \text{m}^3 \text{kg}^{-1} \text{s}^{-2}$.

2.3 Newtonian Mechanics: Orbiting

Consider a particle of mass m , moving under the influence of a central force

$$f_r = -\frac{k}{r^2} = -\frac{d}{dr}V_{\text{pot}}, \quad (2.5)$$

which can be derived from the *standard potential*

$$V_{\text{pot}} = -\frac{k}{r}. \quad (2.6)$$

The motion is planar. The Lagrangian is given by (a superior point denotes differentiation with respect to time)

$$\mathcal{L} \equiv T - V_{\text{pot}} = \frac{m}{2}(\dot{x}^2 + \dot{y}^2) + \frac{k}{r} \quad (2.7)$$

Consider instead plane polar coordinates (r, θ) given by

$$\begin{aligned} x &= r \cos \theta, & \dot{x} &= \dot{r} \cos \theta - r \dot{\theta} \sin \theta, \\ y &= r \sin \theta, & \dot{y} &= \dot{r} \sin \theta + r \dot{\theta} \cos \theta, \end{aligned}$$

whence

$$\dot{x}^2 + \dot{y}^2 = \dot{r}^2 + r^2 \dot{\theta}^2, \quad (2.8)$$

and the Lagrangian in plane polar coordinates is given by

$$\mathcal{L} = \frac{m}{2}(\dot{r}^2 + r^2 \dot{\theta}^2) + \frac{k}{r}. \quad (2.9)$$

The Euler–Lagrange equation (c.f. Appendix A, for a general derivation and basic concepts) for θ is

$$\frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{\theta}} - \frac{\partial \mathcal{L}}{\partial \theta} = 0 \quad \Rightarrow \quad \frac{d}{dt} (mr^2 \dot{\theta}) = 0, \quad (2.10)$$

from which follows conservation of angular momentum L

$$\frac{L}{m} = r^2 \dot{\theta} \equiv r^2 \omega_{\theta}. \quad (2.11)$$

The Euler–Lagrange equation for r reads

$$\frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{r}} - \frac{\partial \mathcal{L}}{\partial r} = 0 \quad \Rightarrow \quad m\ddot{r} - mr\dot{\theta}^2 + \frac{k}{r^2} = 0. \quad (2.12)$$

From the conservation of angular momentum (2.11) we can express $\dot{\theta}$ in terms of the conserved quantity L/m :

$$m\ddot{r} - mr \frac{(L/m)^2}{r^4} + \frac{k}{r^2} = 0 \quad \Rightarrow \quad m\ddot{r} - m \frac{(L/m)^2}{r^3} + \frac{k}{r^2} = 0. \quad (2.13)$$

For the case of Newtonian gravitation (in units where $G_N = c = 1$) we have $k = Mm$ where M is the mass of the gravitational attractor. Multiplying equation (2.13) above by \dot{r} we have

$$m\dot{r}\ddot{r} = m\dot{r} \frac{(L/m)^2}{r^3} - \dot{r} \frac{Mm}{r^2} \quad (2.14)$$

Hence

$$\frac{d}{dt} \left(\frac{1}{2} m \dot{r}^2 \right) = m \frac{d}{dr} \left(-\frac{(L/m)^2}{2r^2} \right) \dot{r} + \frac{d}{dr} \left(\frac{Mm}{r} \right) \dot{r}. \quad (2.15)$$

From the chain rule of differentiation

$$\dot{r} \frac{d}{dr} g(r) = \frac{d}{dt} (g(r)),$$

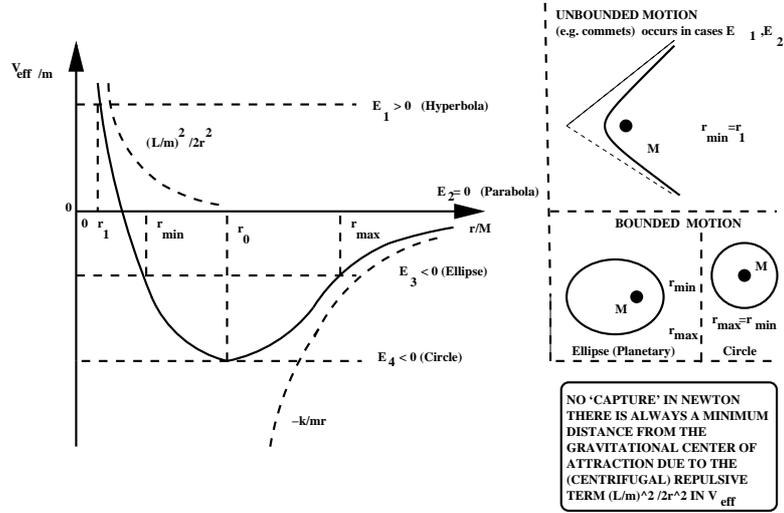


Figure 1: Effective potential and the associated orbits in the Newtonian theory.

we can rewrite equation (2.15) as

$$\frac{d}{dt} \left(\frac{m}{2} \dot{r}^2 + m \frac{(L/m)^2}{2r^2} - \frac{Mm}{r} \right) = 0. \quad (2.16)$$

Therefore we have another conservation law, this time the conservation of total energy:

$$E = \frac{m}{2} \dot{r}^2 + m \frac{(L/m)^2}{2r^2} - \frac{Mm}{r} = \text{const.} \quad (2.17)$$

This leads us to the notion of an EFFECTIVE POTENTIAL:

$$\frac{1}{2} \left(\frac{dr}{dt} \right)^2 = \frac{E}{m} - \frac{V_{\text{eff}}}{m}. \quad (2.18)$$

In words, the effective potential is whatever you have to subtract from the total energy to leave the square of the radial velocity. In our case we have

$$\frac{V_{\text{eff}}}{m} = -\frac{M}{r} + \frac{(L/m)^2}{2r^2}. \quad (2.19)$$

You can see that the effective potential is the original potential modified by a term arising from angular momentum conservation. The effective potential is a useful tool which allows us to map the problem of finding the orbit to a one-dimensional effective problem (the angular problem has been solved by incorporating the effects of angular momentum conservation). Indeed by sketching the effective potential (2.19) we may make a general classification of the various Newtonian orbits (see figure 1).

Depending on the value of the constant total energy we may classify the motion as follows:

1. Bounded motion:
 - (a) elliptical orbits
 - (b) circular orbits
2. Unbounded motion:

- (a) parabolic orbits
- (b) hyperbolic orbits

Note that due to the repulsive term proportional to $(L/m)^2$ (the centrifugal term) in the effective potential which dominates for small r , there is no *capture* in this model.

A general expression for the equation for a conic, i.e. a general Newtonian orbit

$$\frac{1}{r} = \frac{Mm^2}{L^2} \left(1 + \sqrt{1 + \frac{2EL^2}{m^3M^2}} \cos(\theta - \theta_0) \right), \quad (2.20)$$

where θ_0 is a constant of integration (the starting angle, or angle at time zero). The quantity $\sqrt{1 + 2EL^2/m^3M^2}$ is called the *eccentricity* and can be used to classify the motion as follows:

1. $e = 0$ corresponds to $E = -m^3M^2/2L^2$; motion is a CIRCLE
2. $0 < e < 1$ corresponds to $-m^3M^2/2L^2 < E < 0$; motion is an ELLIPSE
3. $e = 1$ corresponds to $E = 0$; motion is a PARABOLA
4. $e > 1$ corresponds to $E > 0$; motion is a HYPERBOLA

2.4 Perihelion Precession of Mercury: Newtonian Version

Approximate the angular motion of Mercury as if it were a circle of average radius r_0 . This corresponds to the minimum of the effective potential:

$$\frac{d}{dr} \frac{V_{\text{eff}}}{m} = 0 \quad \Rightarrow \quad -\frac{M}{r_0^2} + \frac{(L/m)^2}{r_0^3} = 0. \quad (2.21)$$

For Mercury, $M = M_\odot$, m is the mass of Mercury, and we have (in units where $G_N = c = 1$)

$$r_0 = \frac{(L/m)^2}{M} \quad \Rightarrow \quad \frac{L}{m} = \sqrt{Mr_0}. \quad (2.22)$$

The quantities M_\odot , m and r_0 are known experimentally.

Now we use the fact that the amplitude of radial motion of Mercury is small enough so that the effective potential looks like a parabola to an excellent approximation near the minimum. (See figure 2). The parabolic approximation implies, near the minimum of the potential, a harmonic oscillation in the radial direction

$$\frac{V_{\text{eff}}}{m} \simeq \frac{1}{2} \omega_r^2 r^2, \quad (2.23)$$

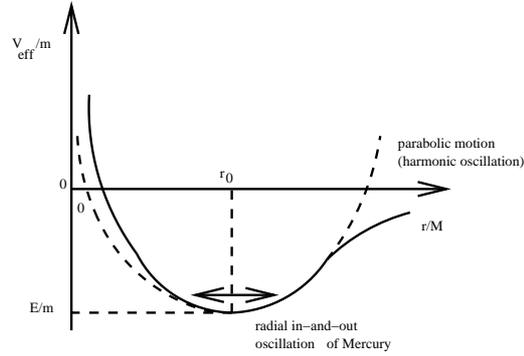
where ω_r is the frequency of radial oscillations. We immediately obtain

$$\frac{d^2}{dr^2} \left(\frac{V_{\text{eff}}}{m} \right) = \omega_r^2. \quad (2.24)$$

From conservation of angular momentum we have that the frequency, ω_θ , of angular motion is obtained as follows

$$\dot{\theta} = \omega_\theta \quad \Rightarrow \quad \frac{L}{m} = r^2 \frac{d\theta}{dt} \equiv r^2 \omega_\theta \quad \Rightarrow \quad \omega_\theta = \frac{(L/m)}{r^2}. \quad (2.25)$$

The possible advance of the perihelion, if Newton was right, would have come from a difference $\omega_\theta - \omega_r > 0$ so that during a complete period of angular motion (rotation) the radial oscillation is not complete and the perihelion precesses.



APPROXIMATE COMPUTATION OF MERCURY'S PERIHELION PRECESSION (IF ANY) IN NEWTONIAN MECHANICS

NET RESULT: NO PRECESSION IF ONLY SUN-MERCURY INTERACTION CONSIDERED; UNPERTURBED ORBIT IS AN ELLIPSE



Figure 2: Approximate computation of the perihelion precession.

Now in the Newtonian case we have from equations (2.19) and (2.24) at $r = r_0$ (i.e. substituting for L/m from equation (2.22))

$$\left. \frac{d^2 V_{\text{eff}}}{dr^2} \right|_{r=r_0} = -\frac{2M}{r_0^3} + \frac{3(L/m)^2}{r_0^4} = -\frac{2M}{r_0^3} + \frac{3M}{r_0^3} = \frac{M}{r_0^3}, \quad (2.26)$$

so that

$$\omega_r = \sqrt{\frac{d^2 V_{\text{eff}}}{dr^2}} = \sqrt{\frac{M}{r_0^3}}. \quad (2.27)$$

Now from equation (2.25) we also have $\omega_\theta = M^{1/2}/r_0^{3/2}$ and therefore in the Newtonian model

$$\omega_r = \omega_\theta, \quad (2.28)$$

and there is no perihelion precession if one considers only the Mercury-Sun interaction, ignoring the effects of the other planets. However, their effects are far too small to generate the observed precession of the orbit of Mercury.

The effective potential method is a very useful tool, which as we shall see, can be used intact in the generally relativistic case in order to get a precessing orbit for Mercury which is very close to the observed value.

Exercise 2.1 Assume that general relativistic corrections yield the following form for effective potential:

$$\frac{U_{\text{eff}}}{m} = \frac{1}{2} - \frac{M}{r} + \frac{(L/m)^2}{2r^2} - \frac{M(L/m)^2}{r^3}. \quad (2.29)$$

1. Find for which $r = r_0$ the minimum of the potential occurs and express L/m in terms of M and r_0 .

2. Show that

$$\left. \frac{d^2(U_{\text{eff}}/m)}{dr^2} \right|_{r=r_0} = \omega_r^2 = \frac{M(r_0 - 6M)}{r_0^3(r_0 - 3M)}, \quad (2.30)$$

where the symbols have the meanings explained in the lectures.

3. By assuming that the conserved angular momentum in this case is given by $L/m = r^2 d\theta/d\tau$, show that

$$\omega_\theta^2 = \left(\frac{d\theta}{d\tau} \right)^2 = \frac{M}{r_0^2(r_0 - 3M)}, \quad (2.31)$$

where think of the τ as the time (as we shall see, in GR this is actually a universal time upon which all observers agree).

4. Consider the case in which ω_θ is close enough to ω_r so that $\omega_\theta + \omega_r \simeq 2\omega_\theta$. With this approximation show that

$$\omega_\theta - \omega_r \simeq \frac{3M}{r_0} \omega_\theta. \quad (2.32)$$

5. What do you conclude from this analysis? Sketch the associated orbit. In what limit does one obtain the Newtonian situation of (almost) zero precession?

6. How many revolutions around the sun does Mercury make in 100 Earth-years? How many degrees of angle are traced out by Mercury in one century?

7. From equation (2.32) and taking into account that all of the ω -expressions have the form $d[\text{angle}]/d\tau$, it immediately follows (why?) that

$$\begin{aligned} \begin{pmatrix} \text{predicted} \\ \text{precession} \\ \text{angle} \end{pmatrix} &= \begin{pmatrix} \text{total angle} \\ \text{of orbital} \\ \text{motion} \end{pmatrix} - \begin{pmatrix} \text{total phase angle} \\ \text{of radial} \\ \text{motion} \end{pmatrix} \\ &= \frac{3M}{r_0} \begin{pmatrix} \text{total angle} \\ \text{of orbital} \\ \text{motion} \end{pmatrix}. \end{aligned} \quad (2.33)$$

From this, compute the predicted perihelion advance of Mercury in degrees per century.

The period of Mercury's orbit is 7.60×10^6 seconds, and that of the Earth is 3.16×10^7 seconds. The mass of the sun is $M = M_\odot = 1.48 \times 10^3$ metres, and the average radius of Mercury's orbit is $r_0 = 5.80 \times 10^{10}$ metres.

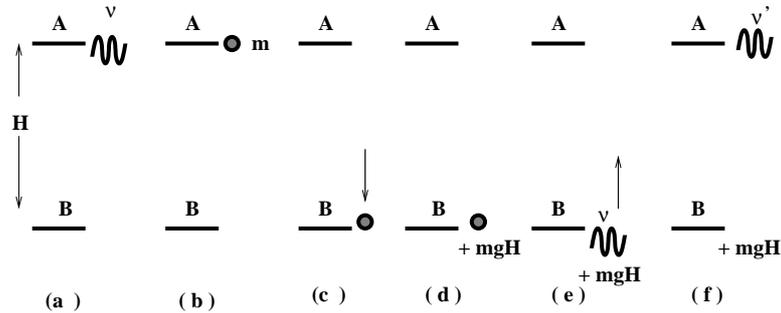
2.5 Newton Predicts Gravitational Redshift

Despite its failure to account for the precession of the perihelion, one can still use a Newtonian approach, augmented with a taste of quantum mechanics, to predict that the frequency of photons is shifted depending on the altitude.

To see this one recalls a basic formula from quantum mechanics: the quantization of the photon energy:

$$E = h\nu \quad (2.34)$$

where $h = 6.626 \times 10^{-34} \text{m}^2 \text{kg s}^{-1}$ is Planck's constant and ν is the frequency of the photon. Consider the following simplified version of an experiment carried out by Pound–Rebka in 1960 and improved by Pound–Snider in 1965, sketched in figure 3. Two observers, A and B, are separated by a height H , assumed *sufficiently small* so that the approximation of uniform gravitational



SIMPLIFIED VERSION OF THE POUND-REBKA-SNIDER
EXPERIMENT TO DETECT GRAVITATIONAL REDSHIFT

Figure 3:

acceleration is valid. Observer A emits a photon of frequency ν , “converts” it (somehow) to a particle of mass m and sends it down to observer B. The latter observer stores the energy mgH that the particle gains on descending the height H as a result of Newtonian gravitation, converts the particle back into a photon of the same energy, and hence the same frequency ν , and sends it back to observer A. The latter receives a photon of a different frequency, ν' , determined by energy conservation:

$$h\nu = h\nu' + mgH = h\nu' + \frac{EgH}{c^2} = h\nu' + \frac{h\nu gH}{c^2}, \quad (2.35)$$

whence the change in frequency is

$$\frac{\delta\nu}{\nu} \equiv \frac{\nu' - \nu}{\nu} = -\frac{gH}{c^2}. \quad (2.36)$$

Notice that since the frequency ν' is smaller than ν , this is called a RED SHIFT, because red light has the smallest frequency of the observable spectrum. This formula can be derived rigorously in the framework of general relativity, but it is surprising that Newtonian gravitation (augmented with quantum mechanical concepts) gives the correct prediction.

Despite this success, as we have seen in the lectures, Newton’s laws failed to account for the observed precession of the perihelion of Mercury which implied the need for a new theory. This is what Einstein did in the *annus mirabilis* (year of wonders) 1905, when he developed Special Relativity, but this theory, as it stood, was incompatible with Newtonian Gravitation. This troubled Einstein a lot, since he wanted to find a consistent way of reconciling gravity with Special Relativity. It took him ten more years of intense research, until he finally came up with the Theory of General Relativity, by publishing his famous equations describing the interaction of matter with gravity in 1915. Ever since our view on the world was bound to change. We shall follow this historical path during the lectures, and for this we need first to develop the mathematics appropriate to general relativity in the context of Einstein’s special theory. This is called covariant formalism or *tensor calculus*. We start by rehearsing some basic concepts in special relativity.

3 Special Relativity Primer

3.1 Historical Note

Special Relativity (Einstein, 1905) grew out of a need to understand some strange properties of Maxwell's theory of Electromagnetism, for instance the squashing of the electric field of a moving atom with a charged nucleus, in the direction of motion. Indeed, solving Maxwell's equations for the electric field of an atom moving with a velocity \vec{v} , you find that it is squashed in the direction of motion by a factor

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}}, \quad (3.1)$$

relative to the (spherically symmetric) electric field of the atom at rest. whilst if one solves these equations for the time an electron takes to orbit around the nucleus in the atom, one finds that it is enlarged by the factor γ relative to the period of orbit T in the case where the atom is at rest.

In Einstein's relativity these properties characterize all moving bodies, and not just the case of Maxwell's equations, and are known as *length contraction* and *time dilation* respectively. We shall not cover Special relativity in detail in these notes. Instead we shall only give a brief summary of the main concepts and formulas that we shall make use of in our approach to General Relativity. Besides, as already mentioned, we shall approach Special Relativity from a covariant view point, which will allow us to move on into the general relativistic framework in a more-or-less straightforward manner.

3.2 Invariant Intervals and World Lines

Points in a relativistic space time with coordinates (t, x, y, z) represent "events".

The invariant wristwatch (proper) time separation $\delta\tau$ between two events located at $(0, 0, 0, 0)$ and (t, x, y, z) is given by

$$\Delta^2 \equiv \delta\tau^2 = t^2 - x^2 - y^2 - z^2 \equiv t^2 - \ell^2 \quad (3.2)$$

for a *timelike* spacetime interval ($t^2 > \ell^2$), where t is the temporal separation in a frame, and $\ell^2 \equiv x^2 + y^2 + z^2$ is the spatial separation in the same frame. Everybody agrees on this wristwatch (proper) time between two events. Note that t in meters is equal to ct with t in seconds.

The invariant proper distance between two events is given by

$$\delta\sigma^2 = x^2 + y^2 + z^2 - t^2 \equiv \ell^2 - t^2 \quad (3.3)$$

for a *space-like* interval ($\ell^2 > t^2$), using the same notation as above. Again, every observer agrees on the proper distance between events.

Finally, events which lie on the surface $\Delta^2 = 0$ are called *light like*. This surface defines a cone, the *light cone*, separating 'future' ($\Delta^2 > 0, \delta\tau > 0$) and 'past' ($\Delta^2 > 0, \delta\tau < 0$) events, lying inside the cone, from 'events lying elsewhere' ($\Delta^2 < 0$), i.e. outside the light cone (see figure 4).

Pictorially, the various paths of a particle in a relativistic space time diagram (**world lines**) are represented as in figure 4.

As we observe from the figure, a time-like (space-like) path *always* lies inside (outside) the light cone, whilst the world-line of a photon, always lies on the light cone.

Exercise 3.1 *If two events are separated by a time-like interval show - by means of space time diagrams- that,*

1. *there exists a Lorentz frame in which they happen at the same point in space.*
2. *in no Lorentz frame they are simultaneous*

Exercise 3.2 *If two events are separated by a space-like interval show - by means of space time diagrams- that,*

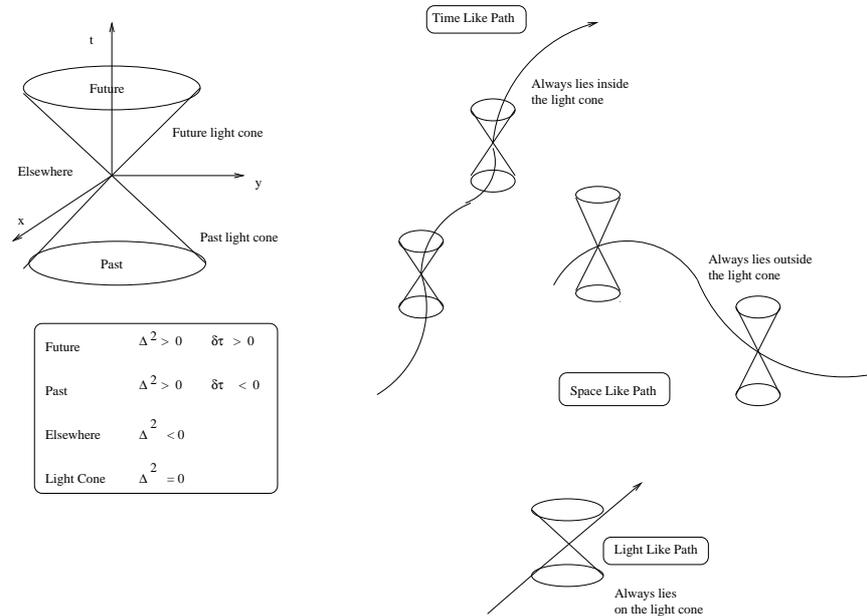


Figure 4: Light Cone, and Various types of world lines in Special Relativity

1. there exists a Lorentz frame in which they are simultaneous.
2. in no Lorentz frame do they occur at the same point in space.

3.3 Principle of Extremal Aging and Conservation Laws

The familiar from the course on Special Relativity *Twin Paradox* leads to a very important concept, that of *Natural Motion* of a relativistic body. Recall that in the twin paradox, one of them stays on Earth, whilst the other *identical* twin travels with her spaceship to a distant constellation, and returns back to Earth. At her great surprise, since she did not know any relativity and, hence, ignored the induced effects of time dilation, she discovers that she is no longer identical with her twin sister on earth, whom she found to have *aged* considerably compared with herself.

The question is which one of the twins had a *natural* motion. According to Newton this can be easily answered. His first Law would say that the twin at rest tends to remain at rest. So for Newton the twin who stayed at her home on earth is the one who moves in a *natural* way. In fact this twin has a natural motion from the point of view of *any* observer who moves at a *constant* speed with respect to the earth frame (assumed inertial and at rest, i.e. ignoring rotation of earth for the purposes of this problem). For such a moving observer the twin who stayed at her home on earth would look moving along a straight line, with constant speed, and hence according to Newton's first Law, she will continue remaining in this natural state.

In contrast, the twin who traveled into space, was required to change her state of motion, by stopping the spaceship at the distant galaxy, and then turning it around to start the return trip to earth. This second twin moved in an *unnatural* way. The fact that special relativity predicts time dilation, in other words that time run slow as measured by the wristwatch of the twin in the spaceship (proper time), as compared with the time measured by clocks on earth, is actually a *defining* property of *natural motion*, which can be stated by means of a principle that we formulate below.

Principle of Extremal Aging: The path a free object takes between two events

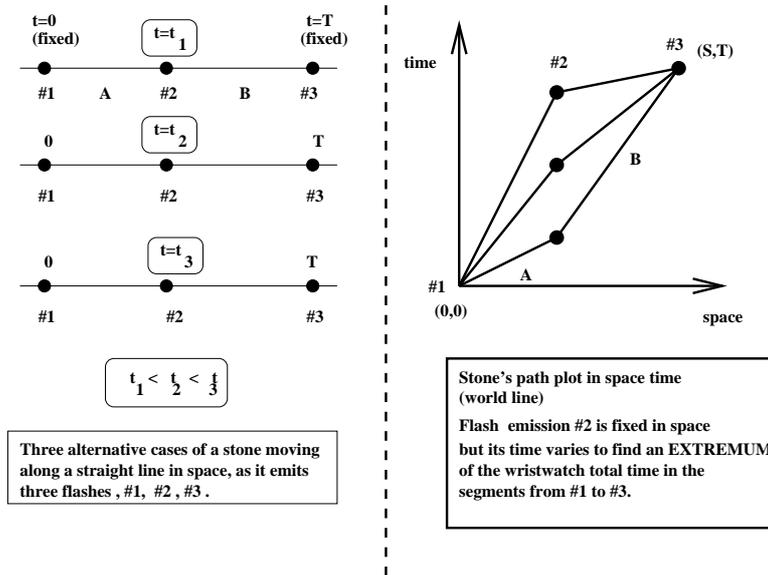


Figure 5: Towards an understanding of the principle of extremal aging in relativity. When the intermediate time t yields an extremal proper time, then, the pertinent world line (in the spacetime diagram) connecting the events #1, #2 and #3 is a straight line.

in spacetime is the path for which the time lapse between these events, recorded on the object's wristwatch (the proper time) is an EXTREMUM.

Note that extrema include both maxima and minima: most of the cases in general relativity are actually maxima. In the case of twins, and for all of the problems that we shall encounter during the course, the proper time is actually at a *maximum* for natural motion, but there are some special cases of natural motion in space time in which it can be at a minimum.

The principle of energy and momentum conservation can be derived from the principle of extremal aging: Consider the simple situation depicted in figure 5, where a stone, moving with constant speed with respect to an inertial observer, emits three flashes of light, #1, #2, #3. We consider the three cases indicated in the figure. The flashes #1 and #3 are emitted at fixed times, with respect to the inertial observer, whilst the emission time of the flash #2 is allowed to vary, but occurs in each case at the same position in space relative to the inertial observer.

Segment A: $\tau_A = (t^2 - \ell^2)^{\frac{1}{2}}$, where ℓ denotes the (one-dimensional) spatial separation between the pertinent events. Take derivative with respect to t :

$$\frac{d\tau_A}{dt} = \frac{t}{(t^2 - \ell^2)^{\frac{1}{2}}} = \frac{t}{\tau_A}.$$

Segment B: $\tau_B = ((T - t)^2 - (S - \ell)^2)^{\frac{1}{2}}$. Take derivative with respect to t :

$$\frac{d\tau_B}{dt} = -\frac{T - t}{((T - t)^2 - (S - \ell)^2)^{\frac{1}{2}}} = -\frac{T - t}{\tau_B}.$$

The total time as measured on the wristwatch is $\tau = \tau_A + \tau_B$, and the principle of extremal aging tells us that

$$\frac{d\tau}{dt} = 0 \Rightarrow \frac{d\tau_A}{dt} + \frac{d\tau_B}{dt} = 0 \Rightarrow \frac{t}{\tau_A} = \frac{T - t}{\tau_B} \quad (3.4)$$

whence

$$\frac{t_A}{\tau_A} = \frac{t_B}{\tau_B} \quad (3.5)$$

and indeed for an arbitrary number of partitions of the segment

$$\frac{t_A}{\tau_A} = \frac{t_B}{\tau_B} = \frac{t_C}{\tau_C} = \dots \quad (3.6)$$

Hence from the principle of extremal aging we have

$$\frac{t}{\tau} = \text{a constant of (relativistic) motion}, \quad (3.7)$$

To determine what this quantity represents physically we express the proper time τ in terms of the temporal and spatial coordinate separation of the pertinent events:

$$\frac{t}{\tau} = \frac{t}{(t^2 - \ell^2)^{\frac{1}{2}}} = \frac{t}{t(1 - (\ell/t)^2)^{\frac{1}{2}}} = \frac{1}{(1 - v^2)^{\frac{1}{2}}} = \frac{E}{m}, \quad (3.8)$$

so the constant is the energy per unit mass. Similarly one can prove the conservation of momentum,

$$\frac{\ell}{\tau} = \frac{v}{(1 - v^2)^{\frac{1}{2}}} = \frac{p}{m} = \text{constant}. \quad (3.9)$$

Exercise 3.3 *By following similar steps as above leading to the conservation of energy, i.e. by making use of the principle of extremal aging, but this time taking derivatives with respect to the (intermediate) space separation ℓ , instead of the time separation t , prove (3.9), that is to say spatial momentum is conserved in a relativistic motion.*

In general, if a particle changes speed, a useful concept is the instantaneous speed, for which we have just found a convenient differential form:

$$\frac{E}{m} = \frac{dt}{d\tau}, \quad \frac{p}{m} = \frac{d\ell}{d\tau}. \quad (3.10)$$

Note that these are all dimensionless quantities.

3.4 Invariant Mass

The infinitesimal separation between two events is

$$(d\tau)^2 = (dt)^2 - (d\ell)^2. \quad (3.11)$$

If we divide through by $(d\tau)^2$ and multiply through by m^2 , a constant, we have

$$m^2 = m^2 \left(\frac{dt}{d\tau} \right)^2 - m^2 \left(\frac{d\ell}{d\tau} \right)^2 \Rightarrow m^2 = E^2 - p^2. \quad (3.12)$$

So the mass is an invariant about which every observer agrees. One may actually use covariant notation to rewrite the above expressions. We shall do this in what follows. This will be essential for our discussion of curved spacetimes. First we need to define some concepts, which we do in the next subsection, as well as in the notes on tensor calculus.

Exercise 3.4 *Show that the space-time path of a massive particle is always time like, whilst that of a massless particle is always light like.*

3.5 Covariant Formulation of Special Relativity

Spacetime is viewed as a four-dimensional metric space, with coordinates $x^\mu = (x^0, x^1, x^2, x^3)$ where $x^0 = ct$ but in our system of units, where $c = 1$, we shall not distinguish between x^0 and t from now on. In the framework of special relativity, spacetime is defined by the invariant line element

$$ds^2 = -d\tau^2 = \eta_{\mu\nu} dx^\mu dx^\nu = -(dt)^2 + (dx^1)^2 + (dx^2)^2 + (dx^3)^2, \quad (3.13)$$

where $\eta_{\mu\nu}$ is the *Minkowski metric*, given in matrix form by

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \equiv \text{diag}(-1, 1, 1, 1). \quad (3.14)$$

In fact $\eta^{\mu\nu}$ in this case is identical to $\eta_{\mu\nu}$. A space time corresponding to the line element (3.13) is *flat*. As we shall see later on, this implies *zero curvature*. Special Relativity is a theory describing kinematics and dynamics in flat Minkowski space times.

The coordinate transformations of Special Relativity (Lorentz transformations) form a group called the *Lorentz group*, which consists of rotations and boosts. These transformations leave the element Δ^2 (or its infinitesimal counterpart ds^2) *invariant*.

We can see that explicitly in the instructive example of a boost of velocity v along the x -direction, say. The transformed coordinates (denoted by a bar) are related to the initial coordinates as follows (in units where $c = 1$):

$$\bar{t} = \gamma(t - vx), \quad \bar{x} = \gamma(x - vt), \quad \bar{y} = y, \quad \bar{z} = z \quad (3.15)$$

In tensor form this transformation is simply $\bar{x}^\mu = \Lambda^\mu{}_\nu x^\nu$ where the Lorentz boost (3.15) is represented as a 4×4 matrix:

$$\Lambda^\mu{}_\nu = \begin{pmatrix} \gamma & -\gamma v & 0 & 0 \\ -\gamma v & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (3.16)$$

Notice that such transformations are a special case of the four vector transformation under a change of coordinates, familiar from our tensor calculus part of the course, $x^\mu \rightarrow (\partial \bar{x}^\mu / \partial x^\nu) x^\nu$. As we shall see later on, these transformations are then a special case of general coordinate transformations which leave not only the ds^2 element invariant, but also the Minkowski metric $\eta_{\mu\nu}$ invariant.

Pictorially, the effect of a Lorentz transformation (say the boost (3.15) for definiteness) is represented in a spacetime diagram by a squashing of the axes as indicated in figure (6). Note carefully that in this transformation the position of the lightcone is unchanged, as it lies half way between the t and x axes (and also the \bar{t} and \bar{x} axes). This reflects the underlying postulate of special relativity that the speed of light is an absolute constant.

Exercise 3.5 *Explain carefully figure 6.*

The *four-velocity* is a contravariant vector which is defined by an appropriate covariantization of the concept of velocity in Gallilei–Newton (G–N) mechanics. In G–N mechanics, time was a universal quantity from which the velocity was defined as $d\vec{x}/dt$; however in special relativity as we have mentioned, the coordinate time is observer-dependent, and the rôle of a universal time on which all observers agree is played by the proper time, τ . This prompts us to define the four-velocity as the contravariant four-vector

$$u^\mu = \frac{dx^\mu}{d\tau}. \quad (3.17)$$

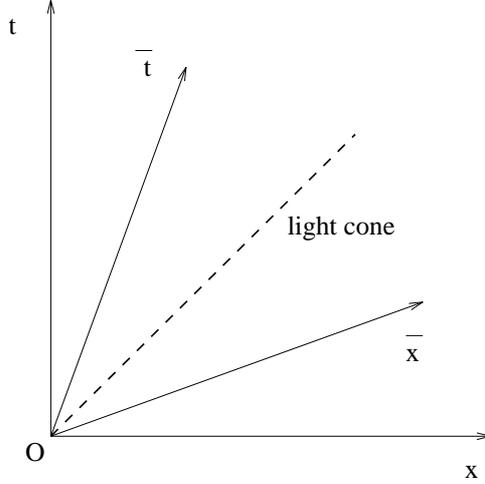


Figure 6: The effect of a Lorentz transformation on a spacetime diagram in special relativity.

The components of the four-velocity are given by

$$\begin{aligned}
 u^0 &= \frac{dt}{d\tau} = \frac{1}{\sqrt{1-v^2}} \equiv \gamma \\
 u^i &= \frac{dx^i}{d\tau} = \frac{dt}{d\tau} \frac{dx^i}{dt} = \gamma v^i, \quad i \in \{1, 2, 3\},
 \end{aligned}
 \tag{3.18}$$

where γ is called the Lorentz factor, and $v^i = dx^i/dt$ is the three-velocity.

Exercise 3.6 Show that $u^\mu u_\mu = -1$.

Following naturally from this definition is that of the *four-momentum*, vector p^μ which is defined by analogy with its G–N counterpart:

$$p^\mu = m u^\mu = (m\gamma, m\gamma v^i) = (E, p^i) \tag{3.19}$$

where E and p^i are respectively the relativistic energy and momentum: note that these are NOT simply the G–N energy and three-momentum. In the non-relativistic limit $v \ll c = 1$, they reduce to the sum of the rest energy and the G–N kinetic energy, and three-momentum respectively, as can be seen by expanding the factors of $\gamma \simeq 1 + v^2/2 + \dots$

The four-acceleration is, by direct analogy with G–N, defined by

$$a^\mu = \frac{du^\mu}{d\tau} = \frac{d^2 x^\mu}{d\tau^2}. \tag{3.20}$$

The four-force is then related to the four-acceleration in the usual way:

$$F^\mu = m a^\mu = \frac{dp^\mu}{d\tau}. \tag{3.21}$$

Exercise 3.7 Find the components of the four-acceleration a^μ and the four-force F^μ .

From the above covariant generalizations, we are in a position to rewrite the energy conservation equation (3.12) in a covariant way:

$$p^\mu p_\mu = \eta_{\mu\nu} p^\mu p^\nu = -E^2 + (p^i)^2 = -m^2. \tag{3.22}$$

The relativistic law of addition of (three) velocities can be derived very easily by using the four-velocity concept, as we shall see below. This derivation is much simpler than the straightforward but extremely tedious application of Lorentz transformations that you might have already come across in the Special Relativity course.

Indeed, consider two frames moving with three velocities \vec{v}_1, \vec{v}_2 with respect to some inertial frame. The corresponding four-velocities are in components: $u_1 = (\gamma_1, \gamma_1 \vec{v}_1), u_2 = (\gamma_2, \gamma_2 \vec{v}_2)$.

First let us change coordinates by going to the rest frame of one of the moving observers, say the one with four velocity u_1 . In that frame one would have $u_1 = (1, 0), u_2 = (\gamma, \gamma \vec{v})$, where \vec{v} is the relative velocity of the two frames, and γ the corresponding Lorentz factor. Since the covariant inner product of u_1 and u_2 is, like any such products between two four vectors, an *invariant*, one can evaluate it by going to the easiest frame, in this case the frame in which one of the frames, say 1, is at rest. The result is then

$$u_1 \cdot u_2 \equiv u_1^\mu u_2^\nu \eta_{\mu\nu} = -\gamma = -(1 - v^2)^{-1/2} \quad (3.23)$$

This is valid in all frames, and hence by going back to the initial frame, one can write:

$$u_1 \cdot u_2 = -(1 - v^2)^{-1/2} = \gamma_1 \gamma_2 (-1 + \vec{v}_1 \cdot \vec{v}_2) \quad (3.24)$$

from which solving with respect to v^2 one obtains the composition law of three velocities in Special Relativity:

$$|\vec{v}|^2 = \frac{(\vec{v}_1 - \vec{v}_2)^2 - (\vec{v}_1 \times \vec{v}_2)^2}{(1 - \vec{v}_1 \cdot \vec{v}_2)^2} \quad (3.25)$$

where we took into account that $(\vec{v}_1 \times \vec{v}_2)^2 = \vec{v}_1^2 \vec{v}_2^2 - (\vec{v}_1 \cdot \vec{v}_2)^2$.

Exercise 3.8 Using tensor calculus, and in particular properties of the antisymmetric symbol in three Euclidean space dimensions, $\epsilon_{ijk}, i, j, k = 1, \dots, 3$, prove that: $(\vec{v}_1 \times \vec{v}_2)^2 = \vec{v}_1^2 \vec{v}_2^2 - (\vec{v}_1 \cdot \vec{v}_2)^2$.

Exercise 3.9 A particle of (rest) mass m and four momentum p_μ in a Minkowski space time is examined by an observer who moves with a four velocity u^μ . Show the following:

- (a) the energy the observer measures is $E = -p_\mu u^\mu$.
- (b) the (rest) mass he attributes to the particle is $m^2 = -p_\mu p^\mu$.
- (c) the three momentum the observer measures has magnitude: $|\vec{p}| = [(p_\mu u^\mu)^2 + p_\mu p^\mu]^{1/2}$.
- (d) the ordinary three velocity the observer measures has magnitude $|\vec{v}| = \left[1 + \frac{p_\mu p^\mu}{(p_\mu u^\mu)^2}\right]^{1/2}$.
- (e) Find the components of the four vector $\mathcal{V}_\mu \equiv -u_\mu - \frac{p_\mu}{p_\nu u^\nu}$ in the observer's rest (Lorentz) frame.

In a similar manner, by using four-velocity formalism, one obtains easily the relativistic formula for the Doppler shift. We leave the derivation as an exercise for the reader.

Exercise 3.10 The Relativistic Doppler Shift: A moving radioactive source emits a photon gamma ray with frequency ν_0 as measured in the source's rest frame. The source is traveling with velocity $\vec{\beta} \equiv v/c$ with respect to some inertial frame. If \hat{n} denotes the unit vector pointing towards the source at the time of emission, as measured by the observer, show that the frequency ν_{obs} the observer measures when the gamma ray reaches him is given by (Doppler shift):

$$\nu_{\text{obs}} = \nu_0 \frac{1}{\gamma(1 + \vec{\beta} \cdot \hat{n})}, \quad (3.26)$$

where γ is the usual Lorentz factor. What is the form of the Doppler shift in the non relativistic limit of low velocities $|\vec{\beta}| \ll 1$?

3.6 Fluids in Special Relativity

3.6.1 Concepts and Definitions

The reason why fluids will be important for our purposes in this course is the fact that in many situations in general relativity the source of the gravitational field can be taken (to a first approximation) to be a fluid, and in fact a perfect fluid (see below). Fluids are a special kind of *continuum*, i.e. a collection of particles, which is on the one hand large enough so that any individual particle characteristics disappear from the bulk dynamics, thereby implying that the collection is characterized by average collective quantities (speed, energy density, pressure, temperature, etc.), and on the other hand is small enough so that the behaviour of the collection is more or less uniform. Such a collection of particles is called an *element*. For instance, consider water in a bounded region of space such as a lake, and suppose we are interested in studying the gravitational field it generates. The properties of the water (energy density, pressure, temperature, etc.) are not the same globally, e.g. the pressure at the bottom of the lake is higher than at the surface. However, one can always find sufficiently small elements in which such properties are uniform, thereby implying that we can find appropriate average quantities (applicable to individual elements) to describe the behaviour of the water in these small regions of space. From this we can deduce the properties of the lake as a whole, and the gravitational field it generates.

For general fluids, in addition to the pressure, there are also forces parallel to the interface between two neighbouring elements. These forces are responsible for *rigidity* of the fluid. Two adjacent elements can push and pull each other, and for the extreme case of fluids called “solids” they can prevent sliding of adjacent elements along their common boundary. In the more familiar cases of “liquid” fluids such anti-sliding (shear or stress) forces are not strong enough to prevent this sliding, and liquids are not rigid.

A *perfect fluid*, which will be of interest to us in this course, is the one in which all shearing (anti-sliding) forces are absent and the only kind of force between neighbouring elements is due to pressure (which acts normal to the interface between two elements). It is the purpose of the next section to give a precise mathematical definition of a perfect fluid in the context of special relativity. For this purpose we shall use the covariant tensor formalism; in this sense this section will constitute an interesting physical application of the material covered in the tensor calculus section of this course.

3.6.2 The Stress-Energy Tensor

In special relativistic particle mechanics we have seen that the energy and the momentum of a particle are represented as components of a four-vector p^μ , which is really a rank- $\binom{1}{0}$ tensor. A natural question arises, therefore, as to how one can represent, in a covariant formulation, the energy and momentum *density* of a fluid. Clearly these quantities, being densities, cannot be the components of a four-vector. To understand this, consider the case of a perfect fluid, with N particles included in an infinitesimal volume element of the fluid with volume $dV_{\text{mcrf}} = dx dy dz$ in the rest frame of the fluid, the so-called *momentarily comoving rest frame* (MCRF). The MCRF is defined as the frame in which all the N particles are momentarily at rest. For an observer \bar{O} , who moves with a velocity v with respect to the MCRF, say in the x -direction, the volume element will appear Lorentz-contracted by the γ -factor (3.1) (see figure 7)

$$dV_{\bar{O}} = \frac{dV_{\text{mcrf}}}{\gamma}. \quad (3.27)$$

The number of particles N inside this volume will be the *same* for all observers, hence the density of particles, n , will transform as follows:

$$n_{\bar{O}} \equiv \frac{N}{dV_{\bar{O}}} = \gamma n_{\text{mcrf}}. \quad (3.28)$$

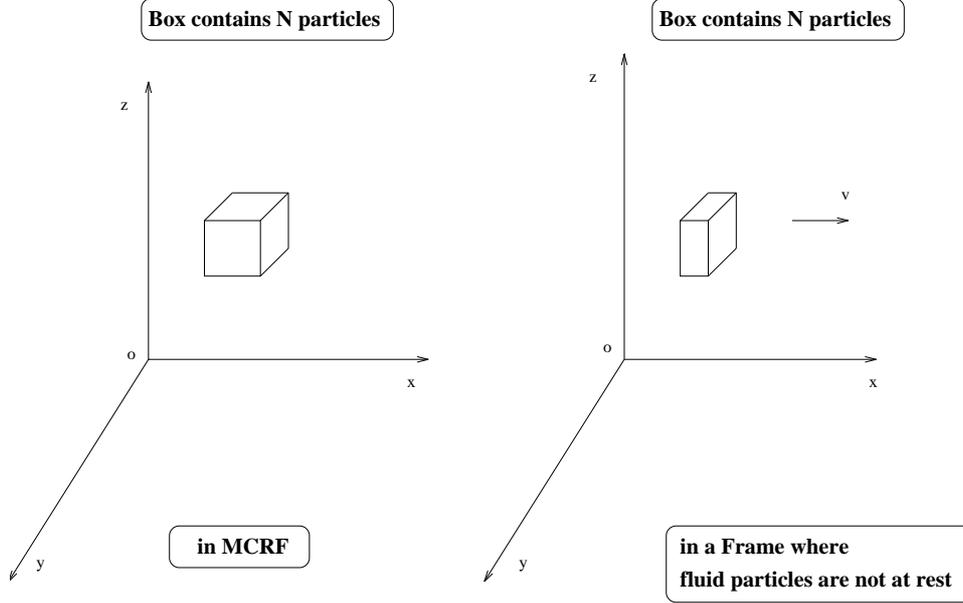


Figure 7: Lorentz contraction along the direction of motion for a fluid volume element. The fluid moves with velocity v along, say, the x -direction, with respect to an inertial observer. The parallelepiped on the left denotes the volume element in the mcrf, whilst that on the right denotes the same volume element as seen by an inertial observer with respect to the fluid.

Clearly the energy of each particle in the MCRF will be its invariant mass m (in units where $c = 1$), since the particles are at rest in that frame. Thus

$$\rho_{\text{mcrf}} = mn_{\text{mcrf}}. \quad (3.29)$$

Given that in the frame $\bar{\mathcal{O}}$ the energy will be (from equation (3.8))

$$E_{\bar{\mathcal{O}}} = m\gamma$$

one obtains that the energy density $\rho_{\bar{\mathcal{O}}}$ is given by

$$\rho_{\bar{\mathcal{O}}} \equiv E_{\bar{\mathcal{O}}}n_{\bar{\mathcal{O}}} = \gamma^2\rho_{\text{mcrf}}. \quad (3.30)$$

Similarly one can argue about the momentum density which also requires two Lorentz γ -factors in the corresponding transformation. These cannot be the transformations of a four-vector (which would require only one γ -factor) but they suggest that the energy and momentum densities of a fluid are components of a rank- $\binom{2}{0}$ tensor, with one factor of γ coming from each factor of $\partial\bar{x}/\partial x$ in the (Lorentz) transformation of the tensor.

We are thus led to define a rank- $\binom{2}{0}$ tensor, called the STRESS-ENERGY TENSOR (or, equivalently, the energy-momentum tensor), whose 00-component is the energy density and whose $0i$ -components are the momentum densities (along the i^{th} spatial direction):

$$T^{\mu\nu} = \left(\begin{array}{c} \text{flux of } \mu\text{-component of four-momentum} \\ \text{across a surface of constant } x^\nu. \end{array} \right). \quad (3.31)$$

In the above, by *flux* we mean the rate of momentum transfer per unit area. The various compo-

nents of the stress-energy tensor are defined as follows:

T^{00} = flux of energy across the surface $t = \text{constant}$ = energy density
T^{0i} = flux of energy across the surface $x^i = \text{constant}$ = energy density \times speed it flows at
T^{i0} = flux of i -momentum across the surface $t = \text{constant}$
T^{ij} = stress (diagonal elements = pressure, and off-diagonal elements = shear).

The dimensions for every component, in our system of natural units where $c = 1$, are $[\text{length}]^{-2}$. The diagonal spatial components of the stress tensor may be expressed, by means of Newton's second law (2.2) in terms of the *pressure* (i -th component of the force per unit area acting on the surface $x^i = \text{constant}$, i.e. a force acting perpendicularly on the surface which defines pressure).

For a perfect fluid all the components of the pressure are the same (isotropic), i -th component of T is $T^{ii} = p$, and there is no shear (i.e. no viscosity). Hence $T^{\mu\nu}$ as a matrix is diagonal, and can be represented by the following 4×4 matrix in the MCRF

$$T^{\mu\nu} = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}. \quad (3.32)$$

To find the stress tensor in a general frame with four-velocity u^μ , it is sufficient to notice that in the MCRF one has $u^0 = 1$ and $u^i = 0$ from which it follows that (3.32) can be written in covariant form as

$$T^{\mu\nu} = p\eta^{\mu\nu} + (p + \rho)u^\mu u^\nu, \quad (3.33)$$

where $\eta^{\mu\nu}$ is the inverse of the Minkowski metric.

Exercise 3.11 *Verify that (3.32) and (3.33) are equivalent.*

Given that (3.33) is a *tensorial* relation in the MCRF it is formally valid in all frames. As we shall see later on this relation can be generalized to arbitrary spacetimes by replacing the Minkowski metric by a general metric. Both equations (3.32) and (3.33) can be taken as a mathematical definition of a perfect fluid.

A particular kind of perfect fluid is *dust*, for which by definition there is no pressure, $p = 0$. Dust will be useful for us in Cosmology.

An important property of the stress-energy tensor, which can already be seen from the special form for the perfect fluid, but applies to general fluids is that it is *symmetric*:

$$\boxed{T^{\mu\nu} = T^{\nu\mu}} \quad (3.34)$$

We can give a physical proof of the symmetry by recalling the definitions of the various components. First look at the T^{0i} components, which express the energy density times mean velocity of energy flow in the i^{th} direction which is equal to the mass density times the i^{th} component of the mean velocity of the energy flow. The latter quantity is just the i^{th} component of the momentum density across the surface $t = \text{constant}$, which is just T^{i0} . To prove the symmetry of the stress components T^{ij} we first go to the MCRF, which is convenient¹. Consider the torque ($\vec{r} \times \vec{F}$ where F is the force) along the z -component exerted on a small cube of side ℓ with the Lorentz MCRF origin at

¹Due to the tensorial nature of T if one proves the symmetry in one frame it is valid in all frames.

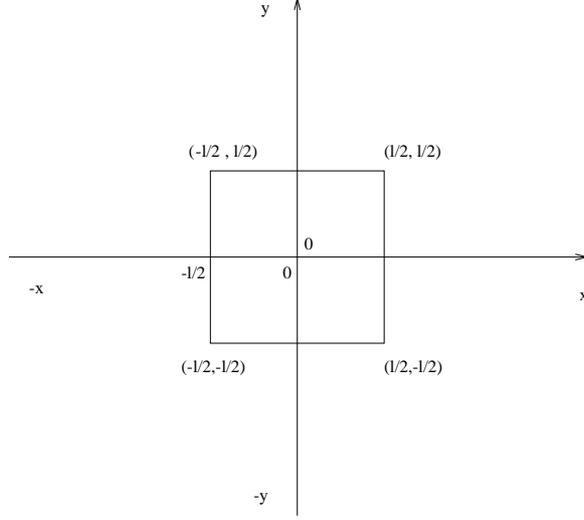


Figure 8: Towards a physical understanding of the symmetry of the stress energy tensor: the stresses must arrange themselves in such a way that there is no torque exerted on the fluid volume element (cube). Here the centre of the Lorentz (mcrf) frame of the fluid is placed at the centre of the cube for convenience.

the centre of the cube, as in figure 8:

$$\begin{aligned}
\tau^z &\equiv \left(\begin{array}{l} y\text{-component of} \\ \text{force on } +x\text{-face} \end{array} \right) \times \left(\begin{array}{l} \text{displacement to } +x \\ \text{face from origin} \end{array} \right) \\
&\quad + \left(\begin{array}{l} y\text{-component of} \\ \text{force on } -x\text{-face} \end{array} \right) \times \left(\begin{array}{l} \text{displacement to } -x \\ \text{face from origin} \end{array} \right) \\
&\quad - \left(\begin{array}{l} x\text{-component of} \\ \text{force on } +y\text{-face} \end{array} \right) \times \left(\begin{array}{l} \text{displacement to } +y \\ \text{face from origin} \end{array} \right) \\
&\quad - \left(\begin{array}{l} x\text{-component of} \\ \text{force on } -y\text{-face} \end{array} \right) \times \left(\begin{array}{l} \text{displacement to } -y \\ \text{face from origin} \end{array} \right) \\
&= -T^{yx} \times \ell^2 \times \left(\frac{\ell}{2}\right) + T^{yx} \times \ell^2 \times \left(\frac{-\ell}{2}\right) - (-T^{xy}) \times \ell^2 \times \left(\frac{\ell}{2}\right) - T^{xy} \times \ell^2 \times \left(\frac{-\ell}{2}\right) \\
&= (T^{xy} - T^{yx})\ell^3.
\end{aligned}$$

Above we have used Newton's second law (2.2), which is valid in the MCRF, to relate the flux of momentum with the force acting on the surface. Since the torque τ^z decreases with decreasing ℓ only as ℓ^3 while the moment of inertia decreases as ℓ^5 then if we had a non-zero torque τ^z this would give each arbitrarily small cube an arbitrarily large angular acceleration, which is absurd. To avoid this inconsistency, one should have an appropriate distribution of stresses in the fluid such that the torques vanish, which can be arranged by demanding a symmetric $T^{ij} = T^{ji}$ as can be seen from the last relation above.

3.6.3 Conservation Laws and the Stress-Energy Tensor

First we discuss the conservation of energy and momentum in a fluid. Consider the situation depicted in figure 9, where we depict a small fluid volume element of side ℓ . The rate of momentum flow across face (4) is $\ell^2 T^{0x}(x=0)$. The rate of momentum flow across face (2) is $-\ell^2 T^{0x}(x=\ell)$. The rate of momentum flow across face (1) is $\ell^2 T^{0y}(y=0)$. The rate of momentum flow across

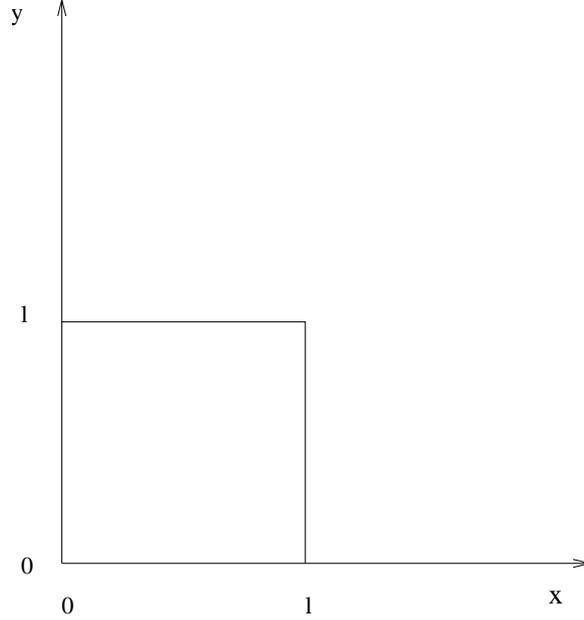


Figure 9: Towards a physical understanding of the conservation laws obeyed by the stress energy tensor. The rectangle denotes a projection on the xy plane of a cubic fluid element.

face (3) is $-\ell^2 T^{0y}(y = \ell)$. The rates of flow across the faces perpendicular to the z -direction are defined similarly (but obviously not depicted in the figure). In the preceding positive flows are out of the cube and negative flows are into the cube. With this convention the sum of these rates must be the rate of increase of energy inside the cube (and considering the limit $\ell \rightarrow 0$):

$$\begin{aligned} \frac{\partial}{\partial t}(T^{00}\ell^3) &= \ell^2[T^{0x}(x=0) - T^{0x}(x=\ell) + T^{0y}(y=0) - T^{0y}(y=\ell) + T^{0z}(z=0) - T^{0z}(z=\ell)] \\ &= -\ell^3 \frac{\partial}{\partial x} T^{0x} - \ell^3 \frac{\partial}{\partial y} T^{0y} - \ell^3 \frac{\partial}{\partial z} T^{0z}, \end{aligned}$$

from which one obtains the following conservation law:

$$\partial_t T^{00} + \partial_i T^{0i} = 0 \quad \implies \quad \partial_\beta T^{\alpha\beta} \equiv T^{\alpha\beta}_{,\beta} = 0. \quad (3.35)$$

Note the use of the convenient comma notation denoting partial differentiation with respect to a coordinate; this will be used from now on.

Suppose that we consider a situation in which $T^{\mu\nu} = 0$ outside a bounded region of space \mathcal{D} and on the boundary of \mathcal{D} ($\partial\mathcal{D}$) itself, and that this region does not change in time. Then Gauss's theorem tells us that for a three-vector \vec{V}

$$\int_{\mathcal{D}} d^3x \partial_i V^i = \oint_{\partial\mathcal{D}} d^2S_i V^i, \quad (3.36)$$

where $\partial\mathcal{D}$ is the (two-dimensional) surface bounding the volume \mathcal{D} and $d^2S_i = d^2x \hat{n}_i$ is the infinitesimal oriented surface element (where \hat{n}_i denotes the unit normal to the surface). This, and the fact that T is zero on the boundary $\partial\mathcal{D}$, implies that when we integrate (3.35) over space we obtain

$$\partial_t \int_{\mathcal{D}} d^3x T^{\alpha 0} = - \oint_{\partial\mathcal{D}} d^2x \hat{n}_i T^{\alpha i} = 0. \quad (3.37)$$

This is actually four equations, one for each value of $\alpha \in \{0, 1, 2, 3\}$. When $\alpha = 0$ the equation gives the conservation of total energy in the volume \mathcal{D} , while when $\alpha = i$ we have total three-momentum conservation:

$$\partial_t E = 0, \quad \partial_t \vec{p} = 0. \quad (3.38)$$

Next we discuss conservation of total angular momentum in connexion with the symmetry of the stress-energy tensor. To this end, consider a Lorentz frame with origin at O^μ and define the rank- $\binom{3}{0}$ tensor

$$\mathcal{J}^{\alpha\beta\gamma} \equiv (x^\alpha - O^\alpha)T^{\beta\gamma} - (x^\beta - O^\beta)T^{\alpha\gamma}, \quad (3.39)$$

where T is the stress-energy tensor. Consider the four-divergence of \mathcal{J} and use the symmetry of T in (3.34) to write

$$\mathcal{J}^{\alpha\beta\gamma}_{,\gamma} = \delta_\gamma^\alpha T^{\beta\gamma} - \delta_\gamma^\beta T^{\alpha\gamma} + (x^\alpha - O^\alpha)T^{\beta\gamma}_{,\gamma} + (x^\beta - O^\beta)T^{\alpha\gamma}_{,\gamma} = T^{\alpha\beta} - T^{\beta\alpha} = 0, \quad (3.40)$$

where we have used the vanishing of the divergence of T in (3.35). One can define the integral over a spacelike three-surface

$$J^{\mu\nu} = \int d^3x \mathcal{J}^{\mu\nu 0} = \int d^3x [(x^\mu - O^\mu)T^{\nu 0} - (x^\nu - O^\nu)T^{\mu 0}]. \quad (3.41)$$

We wish to concentrate on the spatial components of J , where μ and ν run over $\{1, 2, 3\}$. From the Newtonian definition of angular momentum ($\vec{r} \times \vec{p}$) and taking into account that T^{0i} are components of the momentum density, one sees that the ij -components of J is the total angular momentum of the fluid. Using the symmetry of the stress-energy tensor T and hence the vanishing four-divergence of \mathcal{J} one can then apply a similar procedure to the one leading to energy-momentum conservation above to obtain the conservation of total angular momentum. Thus we link the symmetry of T with the conservation of angular momentum in the fluid.

There is an alternative way to arrive at the *integral conservation laws* for energy, momentum and angular momentum, described above, which makes use of a higher-dimensional version of Gauss's theorem, applied directly to four-dimensional integrals over a space time \mathcal{E} . For your information, the proof of the higher-dimensional version of Gauss's theorem is entirely analogous to its three-dimensional counterpart. For the purposes of this course these theorems may be assumed known (without proof), whenever needed.

Consider the vanishing four-divergence of either the stress tensor $T^{\alpha\gamma}_{,\gamma} = 0$ or the covariant angular momentum tensor $\mathcal{J}^{\alpha\beta\gamma}_{,\gamma} = 0$. Integrating these equations over a spacetime volume \mathcal{E} and using the four-dimensional version of Gauss's theorem we have

$$\int_{\mathcal{E}} d^4x T^{\alpha\gamma}_{,\gamma} = \oint_{\partial\mathcal{E}} d^3\Sigma_\gamma T^{\alpha\gamma} = 0, \quad (3.42)$$

and,

$$\int_{\mathcal{E}} d^4x \mathcal{J}^{\alpha\beta\gamma}_{,\gamma} = \oint_{\partial\mathcal{E}} d^3\Sigma_\gamma \mathcal{J}^{\alpha\beta\gamma} = 0, \quad (3.43)$$

where $\partial\mathcal{E}$ is the (three-dimensional) boundary of the spacetime volume \mathcal{E} and $d^3\Sigma_\gamma$ is the oriented infinitesimal (three-dimensional, space-like) hypersurface (i.e. volume) element (which is the generalization of d^2S_i).

The situation concerning the above four-dimensional integrals is depicted in figure (10). From the figure it is clear that we close the volume \mathcal{E} by time like surfaces (denoted by dashed lines in the figure) at infinity. The basic *assumption* is that the latter parts do not contribute to the integral. The results of the four-dimensional Gauss theorem, then, (3.42) and (3.43) imply that the

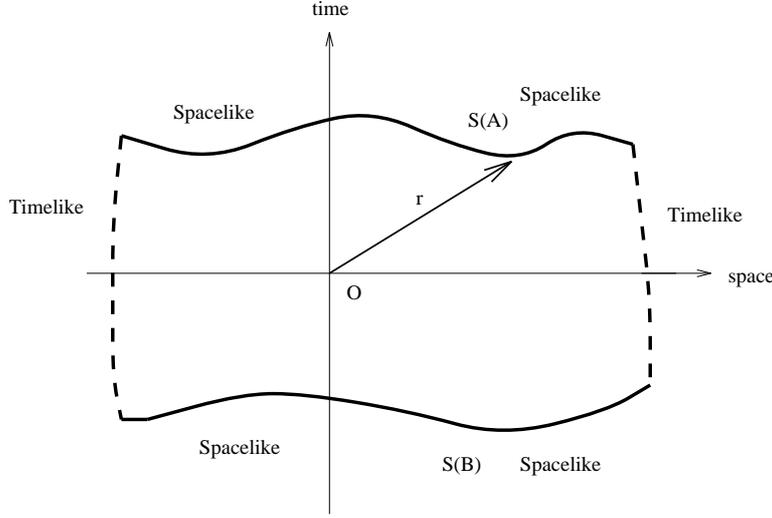


Figure 10: Understanding the integral conservation laws from the point of view of the four-dimensional Gauss's theorem. In the above space-time diagram, $\mathcal{S}(A)$ and $\mathcal{S}(B)$, are space-like hypersurfaces, while the dashed lines are timelike hypersurfaces at spatial infinity, which do not contribute to the integral. All these four hypersurfaces constitute the boundary $\partial\mathcal{E}$ of the four-dimensional spacetime hypervolume \mathcal{E} . The event O denotes the point with respect to which one evaluates the angular momentum of a space-like hypersurface \mathcal{S} .

integrals on the middle of these equations may be written as a difference over the two space-like surfaces $\mathcal{S}(A)$ and $\mathcal{S}(B)$ (cf. figure (10)):

$$\oint_{\mathcal{S}(A)} d^3\Sigma_\gamma T^{\alpha\gamma} - \oint_{\mathcal{S}(B)} d^3\Sigma_\gamma T^{\alpha\gamma} = 0, \quad (3.44)$$

and

$$\oint_{\mathcal{S}(A)} d^3\Sigma_\gamma \mathcal{J}^{\alpha\beta\gamma} - \oint_{\mathcal{S}(B)} d^3\Sigma_\gamma \mathcal{J}^{\alpha\beta\gamma} = 0, \quad (3.45)$$

i.e. the corresponding hypersurface integrals are independent on the hypersurface on which they are evaluated. Choosing the space-like surfaces to be the ones corresponding to constant coordinate time t , then, which is equivalent to considering the index γ taking on only the value 0 in the above relations, we arrive straightforwardly to the integral laws of the conservation of total energy, momentum and angular momentum.

Exercise 3.12 Show that, if T is zero outside and on a boundary $\partial\mathcal{D}$ of some spatial region \mathcal{D} , the following relation is true

$$\frac{\partial^2}{\partial t^2} \int_{\mathcal{D}} d^3x T^{00} x^i x^j = 2 \int_{\mathcal{D}} d^3x T^{ij}, \quad (3.46)$$

where i, j are spatial indices. This is known as the tensor virial theorem.

3.7 Electromagnetism

The tensor calculus we have developed can also be used to write Maxwell's electromagnetism in a compact and covariant manner. The example is very instructive because it demonstrates the

economy of using tensor calculus for writing Maxwell's equations. Maxwell's equations for electric and magnetic fields (\vec{E} and \vec{B} respectively) in the presence of an electric current \vec{J} and charge density ρ read (in units where $c = \mu_0 = \varepsilon_0 = 1$)

$$\begin{aligned}\vec{\nabla} \times \vec{B} - \partial_t \vec{E} &= 4\pi \vec{J}, & \vec{\nabla} \cdot \vec{B} &= 0 \\ \vec{\nabla} \times \vec{E} + \partial_t \vec{B} &= \vec{0}, & \vec{\nabla} \cdot \vec{E} &= 4\pi \rho.\end{aligned}$$

Define an antisymmetric rank- $\binom{2}{0}$ tensor as follows:

$$F^{0i} = E^i \quad F^{ij} = \epsilon^{ijk} B_k, \quad i, j, k \in \{1, 2, 3\}. \quad (3.47)$$

The tensor F is called the *Maxwell tensor*.

Exercise 3.13 Assemble the components of F in a 4×4 matrix and show that this matrix is traceless.

Exercise 3.14 Defining the four-current $J^\mu := (\rho, \vec{J})$, show that two of Maxwell's equations can be written in the compact form

$$F^{\mu\nu}{}_{,\nu} = 4\pi J^\mu,$$

and the other two are a consequence of the so-called Bianchi identity

$$F_{\mu\nu,\lambda} + F_{\nu\lambda,\mu} + F_{\lambda\mu,\nu} = 0. \quad (3.48)$$

Of course, the indices are raised and lowered with the Minkowski metric.

4 Preparing for Curved Space Time

4.1 Comparison Between Newton's Laws and Special Relativity

The first law of Newtonian Mechanics still holds intact. The second Law (2.2) holds in some sense, upon replacing the three-vector of the force by the appropriate four vector F^μ , and the Newtonian time by the universal proper time τ defined above, which is the same for all observers, equation (3.21)

$$F^\mu = \frac{dp^\mu}{d\tau}$$

where p^μ is the four-momentum. The spirit of Newton's second law, of momentum changing under the action of a force is still captured in this special-relativistic extension. The third law, however, fails. The Newtonian treatment assumes that the forces of action and reaction are exerted *simultaneously* through space, and actually that this simultaneity was an objective notion. In Special relativity this notion is abandoned, simply because the notion of simultaneity is observer dependent. Some events that look simultaneous in one frame are not simultaneous in another (e.g. moving frame). This was known to characterize Maxwell's theory, since the electromagnetic interactions propagate with finite speed of light c , and this was already known in the year 1905 when Einstein presented the special theory of relativity. For the same reason Newton's law of gravitation is also abandoned in the novel relativistic treatment.

4.2 The Equivalence principle

Comparing the Newtonian Law of gravitation with the second Law, an immediate question arises: are the 'masses' in these two Laws the same? In the second law one talks about an *inertial mass*, $m^{(I)}$, which in principle is defined as the proportionality constant that connects the acceleration of the particle with the force. This law characterizes, according to Newton, all kinds of force, not only the gravitational one. In other words, one may imagine a situation of a particle in space, far

away from the gravitational centre of attraction of Earth or any other celestial body, so that any gravitational force on the body is negligible. On the contrary, in the Law of Gravity, the mass entering the pertinent formula (2.3), is the so-called ‘gravitational mass’, $m^{(G)}$, the mass of a body which finds itself under the influence of the gravitational field of a nearby gravitational centre. A priori these two masses do not have to be the same. The ratio $m^{(I)}/m^{(G)}$ could, in principle, depend on the chemical composition of the body.

However, already in 1889, the experiment of L. von Eötvös had shown that this ratio was pretty close to one (to an accuracy 10^{-9}) for such different materials as gold and aluminum. This prompted Einstein to define the *weak form of Equivalence Principle*:

The Weak Form of the Equivalence Principle: *The inertial and gravitational masses of a body, as entering the respective Newtonian Laws, are identical, irrespective of the body’s chemical composition.*

This principle has dramatic consequences for the form of Newton’s law of gravitation. Indeed, assume that several particles exist in a *restricted region of space*, which is *small enough* so that the gravitational field on the particles can be considered more or less uniform. Assume also that there are interaction forces \vec{F}_{ij} between the i, j particles. In this case, a combination of the weak equivalence principle, second Newtonian Law (2.1), and the law of gravitation (2.3) implies for the acceleration of the i -th particle in the sample:

$$\ddot{\vec{x}}_i = \frac{1}{m_i} \sum_j \vec{F}_{ij} + \vec{g}$$

where \vec{g} is the (uniform) acceleration due to gravity that the i -th particle experiences.

Suppose now that we change coordinates to those relative to an observer falling freely in the gravitational field. Such a falling-freely observer would have position

$$\vec{z} = \vec{z}_0 + \vec{v}t + \frac{1}{2}\vec{g}t^2$$

, since in our case the gravitational acceleration is assumed uniform.

Relative to this freely-falling observer the i -th particle has coordinates: $\vec{y}_i = \vec{x}_i - \vec{z}_i$, and hence

$$\ddot{\vec{y}}_i = \ddot{\vec{x}}_i - \ddot{\vec{z}}_i = \frac{1}{m_i} \sum_j \vec{F}_{ij} \quad (4.1)$$

i.e. by changing to these coordinates the gravitational field *vanished* completely, and the formula (4.1) assumed the form it would assume in the theory without gravitation.

This fact depends on a crucial assumption, that the region of space on which the experiment under consideration took place was *small enough* so that the gravitational field could be considered uniform. These simple facts led Einstein to formulate the *strong version of the Equivalence Principle*:

Strong form of the Equivalence Principle: *At every space-time point, in an arbitrary gravitational field, it is possible to choose a locally inertial (‘free-float’) coordinate frame, such that within a sufficiently small region of space and time around the point in question, the laws of Nature are described by special relativity, i.e. are of the same form as in unaccelerated Cartesian coordinate frames in the absence of Gravitation.*

In other words, locally one can always make a coordinate transformation such that the space time looks *flat*. This is not true globally, and in the next section we are going to demonstrate this with a simple thought experiment.

The Equivalence principle (in its strong form) serves a similar purpose to that of the *Correspondence Principle* between Classical and Quantum Mechanics. The latter is the vehicle from classical to quantum Physics, and in this spirit the Strong Equivalence Principle may be considered our vehicle from Special Relativity (theory of flat space times) to General Relativity (theory of curved space times).

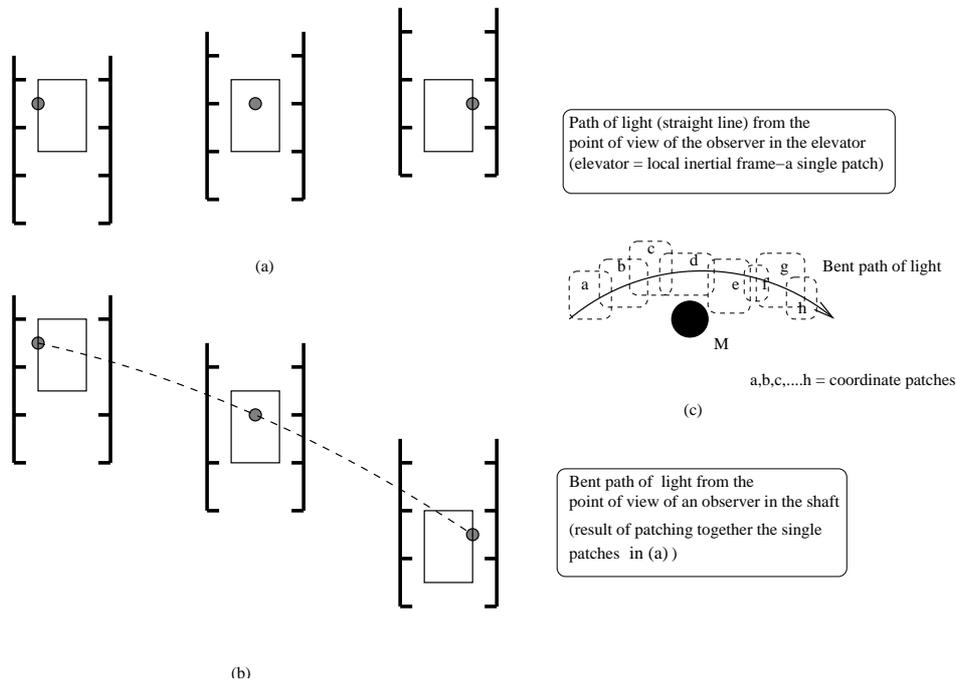


Figure 11: Einstein's elevator (thought) experiment to demonstrate qualitatively the bending of light in a gravitational field from the Strong Equivalence Principle.

4.3 Some important consequences of the Equivalence Principle

Using the strong equivalence principle we can understand qualitatively two important facts about the behaviour of light in gravitational fields, namely, (i) *the gravitational redshift* which, as we saw above, can be viewed as a consequence of Newtonian gravitation and energy conservation, and also (ii) *the bending of light*.

The *gravitational redshift* can be easily understood in the spirit of the strong equivalence principle as follows. According to this principle the effects of a uniform gravitational field, which is a valid approximation in a sufficiently small region of spacetime, can be considered equivalent to the effects of a uniform acceleration in a coordinate frame in the absence of gravitation. Consider therefore a uniformly accelerating space rocket and two experimenters, A and B in it, separated by distance H (with B nearer the nose of the rocket) in the direction of acceleration, as measured in the reference frame of the rocket. One of the experimenters, A sends a photon to the other one; assume for convenience that the rocket is at rest at the time the photon is emitted. The photon will travel a distance H in time $t = H$ (remember $c = 1$ in our units). The other experimentalist will receive that photon when he is traveling at a speed $v = gt = gH$ which of course implies that the photon will be Doppler-shifted. The Doppler-shift z is related to the wavelength change by (c.f. the non-relativistic limit of the Doppler shift (3.26) obtained in the pertinent exercise):

$$1 + z = \frac{\lambda_A}{\lambda_B} = \frac{\nu_B}{\nu_A} = 1 + gH \quad (4.2)$$

from which follows the redshift equation (2.36), for the non-relativistic limit of sufficiently small H , which also guarantees uniform gravitational fields.

Secondly, the *bending of light* can be seen qualitatively by Einstein's elevator thought-experiment, see figure 11. Consider an elevator falling in a shaft under the influence of gravity alone, with an unfortunate observer inside. At the time the elevator starts moving downwards a photon

is emitted horizontally (in the observer's frame) from one side of the elevator towards the other. From the point of view of the observer the photon will travel in a straight line as a consequence of the Strong Equivalence Principle and the fact that for the accuracy required by the observer, the dimensions of the elevator are small (so the spacetime in the elevator is essentially flat). However, from the point of view of an observer in the shaft, the light will bend as shown in figure 11. Of course, it is a non-trivial statement that this will actually happen, because knowledge of the precise orbit of light in the presence of a gravitational field depends on the detailed dynamics. The strong equivalence principle implies that our spacetime can be considered as a patchwork of many small areas of flat spacetime. This patching depends on the dynamics, and as we shall see, Einstein's equations will tell us exactly what orbit a photon in a gravitational field will follow. This will be done in later sections. At the moment we shall try to understand qualitatively what curvature of spacetime means and why special relativity is incompatible with gravitation.

4.4 The Gravitational Redshift as Implying Incompatibility of Special Relativity and Gravitation

A good way to start understanding why Einstein's special theory was not the correct theory of gravitation is the above-mentioned gravitational redshift, which as we have seen follows from either Newtonian gravitation and the basic principle of energy conservation, or from the strong equivalence principle alone. Indeed, take as an experimental fact that two observers of the Pound–Rebka–Snider experiment of figure 3 observe the redshift (2.36). Assume for convenience that the observer B emits n cycles of light with frequency $\omega_B = 2\pi\nu_B$ in a time $\delta\tau_B$, i.e.

$$2\pi n = 2\pi\nu_B\delta\tau_B . \quad (4.3)$$

It is important that the experimenters are static with respect to the Lorentz frame of the Earth and with respect to each other. This is why their time is identified with the proper time. The observer A will receive the same n cycles of light with a frequency ω_A different from ω_B (see equation (2.36)) in a time $\delta\tau_A$:

$$2\pi n = 2\pi\nu_A\delta\tau_A . \quad (4.4)$$

If special relativity were right, the pertinent spacetime graph for this experiment is given in figure 12. Because the experiment is static and the gravitational field is assumed also static and uniform within the conditions of the experiment, the spacetime graph should be a parallelogram (see figure 6(a), if you assume the light follows a path along the lightcone, i.e. a straight line at 45° in the $t - z$ plane, where z is the direction of motion) or at most the world-lines of the photons, if the gravitational bending of light is taken into account, should be congruent (see figure 6(b)). In both cases this would imply that $\delta\tau_B = \delta\tau_A$, which, in view of (4.3),(4.4), and the fact that the observer A receives the same n cycles of light as B , would contradict the redshift result (2.36). This argument in favour of abandoning Special Relativity in our discussion of a curved space time is due to Schild (1960).

5 Curving

5.1 Einstein's view of Gravitational 'Force'

We are now well equipped to start our tour of the theory of curved spacetimes and gravitation. Einstein's basic assumption, which revolutionized our view of the Universe, was that the concept of a Newtonian gravitational "force" should be replaced by the concept of a "curved" spacetime. As we shall see in subsequent parts of these notes, any *non-trivial mass distribution* in space time is responsible for inducing a *non-trivial curvature*, which is a direct consequence of the dynamics encoded in Einstein's equations. Thus, according to Einstein, and in contrast to Newtonian mechanics, in the case of a satellite of mass m which *orbits* around a massive body of mass M ,

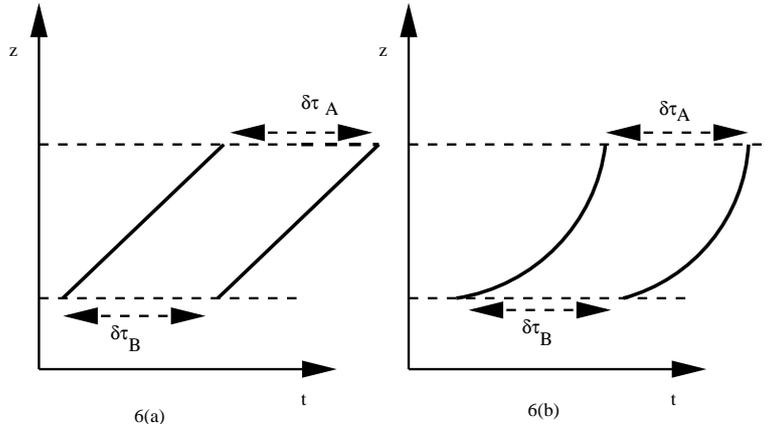


Figure 12: Schild's spacetime diagram argument on the incompatibility of the gravitational redshift phenomenon with special relativity.

there are *no forces* exerted on the satellite, but the latter defines a 'free float' inertial frame in the curved space-time environment of the massive body. In fact, the satellite follows a *geodesic* in the curved space time induced by the Massive body of mass M .

In this section we shall attempt to understand and quantify these concepts in a relative simple manner. First we shall discuss some 'empirical' and/or 'experimental' evidence for the curvature of spacetime, and then we shall proceed in formulating mathematically the concepts of: (i) geodesics, which is the closest concept to the 'straight lines in flat-space time Newtonian mechanics', (ii) that of parallel transport along a geodesic, and finally (iii) spacetime curvature, by first constructing the Riemann Curvature Tensor, as being related to properties of neighbouring geodesics in a spacetime with non-trivial metric tensor, specifically the so-called 'geodesic deviation', and then studying its most important properties. The use of results from our tensor calculus part of the lectures, will be an invaluable tool for a complete understanding of this chapter of the notes.

5.2 Preface to Curvature: Tidal Accelerations

Evidence for the non-trivial curvature of spacetime can be obtained by considering the following thought experiment, due to Einstein, which is depicted in figure 13. Consider a *long and narrow* railway coach, with two test particles in it, A and B. Consider the two cases depicted in the figure 13: (a) the case of a *long and narrow* horizontal railway coach, with two test particles A and B in its two ends. The coach is freely falling, keeping its horizontal disposition, toward Earth as in figure 13a. The two test particles are originally released side by side, but are both attracted toward the centre of Earth, and hence they move closer together, as measured by an observer inside the railway coach. This motion is not related to the gravitational attraction between the two test particles, which by assumption is negligible. It is entirely due to the non-uniform gravitational field of Earth, since the coach is long enough so that such non-uniform field effects are appreciable. (b): the case of the railway coach falling freely *vertically* (i.e. along the radial direction of Earth), in which the test particles find themselves one above the other, as in figure 13b. For such vertical separations, according to Newtonian analysis, the gravitational accelerations of A and B are in the same direction towards Earth, however the particle B nearer Earth is more strongly attracted and gradually leaves the other behind: *the two particles move further apart as observed inside the coach*.

From these two cases we conclude that the large railway coach is not a free-float frame. An observer inside the railway coach in either case, sees the pair of test particles *accelerate* toward one

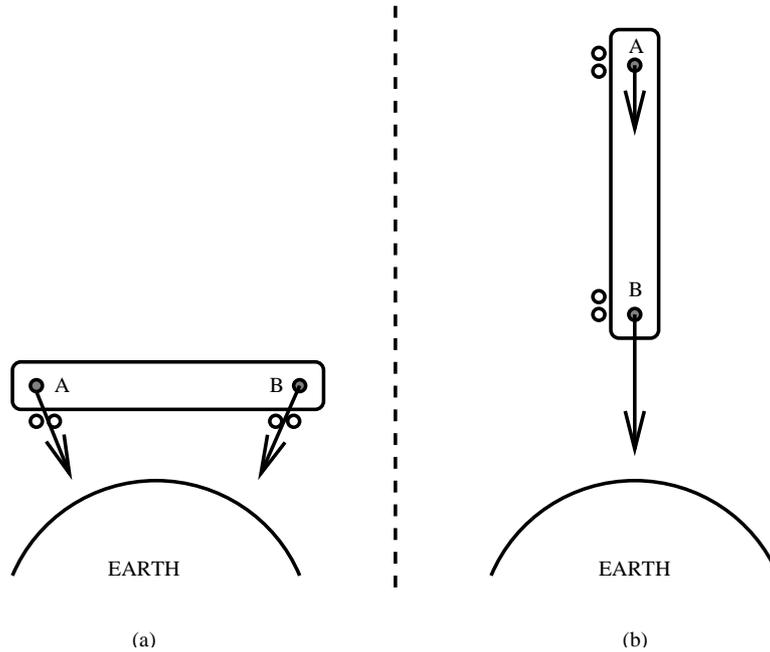


Figure 13: Einstein's thought experiment to demonstrate non-trivial curvature of space time.

another or away from one another. Such relative motions are called **tidal accelerations**, because they arise from the same kind of non-uniform gravitational field (in this case that of Moon), which accounts for ocean tides on earth. This is considered as evidence of curvature of spacetime. The concept of tidal accelerations can be quantified mathematically using geodesic deviation, which we now proceed to analyse in detail, in conjunction with other related topics.

5.3 Curves in a General Relativistic Framework

The world lines of a particle, familiar from Special Relativity, is a concept that can be extended intact in a curved space time. Particles follow such curves in arbitrary space times. Such curves are parametrised by a real parameter λ , $\mathcal{P}(\lambda)$. In Newtonian mechanics curves are parametrized by the time t but here we wish to be more general, and use an arbitrary real parameter for our curve parametrization.

Familiar from our Newtonian mechanics is also the concept of a velocity $d\mathcal{P}(t)/dt$, or in our more general parametrization $d\mathcal{P}(\lambda)/d\lambda$. The velocity is a vector in three-dimensional Euclidean space time. In the general relativistic setting, a curve in a given coordinate system is parametrized by the components of the coordinate four vectors $x^\mu(\lambda)$. In analogy with the three-dimensional Euclidean space, the quantities

$$t^\mu \equiv \frac{dx^\mu(\lambda)}{d\lambda} \quad (5.1)$$

are four vectors, called the *tangent vectors* of the curve at a point with coordinates $x^\mu(\lambda)$.

Exercise 5.1 Show that under a coordinate transformation $x^\mu \rightarrow \bar{x}^\mu(x)$, the quantities t^μ (5.1) do transform as contravariant four vectors.

The tangent vectors of a curve at a given point *lie on a single plane*, called the *tangent plane*.

The curves can be classified in a similar manner as the world lines in Special Relativity, since according to the Strong Equivalence Principle one can always choose a local frame where the spacetime is flat. Thus there are also three types of curves in general relativity:

1. time-like, which always lies inside the local lightcone
2. space-like, which always lies outside the local lightcone
3. null, always lying on the local lightcone: the path of a photon is a null (extremal) curve (or null geodesic as we shall see in a later subsection).

5.4 Invariant Interval between two Events in Curved Space Time and the concept of Metric

An important concept of the Special Relativity, which is also extended to arbitrarily curved space times is that of the *invariant interval* Δ^2 (3.2). The interval Δ^2 is invariant under Lorentz transformations in flat space times, as we learn from our Special Relativity course. This invariance can be extended into arbitrarily general coordinate transformations $x^\mu \rightarrow \bar{x}^\mu$, *provided* one introduces the *metric tensor* in the expression for Δ^2 , or its infinitesimal form ds^2 pertaining to two neighbouring points ('events') in space time.

Exercise 5.2 Consider the Minkowski space invariant integral between two points infinitesimally close to each other. Then, the interval between the neighbouring points x^μ and $x^\mu + dx^\mu$ is: $ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu$. We now perform the following general coordinate transformation: $x^\mu \rightarrow \bar{x}^\mu$. Assuming invariance of ds^2 show that in the new coordinate frame ds^2 is written in the form:

$$ds^2 = g_{\mu\nu}(\bar{x}) d\bar{x}^\mu d\bar{x}^\nu \quad (5.2)$$

and express $g_{\mu\nu}$ in terms of $\partial x^\mu / \partial \bar{x}^\nu$.

The quantity $g_{\mu\nu}$ is called the metric, and is a second rank covariant symmetric tensor

$$g_{\mu\nu} = g_{\nu\mu} \quad (5.3)$$

This guarantees the invariance of the expression in the right-hand-side of (5.2).

Exercise 5.3 Show that $ds^2 = g_{\mu\nu}(\bar{x}) d\bar{x}^\mu d\bar{x}^\nu$, where $g_{\mu\nu}$ is a second rank covariant tensor, is invariant under arbitrary general coordinate transformations.

In this way one can also define the angle between two directions dx^μ and $d\bar{x}^\nu$ in a general relativistic setting:

$$\cos(dx^\mu, d\bar{x}^\nu) \equiv \frac{g_{\mu\nu} dx^\mu d\bar{x}^\nu}{\sqrt{ds^2 d\bar{s}^2}} \quad (5.4)$$

This definition reduces to the familiar one of $\cos\theta$ in the case of flat three-dimensional Euclidean space, where θ the angle between the two directions.

The metric $g_{\mu\nu}$ is a real symmetric matrix so that the square length of a vector is a real number. By changing coordinate systems the form of the metric changes in general, but there are some properties of it which do not change, the simplest of them being its *signature*. This is defined as follows: the metric $g_{\mu\nu}$, as a second rank covariant tensor, can be represented as a real symmetric matrix (4×4 in four dimensional space times, $d \times d$ for d -dimensional space times). As such, the metric can be diagonalized. The signature of the metric is simply the number n^- of negative eigenvalues and the number n^+ of positive eigenvalues. Symbolically $(n^-, n^+) = (-, -, \dots, -, +, \dots, +)$.

It should be noticed that there are space times where there is a signature change in certain regions. An example is the Black-Hole (Schwarzschild Space time) which we shall study in subsequent sections. By continuity of the metric tensor such space times are characterized by points

(or surfaces in general) where the metric vanishes. These are usually coordinate singularities, as we shall see, which express a bad choice of coordinates parametrizing the space time. They are called *horizons* as we shall discuss in detail later on.

Exercise 5.4 Determine the signature of the metric for the following space times:

- (i) A two dimensional unit sphere characterized by the line element $ds^2 = d\theta^2 + \sin^2\theta d\phi^2$.
- (ii) A four dimensional Minkowski space time (space time of Special Relativity).
- (iii) A three-dimensional Schwarzschild space time described by the line element: $ds^2 = -(1 - 2M/r)dt^2 + (1 - 2M/r)^{-1}dr^2 + r^2d\phi^2$. Where does the horizon occur ?

The contravariant second rank tensor $g^{\mu\nu}$ is defined as the *inverse* of $g_{\mu\nu}$, i.e.:

$$g_{\alpha\mu}g^{\mu\beta} = \delta_{\alpha}^{\beta} \quad (5.5)$$

where δ_{α}^{β} is the Kronecker δ symbol, i.e. $\delta_{\alpha}^{\beta} = 1$ if $\alpha = \beta$, and $\delta_{\alpha}^{\beta} = 0$ if $\alpha \neq \beta$.

From this property it follows immediately that: $g_{\mu\nu}g^{\mu\nu} = d$, for a d -dimensional space time.

The way the metric appears above, i.e. in defining the invariant length Δ^2 (or equivalently ds^2 for infinitesimal separations) between two events in a curved space time, is actually generalized to defining the length of any vector in this space time. Let ξ^{α} such a four vector. Then its length in a space time with metric $g_{\mu\nu}$ is defined by the (invariant) inner product:

$$\xi^2 \equiv g_{\mu\nu}\xi^{\mu}\xi^{\nu} \quad (5.6)$$

Exercise 5.5 Show that the length of a vector, defined in (5.6) is indeed invariant under a general coordinate transformation.

In general the metric appears in the definition of the invariant inner product between two four vectors in a curved space time

$$A \cdot B \equiv (A, B) \equiv A^{\mu}B^{\nu}g_{\mu\nu} \quad (5.7)$$

In this way, the ‘angle’ relation (5.4) can be generalized to two arbitrary four vectors A^{μ} and B^{ν} , in a space time with metric $g_{\mu\nu}$. One defines the angle independent of the scaling of the two vectors, by a direct generalization of (5.4), as:

$$\cos(A^{\mu}, B^{\nu}) = \frac{g_{\mu\nu}A^{\mu}B^{\nu}}{|A||B|}, \quad |A| \equiv \sqrt{g_{\mu\nu}A^{\mu}A^{\nu}}, |B| \equiv \sqrt{g_{\mu\nu}B^{\mu}B^{\nu}}. \quad (5.8)$$

Exercise 5.6 A conformal transformation of a metric is defined as the one under which the metric transforms as: $g_{\mu\nu} \rightarrow f(x^{\alpha})g_{\mu\nu}$, where f is a real scalar function of x^{α} , $\alpha = 0, \dots, 3$. Show that for an arbitrary choice of the function $f(x^{\mu})$ a conformal transformation preserves all angles in a general relativistic framework.

Invariant inner products of tensors are also defined with the help of the metric tensor, e.g. in the case of two second rank covariant tensors, their invariant inner product is $T_{\mu\nu}T_{\alpha\beta}g^{\mu\alpha}g^{\nu\beta}$ etc. We stress that such products of tensors have no free indices, i.e. all the indices have been contracted, and hence they are scalar, i.e. invariant in all frames.

5.5 The Geodesic Equation for curved space times

5.5.1 Newtonian Dynamics in Euclidean Space

In Newtonian Mechanics the first law assures that a particle, on which there are no forces acting, will continue to move in a straight line, which is the *shortest path* connecting two points in flat space times, where Newtonian Mechanics applies. This concept finds a natural extension in curved space

times, upon replacing the concept of a straight line with that of a *geodesic*. Again the geodesic is the shortest path between two points in curved space time. An important ingredient in Einstein's theory of gravitation is that a particle moving on a geodesic is *force-free*, and hence the particle is floating freely. In this subsection we shall discuss the mathematical formulation of the important concept of a geodesic curve.

Since the geodesic is the shortest curve connecting two points in a curved spacetime, with metric tensor $g_{\mu\nu}$, it is obtained by extremizing (minimizing) the arc length s for given initial (s_i) and final (s_f) points:

$$s = \int_{s_i}^{s_f} ds = \text{extremum} = \int_{\lambda_i}^{\lambda_f} d\lambda \sqrt{g_{\alpha\beta} \frac{dx^\alpha}{d\lambda} \frac{dx^\beta}{d\lambda}} \quad (5.9)$$

Extremization of (5.9) is a variational problem which has mathematically the same form as the Hamilton's principle with Lagrangian: $L = \sqrt{g_{\alpha\beta} \frac{dx^\alpha}{d\lambda} \frac{dx^\beta}{d\lambda}}$. This procedure implies the Euler–Lagrange equations (c.f. Appendix A):

$$\frac{d}{d\lambda} \frac{\partial L}{\partial x'^\alpha} - \frac{\partial L}{\partial x^\alpha} = 0. \quad (5.10)$$

where primes denote differentiation with respect to λ , $x'^\alpha = dx^\alpha(\lambda)/d\lambda$. Writing

$$L = \sqrt{g_{\alpha\beta} x'^\alpha x'^\beta} \equiv \sqrt{F} \quad (5.11)$$

the Euler–Lagrange equations reduce to

$$\frac{d}{d\lambda} \left[\frac{1}{\sqrt{F}} \frac{\partial F}{\partial x'^\alpha} \right] - \frac{1}{\sqrt{F}} \frac{\partial F}{\partial x^\alpha} = 0, \quad (5.12)$$

from which one obtains:

$$\frac{1}{2F\sqrt{F}} \left[-\frac{dF}{d\lambda} \frac{\partial F}{\partial x'^\alpha} + 2F \frac{d}{d\lambda} \frac{\partial F}{\partial x'^\alpha} - 2F \frac{\partial F}{\partial x^\alpha} \right] = 0, \quad (5.13)$$

or

$$-\frac{dF}{d\lambda} [g_{\alpha\beta} x'^\beta] + 2F \frac{d}{d\lambda} [g_{\alpha\beta} x'^\beta] - F g_{\mu\nu, \alpha} x'^\mu x'^\nu = 0. \quad (5.14)$$

For this extremal curve we can choose the parameter λ to be proportional to the arc length. Hence

$$ds = L d\lambda \quad \Rightarrow \quad \lambda \sim s \quad (5.15)$$

The parameter λ (and hence s), which satisfies the above restriction is called an *affine parameter*, and the path which solves the associated equation (5.14) is called an *affine geodesic*.

From (5.9),(5.11) we obtain in this case that $F = \text{constant}$, implying $dF/d\lambda \propto dF/ds = 0$. Hence, the first term in (5.14) vanishes, yielding:

$$\frac{dg_{\alpha\beta}}{d\lambda} x'^\beta + g_{\alpha\beta} \frac{dx'^\beta}{d\lambda} - \frac{1}{2} g_{\mu\nu, \alpha} x'^\mu x'^\nu = 0, \quad (5.16)$$

which implies

$$g_{\alpha\beta} \frac{d^2 x^\beta}{d\lambda^2} + \frac{1}{2} g_{\alpha\beta, \sigma} x'^\sigma x'^\beta + \frac{1}{2} g_{\alpha\sigma, \beta} x'^\beta x'^\sigma - \frac{1}{2} g_{\beta\sigma, \alpha} x'^\beta x'^\sigma = 0. \quad (5.17)$$

Multiplying this with the inverse metric $g^{\alpha\rho}$, and using (5.5), we obtain

$$\frac{d^2 x^\rho}{d\lambda^2} + \frac{1}{2} g^{\alpha\rho} [g_{\alpha\beta, \sigma} + g_{\alpha\sigma, \beta} - g_{\beta\sigma, \alpha}] \frac{dx^\beta}{d\lambda} \frac{dx^\sigma}{d\lambda} = 0, \quad (5.18)$$

which can be written as

$$\frac{d^2 x^\rho}{d\lambda^2} + \Gamma^\rho_{\beta\sigma} \frac{dx^\beta}{d\lambda} \frac{dx^\sigma}{d\lambda} = 0, \quad (5.19)$$

where the objects $\Gamma^\rho_{\beta\sigma}$ are called the *Christoffel symbols*, with

$$\Gamma^\rho_{\beta\sigma} \equiv \frac{1}{2} g^{\alpha\rho} [g_{\alpha\beta,\sigma} + g_{\alpha\sigma,\beta} - g_{\beta\sigma,\alpha}]. \quad (5.20)$$

The curves which satisfy equation (5.19) are called *geodesics* and are the extremal (shortest) paths.

To find a connection between geodesic curves in a space time with metric $g_{\alpha\beta}$ and force-free motion, we next consider a *force-free* particle in Newtonian mechanics: the Lagrangian function is given by

$$L = \frac{1}{2} m v^2 = \frac{1}{2} m \frac{dx^a}{dt} \frac{dx^b}{dt} g_{ab}, \quad (5.21)$$

and the Euler–Lagrange equations $(d/dt)(\partial L/\partial \dot{x}^a) - \partial L/\partial x^a = 0$ give (the dot denotes differentiation with respect to time t):

$$\frac{d}{dt} \left[m \frac{dx^b}{dt} g_{ab} \right] - \frac{m}{2} g_{ab,r} \frac{dx^r}{dt} \frac{dx^b}{dt} = 0, \quad (5.22)$$

or (cancelling the factors of m)

$$\ddot{x}^b g_{ab} + \dot{x}^b \dot{x}^r g_{ab,r} - \frac{1}{2} g_{rb,a} \dot{x}^r \dot{x}^b = 0. \quad (5.23)$$

This equation can be put into a form of a geodesic equation (5.19), with the rôle of the affine parameter played by the time t itself. This is left as an exercise.

Exercise 5.7 *Deduce the formula for the Christoffel symbols (5.20) from the equations (5.23).*

This is self consistent, given that for a force-free motion in Newtonian mechanics the velocity $v = ds/dt$ is constant (c.f. Newton’s first Law -also seen as a consequence of conservation of energy), and thus the time t is one of the affine parameters λ which is proportional to the arc length s .

What the Newtonian mechanics analogue teaches us is that there is an equivalent way of obtaining the geodesic equation (5.19) by extremizing the square of the proper distance ds^2 instead of ds , viewing the latter as a “lagrangian”. This is more convenient in the general relativistic case, where this square is not necessarily *positive definite*.

To summarize, in this latter method, one considers the Lagrange equations for the Lagrangian (c.f. Appendix A): $\mathcal{L} = \frac{1}{2} g_{\mu\nu} \frac{dx^\mu}{d\lambda} \frac{dx^\nu}{d\lambda}$, viewing the affine parameter λ as the ‘time’ parameter:

$$\frac{d}{d\lambda} \frac{\partial \mathcal{L}}{\partial x'^\mu} - \frac{\partial \mathcal{L}}{\partial x^\mu} = 0, \quad x'^\mu \equiv \frac{dx^\mu}{d\lambda} \quad (5.24)$$

and normalizes - by appropriate manipulations - the coefficient of the $\frac{d^2}{d\lambda^2} x^\mu$ terms to unity, so that the resulting equations acquire the geodesic form (5.19). From such expressions, then, one can read the pertinent Christoffel symbols for the problem at hand.

Exercise 5.8 *By using the Lagrange equation method, compute the geodesic equations for the spacetime described by the line element:*

$$ds^2 = R_0^2 (d\theta^2 + \sin^2 \theta d\phi^2), \quad R_0 = \text{constant}$$

If we choose the parameter λ to be the arc length s itself, then we may recapitulate the results (5.19), (5.23) in the following Law:

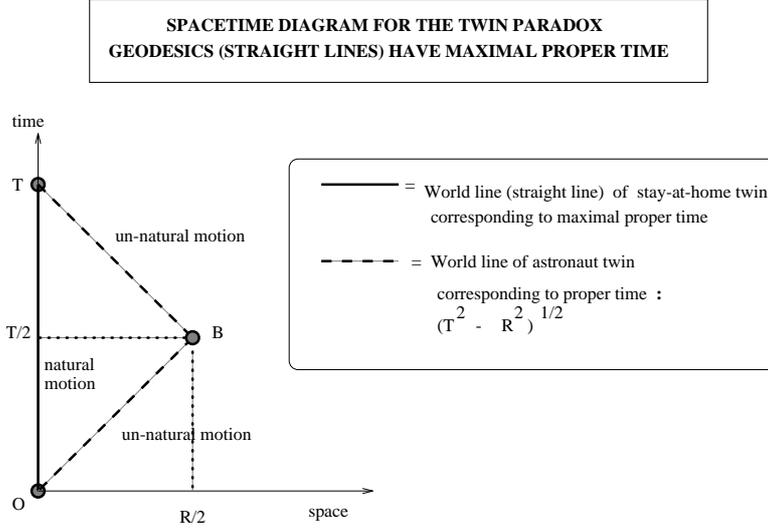


Figure 14: In Special Relativity (flat spacetimes) the worldlines of a particle following a natural motion are straight lines, corresponding to maximal wristwatch (proper) time. These are the geodesics of the flat space time. In curved space times the maximal proper time curves are also geodesics, but are not straight lines.

A force-free particle moves on a geodesic

$$\frac{d^2 x^\rho}{ds^2} + \Gamma^\rho_{\beta\sigma} \frac{dx^\beta}{ds} \frac{dx^\sigma}{ds} = 0, \quad (5.25)$$

of the surface to which it is constrained. Its path, therefore, is the shortest curve between any two-points lying on it.

5.5.2 Einstein Dynamics in Curved Space: Geodesics from Extremal Aging

In contrast to the *minimal* (spacelike) curve which characterized the Newtonian case, in curved spacetimes, in the context of Einsteinian theory, one is dealing with *maximal* timelike worldlines according to the principle of extremal aging covered previously. Formally the two cases can be reconciled if one uses the terminology *extremization of the path*. Indeed the proper time interval connecting an initial (τ_i) and final (τ_f) event is given by

$$\tau = \int_{\tau_i}^{\tau_f} d\tau = \text{extremum} = \int_{\lambda_i}^{\lambda_f} d\lambda \sqrt{-g_{\alpha\beta} \frac{dx^\alpha}{d\lambda} \frac{dx^\beta}{d\lambda}} \quad (5.26)$$

Notice the minus sign inside the square root as compared with the spacelike curve (5.9), which is due to the timelike character of the path. According to the principle of extremal aging (which can be extended to arbitrarily curved spacetimes), a natural motion is one for which the above interval is extremized (maximized) from which one can derive the geodesic equation in a formally similar way as in the Newtonian case studied above.

In the flat spacetime case, which is the space time of Special Relativity, the above extremization implies that the worldline of the particle under consideration, which follows the natural motion of maximal proper time (e.g. the stay-at-home twin in the twin paradox mentioned in section 3, or the stone in figure 5) will be a *straight line* connecting the relevant events (see figure 14, where we

give the spacetime diagram of the twin paradox: From the figure it is clear that the worldline of the astronaut twin, is kinked, and corresponds to a wristwatch (proper) time $\tau^2 = T^2 - R^2$ which is not maximal. In contrast the stay-at-home twin has a maximal proper time and her worldline is straight). Straight lines are the geodesics of flat space times.

5.6 Parallel Transport along a curve and the Covariant Derivative

From the geodesic equations, one can obtain a definition of **parallelism** in an arbitrarily curved space time. Indeed, consider the tangent vector $t^\alpha = dx^\alpha/d\lambda$ of a geodesic curve $x^\mu(s)$ where the affine parameter is identified with the arc length itself. Then, from (5.19) one has:

$$\frac{D}{Ds} t^\mu \equiv \frac{dt^\mu}{ds} + \Gamma^\mu_{\alpha\beta} t^\alpha \frac{dx^\beta}{ds} = 0. \quad (5.27)$$

The middle side of (5.27) defines the *covariant derivative* D/Ds of a contravariant vector, in this particular case the tangent vector. This is an important concept which we now analyze in some detail.

A geometrical construction using **parallel transport** of an arbitrary four vector along a curve $x^\mu(\lambda)$ is given in figure 15. This defines geometrically the *Covariant Derivative* of an arbitrary four vector. One can proceed in this geometrical way to obtain the components of the covariant derivative in an analytic manner, when acting on an arbitrary vector. We shall not do so in this course, but instead we shall state the pertinent analytic formulæ for the covariant derivative, which will be the only thing the reader should remember.

Let T^μ an arbitrary contravariant vector. In components, the covariant derivative of T^μ reads:

$$DT^\mu = dT^\mu + \Gamma^\mu_{\alpha\beta} T^\alpha dx^\beta = \left(T^\mu_{,\alpha} + \Gamma^\mu_{\beta\alpha} T^\alpha \right) dx^\alpha \quad (5.28)$$

where we use the symmetry property of the Christoffel symbol $\Gamma^\alpha_{\mu\nu} = \Gamma^\alpha_{\nu\mu}$.

On the other hand, for a covariant vector the corresponding relation is:

$$DT_\mu = dT_\mu - \Gamma^\alpha_{\mu\beta} T_\alpha dx^\beta = \left(T_{\mu,\alpha} - \Gamma^\nu_{\mu\alpha} T_\nu \right) dx^\alpha \quad (5.29)$$

From these one obtains the following expression for the covariant derivative (denoted from now on by $;\alpha$, as a generalization of the $,\alpha$ which denotes ordinary partial differentiation):

$$\begin{aligned} T^\mu_{;\nu} &= T^\mu_{,\nu} + T^\alpha \Gamma^\mu_{\alpha\nu} , \\ T_{\mu;\nu} &= T_{\mu,\nu} - T_\alpha \Gamma^\alpha_{\mu\nu} , \end{aligned} \quad (5.30)$$

These concepts may be extended to an arbitrary rank tensor in fact. For example, for second rank tensors the components of the corresponding covariant derivatives are:

$$\boxed{\begin{aligned} T^{\mu\nu}_{;\beta} &= T^{\mu\nu}_{,\beta} + T^{\alpha\nu} \Gamma^\mu_{\alpha\beta} + T^{\mu\alpha} \Gamma^\nu_{\alpha\beta} , \\ T_{\mu\nu};\beta &= T_{\mu\nu,\beta} - T_{\alpha\nu} \Gamma^\alpha_{\mu\beta} - T_{\mu\alpha} \Gamma^\alpha_{\nu\beta} , \\ T^\mu_{\nu};\beta &= T^\mu_{\nu,\beta} + T^\alpha_\nu \Gamma^\mu_{\alpha\beta} - T^\mu_\alpha \Gamma^\alpha_{\nu\beta} . \end{aligned}} \quad (5.31)$$

and similarly for higher rank tensors.

It should be noted that for scalars (rank 0 tensors) the covariant derivative coincides with the ordinary one, i.e. one has:

$$\boxed{\phi_{;\alpha} = \phi_{,\alpha}} \quad (5.32)$$

We next remark that the vanishing of the covariant derivative in (5.27) defines the **parallel transport**, and is applicable to any parameter s and vector t^μ (see figure 15). *Two vectors at infinitesimally close points of a curved space time are parallel if and only if the covariant derivative (5.27) vanishes.*

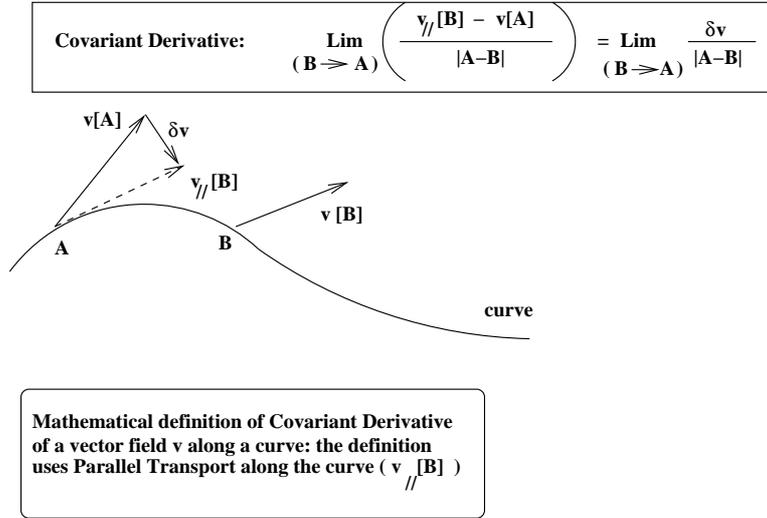


Figure 15: Definition of Covariant Derivative (by means of parallel transport v_{\parallel}^{α}) of the vector v^{α} along the curve $x^{\alpha}(s)$.

The geodesic equation (5.19) is an example of parallelism, given that the tangent vector of a geodesic, $t^{\alpha} = dx^{\alpha}/d\lambda$ remains parallel to itself. *Hence the Newtonian geodesic (5.19) is not only the shortest curve between two points but also the straightest. We remind the reader that in general relativity the concept of ‘shortest curve’ is replaced by that of ‘extremal curve’ for reasons explained previously, however that of ‘straightest curve’ remains.*

One can formally understand the need for introducing the concept of the covariant derivative, and not the simple partial derivative $\partial/\partial x^{\mu}$, in order to define parallelism by the following fact: consider the value of the components of a tensor, e.g. say contravariant of rank n , $T^{i_1 i_2 \dots i_n}$, at two points separated by an infinitesimal distance ds . The difference: $T^{i_1 \dots i_n}(x^i + dx^i) - T^{i_1 \dots i_n}(x^i) = T^{i_1 \dots i_n}_{,i} dx^i$, where we applied a simple Taylor expansion, *is not* a tensor. This is due to the fact that tensors in two different points of a curved space time obey different transformation laws, under a change in coordinates, and hence the above simple difference is not a tensor. On the other hand the use of covariant differentiation defined above, implies that the tensor transformation properties remain intact.

From this it follows that the Christoffel symbol is not a tensor by itself, because it is the full covariant derivatives (5.28), (5.29) that behave properly as tensors in a curved space time.

An important property of the metric tensor is that its covariant derivative vanishes:

$$g_{\alpha\beta}{}_{;\mu} = g^{\alpha\beta}{}_{;\nu} = 0 . \tag{5.33}$$

This property is known as the ‘metricity postulate’.

Exercise 5.9 *Verify the metricity postulate (5.33) above, and show that it is valid in all coordinate systems.*

We are now well equipped to proceed into a quantification of the curvature of space time. We do so in the next subsection by means of the so-called Geodesic Deviation, a concept applying to families of geodesics.

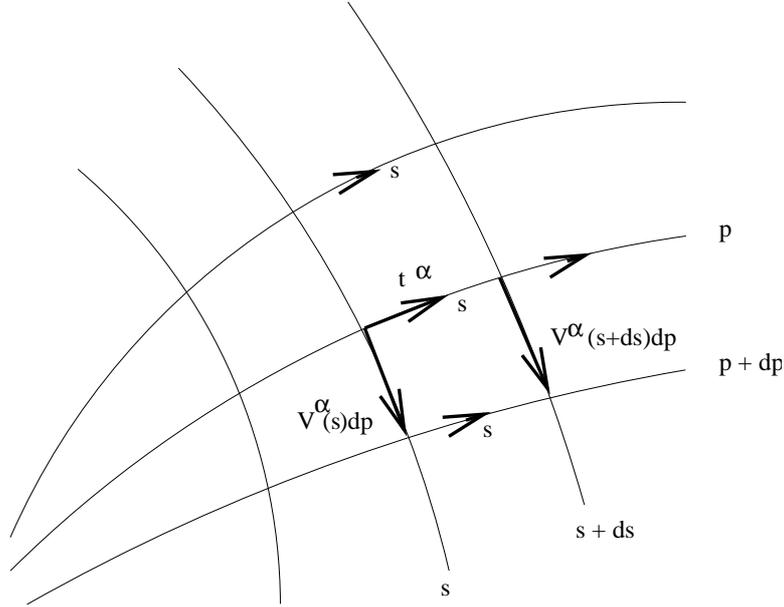


Figure 16: The family of geodesics $x^\alpha(s, p)$.

5.7 Quantifying Space-Time Curvature: The Riemann Curvature Tensor from Geodesic Deviation

We can now proceed to construct a measure for the curvature of space time, the Riemann Curvature Tensor, by making use of the concept of Geodesic Deviation, which we explain below.

Consider families of geodesics as in figure 16 corresponding to affine parameters p and s . The parameter p labels different geodesics, whilst the parameter (arc length) s fixes the different points of the same geodesic. We define the unit tangent vectors by

$$\frac{\partial x^\alpha}{\partial s} = t^\alpha, \quad (5.34)$$

$$\frac{\partial x^\alpha}{\partial p} = v^\alpha. \quad (5.35)$$

The vector t^α is just the tangent vector of a geodesic, whilst the quantity $v^\alpha dp$ is just the displacement vector of two neighbouring geodesics (see figure 16). In the Newtonian case, where the above geodesics are the paths of a force-free particle, with s the time, the unit tangent vector t^α points in the direction of the velocity.

We already know from the previous subsection that geodesics are defined as curves along which a tangent vector remains “parallel” to itself, i.e. its covariant derivative along the geodesic vanishes. We can quantify the curvature of these geodesics, and hence the spacetime, i.e. to see whether we are dealing with a plane, or more generally with a flat space time, or not, by looking at the *second* covariant derivative of the vector v^α . This is analogous to the pertinent concept in real analysis, where the gradient of a curve is its first derivative, and the second derivative quantifies the ‘curvature’. However, as we have explained previously, in curved space times one should not consider ordinary derivatives, but rather the *covariant derivative*, because it is the latter quantity that preserves the tensorial transformation properties.

To understand better the physical meaning of the above considerations we should recall that in a previous subsection we have discussed as evidence for curvature (non flatness) of space time the existence of ‘tidal accelerations’ of particles on nearby geodesics; these relative accelerations

between two ‘particles’ moving on neighbouring geodesics, are precisely encoded in the *second* covariant derivative of the vector v^a with respect to the arc length s ; in fact, $v^a dp$ is just the relative separation of the particles, and ds is proportional to a time differential, since as we have discussed in the previous subsection, in the case of force-free motion, the proper time (which, in the case of Newtonian Mechanics, is identified with the universal Newtonian time) plays the rôle of the affine parameter of a geodesic.

In what follows we shall therefore compute the second covariant derivative $D^2 v^a / Ds^2$, which is a measure of ‘Geodesic Deviation’, and derive from it a mathematical expression for an entity that quantifies the ‘amount of non-flatness’ of space time, the so-called ‘Riemann Curvature Tensor’.

From the definitions (5.34), we have

$$\frac{\partial t^\alpha}{\partial p} = \frac{\partial^2 x^\alpha}{\partial p \partial s} = \frac{\partial}{\partial s} \frac{\partial x^\alpha}{\partial p} = \frac{\partial v^\alpha}{\partial s}. \quad (5.36)$$

From the geodesic equations,

$$\frac{D}{Ds} t^\mu \equiv \frac{dt^\mu}{ds} + \Gamma^\mu_{\alpha\beta} t^\alpha \frac{dx^\beta}{ds} = 0. \quad (5.37)$$

For the covariant derivative with respect to p we have,

$$\frac{D}{Dp} t^\alpha = \frac{dt^\alpha}{dp} + \Gamma^\alpha_{\mu\nu} t^\mu \frac{dx^\nu}{dp}, \quad (5.38)$$

whence

$$\frac{Dt^\alpha}{Dp} = \frac{\partial^2 x^\alpha}{\partial p \partial s} + \Gamma^\alpha_{\mu\nu} t^\mu v^\nu. \quad (5.39)$$

For v^α we have

$$\frac{Dv^\alpha}{Ds} = \frac{dv^\alpha}{ds} + \Gamma^\alpha_{\mu\nu} v^\mu \frac{dx^\nu}{ds} = \frac{\partial x^\alpha}{\partial p \partial s} + \Gamma^\alpha_{\mu\nu} v^\mu t^\nu, \quad (5.40)$$

so that, since $\Gamma^\alpha_{\mu\nu} = \Gamma^\alpha_{\nu\mu}$

$$\frac{Dt^\mu}{Dp} = \frac{Dv^\mu}{Ds}, \quad (5.41)$$

which should be compared with equation (5.36). We now wish to compute the second covariant derivative to measure whether transport along the geodesics is planar, or whether there is deviation:

$$\begin{aligned} \frac{D^2 v^\alpha}{Ds^2} &= \frac{D}{Ds} \frac{Dt^\alpha}{Dp} = \frac{D}{Ds} \left[\frac{\partial t^\alpha}{\partial p} + \Gamma^\alpha_{\mu\nu} t^\mu v^\nu \right] \equiv \frac{D}{Ds} [a^\alpha] \\ &= \frac{\partial}{\partial s} \left[\frac{\partial t^\alpha}{\partial p} + \Gamma^\alpha_{\mu\nu} t^\mu v^\nu \right] + \Gamma^\alpha_{\mu\nu} a^\mu t^\nu \\ &= \frac{\partial^2 t^\alpha}{\partial s \partial p} + \Gamma^\alpha_{\mu\nu,\beta} t^\beta t^\mu v^\nu + \Gamma^\alpha_{\mu\nu} \left[\frac{\partial t^\mu}{\partial s} v^\nu + t^\mu \frac{\partial v^\nu}{\partial s} \right] + \Gamma^\alpha_{\rho\tau} \left[\frac{\partial t^\rho}{\partial p} + \Gamma^\rho_{\mu\nu} t^\mu t^\nu \right] t^\tau. \end{aligned} \quad (5.42)$$

From the geodesic equation (5.37) it follows that

$$\begin{aligned} 0 &= \frac{D}{Dp} \frac{Dt^\alpha}{Ds} = \frac{D}{Dp} \left[\frac{\partial t^\alpha}{\partial s} + \Gamma^\alpha_{\mu\nu} t^\mu t^\nu \right] \\ &= \frac{\partial^2 t^\alpha}{\partial p \partial s} + \Gamma^\alpha_{\mu\nu} \frac{\partial t^\mu}{\partial s} t^\nu + \frac{\partial}{\partial p} [\Gamma^\alpha_{\mu\nu} t^\mu t^\nu] + \Gamma^\alpha_{\rho\sigma} \Gamma^\rho_{\mu\nu} t^\mu t^\nu v^\sigma \\ &= \frac{\partial^2 t^\alpha}{\partial p \partial s} + \Gamma^\alpha_{\mu\nu} \frac{\partial t^\mu}{\partial s} t^\nu + \Gamma^\alpha_{\mu\nu,\beta} t^\mu t^\nu v^\beta + \Gamma^\alpha_{\rho\sigma} \Gamma^\rho_{\mu\nu} t^\mu t^\nu v^\sigma + \Gamma^\alpha_{\mu\nu} \frac{\partial t^\mu}{\partial p} t^\nu + \Gamma^\alpha_{\mu\nu} t^\mu \frac{\partial t^\nu}{\partial p}. \end{aligned} \quad (5.43)$$

Substituting this result into equation (5.43) we obtain

$$\frac{\partial^2 t^\alpha}{\partial s \partial p} = -\Gamma^\alpha_{\mu\nu,\beta} v^\beta t^\mu t^\nu - \Gamma^\alpha_{\mu\nu} \left[\frac{\partial t^\mu}{\partial p} t^\nu + t^\mu \frac{\partial t^\nu}{\partial p} \right] - \Gamma^\alpha_{\rho\tau} \left[\frac{\partial t^\rho}{\partial s} + \Gamma^\rho_{\mu\nu} t^\mu t^\nu \right] v^\tau. \quad (5.44)$$

But we already know (c.f. (5.36)) that $\partial t^\mu / \partial p = \partial v^\mu / \partial s$, hence

$$\frac{\partial^2 t^\alpha}{\partial s \partial p} + \Gamma^\alpha_{\mu\nu,\beta} v^\beta t^\mu t^\nu + \Gamma^\alpha_{\mu\nu} \left[\frac{\partial t^\mu}{\partial s} v^\nu + t^\mu \frac{\partial v^\nu}{\partial s} \right] + \Gamma^\alpha_{\mu\nu} \frac{\partial t^\mu}{\partial p} t^\nu + \Gamma^\alpha_{\rho\tau} \Gamma^\rho_{\mu\nu} t^\mu t^\nu v^\tau = 0, \quad (5.45)$$

and hence

$$\frac{\partial^2 t^\alpha}{\partial s \partial p} + \Gamma^\alpha_{\mu\nu} \left[\frac{\partial t^\mu}{\partial s} v^\nu + t^\mu \frac{\partial v^\nu}{\partial s} \right] + \Gamma^\alpha_{\mu\nu} \frac{\partial t^\mu}{\partial p} t^\nu = -\Gamma^\alpha_{\mu\nu,\beta} v^\beta t^\mu t^\nu - \Gamma^\alpha_{\rho\tau} \Gamma^\rho_{\mu\nu} t^\mu t^\nu v^\tau. \quad (5.46)$$

Substituting these results into equation (5.42) we obtain for the geodesic deviation:

$$\begin{aligned} \frac{D^2 v^\alpha}{Ds^2} &= \Gamma^\alpha_{\mu\nu,\beta} t^\beta t^\mu v^\nu - \Gamma^\alpha_{\mu\nu,\beta} v^\beta t^\mu t^\nu + \Gamma^\alpha_{\rho\beta} \Gamma^\rho_{\mu\nu} t^\mu v^\nu t^\beta - \Gamma^\alpha_{\rho\beta} \Gamma^\rho_{\mu\nu} t^\mu t^\nu v^\beta = \\ &= t^\beta t^\mu v^\nu [\Gamma^\alpha_{\mu\nu,\beta} - \Gamma^\alpha_{\mu\beta,\nu} + \Gamma^\alpha_{\rho\beta} \Gamma^\rho_{\mu\nu} - \Gamma^\alpha_{\rho\nu} \Gamma^\rho_{\mu\beta}]. \end{aligned} \quad (5.47)$$

The object in the brackets is called the Riemann curvature tensor, and is denoted by $R^\alpha_{\mu\beta\nu}$:

$$\boxed{R^\alpha_{\mu\beta\nu} \equiv \Gamma^\alpha_{\mu\nu,\beta} - \Gamma^\alpha_{\mu\beta,\nu} + \Gamma^\alpha_{\rho\beta} \Gamma^\rho_{\mu\nu} - \Gamma^\alpha_{\rho\nu} \Gamma^\rho_{\mu\beta}.} \quad (5.48)$$

In obtaining the object in square brackets we have changed the dummy indices appropriately and taken into account the symmetry of the Christoffel symbols: $\Gamma^\alpha_{\mu\nu} = \Gamma^\alpha_{\nu\mu}$. Note that the Riemann tensor can be remembered more easily as:

$$\boxed{R^\alpha_{\mu\beta\nu} \equiv \Gamma^\alpha_{\mu\nu,\beta} + \Gamma^\alpha_{\rho\beta} \Gamma^\rho_{\mu\nu} - \langle \beta\nu \rangle,} \quad (5.49)$$

where the symbol $\langle \beta\nu \rangle$ indicates that the preceding terms should be repeated with β and ν interchanged.

The curvature tensor gives a measure of the change in the separation of neighbouring geodesics, or in the language of mechanics, the relative acceleration of two particles moving toward one another on neighbouring paths (cf. ‘‘tidal accelerations’’ discussed previously).

More generally, it can be shown that the curvature tensor (5.48) appears as a measure of the *path dependence* of the parallel displacement between two points in a curved space time. In a flat space time the parallel displacement should be independent of the path chosen, but this is not true in a space time with a non-trivial curvature. In that case one may show actually that the Riemann Curvature Tensor is related to the commutator of the covariant derivatives acting on a vector, say, T_μ :

$$\boxed{T_{\mu;\nu;\sigma} - T_{\mu;\sigma;\nu} = R^\alpha_{\mu\nu\sigma} T_\alpha} \quad (5.50)$$

or, more generally, for an arbitrary $\binom{N}{M}$ tensor $T_{\beta\dots}^{\alpha\dots}$:

$$\boxed{T_{\beta\dots;\nu\kappa}^{\alpha\dots} - T_{\beta\dots;\kappa\nu}^{\alpha\dots} = R^\sigma_{\beta\nu\kappa} T_{\sigma\dots}^{\alpha\dots} + \dots - R^\alpha_{\sigma\nu\kappa} T_{\beta\dots}^{\sigma\dots} - \dots} \quad (5.51)$$

where we have omitted similar terms for the other $N - 1$ upper, and $M - 1$ lower indices. The above relations (5.50),(5.51) may also be taken as a *definition* of the curvature tensor.

Exercise 5.10 *By using the definition of the covariant derivative, show that the left-hand-side of (5.50) does indeed give the expression (5.48) for the Riemann Tensor $R^\alpha_{\beta\gamma\rho}$ in terms of Christoffel symbols (Hint: go to a local frame of coordinates, in which the Christoffel symbols vanish, $\Gamma^\alpha_{\rho\sigma} = 0$, but not their first derivatives, $\Gamma^\alpha_{\rho\sigma,\nu} \neq 0$, and prove the validity of this property there. Then, deduce its validity in any frame by covariance reasons, which you should explain).*

5.8 Properties of the Curvature Tensor

From its definition (5.48) one observes the following symmetry properties of the Riemann Curvature tensor:

$$\boxed{R_{\alpha\beta\mu\nu} = -R_{\beta\alpha\mu\nu} = -R_{\alpha\beta\nu\mu} = R_{\mu\nu\alpha\beta}.} \quad (5.52)$$

One also has the relation:

$$\boxed{R_{\alpha\mu\nu\rho} + R_{\alpha\nu\rho\mu} + R_{\alpha\rho\mu\nu} = 0.} \quad (5.53)$$

Because of these symmetry properties it can be shown that in a d -dimensional space time the Riemann Curvature tensor has $d^2(d^2 - 1)/12$ components. This stems from the following (the following proof is not compulsory and may be omitted in first reading): the *antisymmetry* (5.52) of the Riemann tensor $R_{\mu\nu\rho\sigma}$ with respect to the $(\mu\nu)$ and $(\rho\sigma)$ pairs of indices implies that there are $\mathcal{M} = \frac{1}{2}d(d - 1)$ ways of choosing non-trivial pairs $(\mu\nu)$, and similarly \mathcal{M} ways of choosing $(\rho\sigma)$ pairs. Moreover, due to the fact that the Riemann tensor is *symmetric* with respect to the interchange of pairs $(\mu\nu)$ and $(\rho\sigma)$, there are $\frac{1}{2}\mathcal{M}(\mathcal{M} + 1)$ independent ways of choosing $\mu\nu\rho\sigma$ when the pair symmetries are considered. Finally, the cyclic symmetry (5.53), furnishes a number of extra constraints, which equals the number of combinations of 4 objects from d objects, i.e. $\binom{d}{4} = \frac{d!}{4!(d-4)!}$ (notice that this formula gives zero for $d < 4$ as it should, since in that case the cyclic symmetry (5.53) gives no additional constraints). Thus, the number of independent components of the Riemann tensor in d -dimensional spacetimes is:

$$\frac{1}{2}\mathcal{M}(\mathcal{M} + 1) - \binom{d}{4} = \frac{d^2(d^2 - 1)}{12} \quad (5.54)$$

Thus, a $d = 1$ -dimensional space is always flat, because the Riemann tensor has 0 independent components, in $d = 2$ -dimensional space time the Riemann tensor has only one independent component, whilst for the physically interesting case of four-dimensional space time the Riemann tensor has a maximum of 20 algebraically independent components.

Another important property of the Riemann Curvature tensor that we simply list here (without proof) is the set of so-called *Bianchi identities*

$$\boxed{R_{\mu\nu\rho\sigma;\tau} + R_{\mu\nu\sigma\tau;\rho} + R_{\mu\nu\tau\rho;\sigma} = 0.} \quad (5.55)$$

The reader is invited to compare this identity, with the corresponding Bianchi identity (3.48) of Electromagnetism ².

Exercise 5.11 *Verify the properties (5.52) and (5.53) of the Riemann Curvature Tensor.*

From the Riemann tensor one may define two more physically important tensors by contracting some of its indices. One is a second rank symmetric covariant tensor, called the *Ricci tensor*:

$$\boxed{R_{\mu\nu} = R_{\nu\mu} = R_{\mu\alpha\nu}^{\alpha} = -R_{\mu\nu\alpha}^{\alpha}} \quad (5.56)$$

and the other is a scalar, called the *curvature scalar*

$$\boxed{R = g^{\mu\nu} R_{\mu\nu}} \quad (5.57)$$

which is therefore an invariant (under a change of coordinates) characterization of the geometry of space time. The symmetry of the Ricci tensor follows from the symmetry properties of the Riemann Curvature tensor.

²Although it is not part of the undergraduate course, it may be useful to mention that the Maxwell tensor plays actually the rôle of the 'curvature' of the electromagnetic potential A_{μ} , which is defined as $F_{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu}$.

One may obtain useful identities for the Ricci tensor, and curvature scalar, which we shall make use of in later sections of these notes, when we discuss Einstein's equations. By contracting appropriately the Bianchi identities (5.55), using the covariant constancy of the metric (5.33), one obtains:

$$R_{\nu\rho\sigma;\tau} + R_{\nu\rho;\sigma\tau} - R_{\nu\sigma;\rho\tau} = 0 \quad (5.58)$$

and

$$\boxed{\left(R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R\right)_{;\nu} = 0} \quad (5.59)$$

This last identity will be very important in guiding us to the correct equations that link gravity and matter, as we shall see in the next section.

Exercise 5.12 Determine the curvature scalar of a two-dimensional unit sphere.

Exercise 5.13 Determine the curvature scalar of the two-dimensional metric described by the line element:

$$ds^2 = dv^2 - v^2 du^2 \quad (5.60)$$

Exercise 5.14 Determine the curvature scalar of the two-dimensional space time described by the following metric line element:

$$ds^2 = -(1 - 2M/r)dt^2 + (1 - 2M/r)^{-1}dr^2 \quad (5.61)$$

where M is a constant. What do you observe for the behaviour of the curvature scalar?

Exercise 5.15 Starting from the Bianchi identities (5.55), and explaining carefully what contractions you make, prove the Bianchi identities (5.58), (5.59).

Exercise 5.16 Show that for an arbitrary tensor $A^{\mu\nu}$ of rank $\binom{2}{0}$ the following relation is true:

$$A^{\mu\nu}{}_{;\mu\nu} = A^{\mu\nu}{}_{;\nu\mu}$$

6 Einstein's Equations: the interplay between matter and gravity

6.1 Physics in curved spacetimes

Having formulated mathematically the concepts of *natural motion* of a particle in a *curved spacetime* by means of *geodesics*, and that of the *curvature* of spacetime by means of the Riemann Curvature Tensor, we are now well equipped to proceed with a formulation of the *dynamical* equations that would describe Einstein's theory of Gravitation and its interaction/relation with matter.

As we have already mentioned briefly at the beginning of the last section, according to Einstein, a non-trivial distribution of matter results in a non-trivial curvature of spacetime. This result is quantified in the so-called Einstein's equations for Gravity which lie at the core of General Relativity.

Before writing these equations down, it will be useful to summarize some of the basic notions of physics in curved spacetimes, which can be inferred from our studies so far:

1. Spacetime, which is defined as the set of all events, is a four-dimensional manifold endowed with a metric.

2. The metric is measurable by rods and clocks. For instance, the distance along a rod of *infinitesimal length* is given by the inner product $ds = \sqrt{g_{\mu\nu}dx^\mu dx^\nu}$, where $g_{\mu\nu}$ is the metric tensor (a symmetric rank $\binom{0}{2}$ invertible tensor). On the other hand, the time measured by a clock that experiences two events *closely separated in time* is given by the inner product $d\tau = \sqrt{-g_{\mu\nu}dx^\mu dx^\nu}$.
3. *Locally* in spacetime, i.e. within a *sufficiently small* coordinate patch, one can always find an appropriate frame of coordinates for which the spacetime will look *flat*, i.e. the metric can be put in the Minkowski (or Lorentz) form $\eta_{\mu\nu}$. This statement is not true globally, and evidence for this fact is provided by the *tidal accelerations* due to the non-uniformity of the gravitational field for large enough coordinate patches. *Globally* spacetime is the result of appropriately patching together the various local coordinate frames, and the way to do this is encoded in the dynamical equations underlying the theory of General Relativity.
4. Free falling massive particles in curved spacetimes move on *timelike* geodesics of the spacetime. Massless particles move on *lightlike* geodesics.
5. Any physical law which could be expressed in tensor notation in Special Relativity retains exactly the same form in a locally inertial (free-floating) frame of a curved spacetime. This is an alternative version of the *strong equivalence principle*, which is a very basic principle of General Relativity.

Let us see now how these facts help us in understanding the curved-spacetime formalism. Consider the strong equivalence principle, last statement above, in connection with the form of certain properties of the stress-energy tensor of fluids in curved spacetimes. The conservation law (3.35) of Special Relativity, which applies to flat spacetimes, can be generalized in curved spacetimes by replacing the ordinary partial derivative by the *covariant derivative*,

$$\boxed{T^{\mu\nu}{}_{;\nu} = 0} \tag{6.1}$$

This is actually a general rule, which stems directly from the strong form of the equivalence principle, which can be remembered by the name *the comma goes to semicolon* rule, for reasons that are obvious.

For a perfect fluid, in a curved spacetime corresponding to metric $g_{\mu\nu}$, in a coordinate system which moves with respect to the MCRF with four-velocity u^μ , the expression for $T_{\mu\nu}$ is obtained from (3.33) by replacing the Minkowski metric $\eta_{\mu\nu}$ by the curved one $g_{\mu\nu}$. This is the only formal change, i.e., one can write:

$$T^{\mu\nu} = pg^{\mu\nu} + (p + \rho)u^\mu u^\nu \tag{6.2}$$

where ρ, p denote the energy density and pressure respectively.

6.2 Einstein's Equations

6.2.1 Newton's Theory revisited

In this subsection we shall present the reader with a heuristic 'derivation' of the equations that link the distribution of matter to the curvature of spacetime, the so-called Einstein's equations. These equations are central to the theory of General Relativity, and their solutions, which give various configurations of the gravitational field, once the matter part is known, will be the basis for our discussion in the remaining part of the course.

To understand physically the equations we start once again by the Newtonian theory of gravity, which although inadequate to describe the correct theory of gravity, however includes various elements that are crucial in the derivation of the correct equations in the context of the general theory of relativity. According to Newton, the acceleration of a particle in a gravitational field,

corresponding to a potential $\Phi(\vec{x}) = -G_N m \frac{\vec{x}-\vec{y}}{|\vec{x}-\vec{y}|^3}$, due to a point mass m at position \vec{x} in space, is given by (cf. (2.3)):

$$\ddot{\vec{x}} = -\vec{\nabla}\Phi(\vec{x}). \quad (6.3)$$

It should be noted that this relation is also valid in case one has a distribution of density $\rho(\vec{y})$, and not just a single mass m . In that case Newton's law of gravitation reads: $\ddot{\vec{x}} = -G_N \int d^3y \rho(\vec{y}) \frac{\vec{x}-\vec{y}}{|\vec{x}-\vec{y}|^3}$, which is formally can be cast in the form (6.3), if we recall the result from the calculus course that $\vec{\nabla}_x \left(\frac{1}{|\vec{x}-\vec{y}|} \right) = -\frac{\vec{x}-\vec{y}}{|\vec{x}-\vec{y}|^3}$. From the fact that $\nabla^2 \left(\frac{1}{|\vec{x}-\vec{y}|} \right) = -4\pi\delta^{(3)}(\vec{x}-\vec{y})$, one obtains that in Newtonian theory the potential $\Phi(x)$ obeys the following differential equation:

$$\nabla^2\Phi(\vec{x}) = 4\pi G_N \rho(\vec{x}) \quad (6.4)$$

The relations (6.3), (6.4) are also used as *definition* of Newton's theory of gravity. As we shall see they play a very important rôle in our context, because: (i) they will guide us in writing down the corresponding equations that define Einstein's theory of Gravity, and (ii) they help in checking whether Einstein's theory has a limit in which the Newtonian theory is recovered, which should happen for distances far away from the centres of gravitational attraction, given that Newton's theory seem to work pretty well for all practical purposes for such far away cases.

6.2.2 Einstein's equations

This subsection consists of some simple-minded guesswork, which provides a 'heuristic' "derivation" of Einstein's equations. Note that there is no rigorous proof of these equations, and this guesswork, which we shall outline below, was more or less the original derivation of the gravitational equations by Einstein. It can be shown (c.f. sub-section 6.2.3) that once derived these equations can be obtained from an elegant, and invariant - under general coordinate transformations - action, the so-called *Einstein-Hilbert action*, via an action principle (c.f. Appendix A for basic concepts). Instead, in what follows in this subsection we shall try to justify Einstein's equations by heuristic methods.

From Newton's theory (6.3), (6.4) it becomes clear that the equations for the fundamental degree of freedom of gravitation, the potential $\Phi(\vec{x})$, are linear, and have the energy (mass) density ρ on their right-hand side. In Einstein's theory of gravity, one has a second rank tensor field, the metric $g_{\mu\nu}$ which is the fundamental degree of freedom of gravitation. Moreover, as we have seen from our analysis of fluids, the energy density of a fluid is part of a symmetric second rank tensor, the energy momentum tensor T . If the theory of gravitation, therefore, is to be covariant, i.e. exhibits the correct transformation properties between coordinates frames, which should be expected, one is tempted to generalize Newtonian gravity by describing the matter contribution on the right-hand-side of the pertinent equations by the stress-energy tensor $T_{\mu\nu}$ (or $T^{\mu\nu}$, depending on whether one uses covariant or contravariant tensors). Thus, the equations should look like $\mathcal{O}_{\mu\nu} = T_{\mu\nu}$ (we adopt the covariant notation for definiteness, the analysis is similar in the contravariant case). The tensor $\mathcal{O}_{\mu\nu}$ must be a second rank covariant symmetric tensor, otherwise the equations would not have the correct transformation properties (recall that the stress tensor is symmetric). An important restriction in the choice of this tensor is provided by the *conservation law* of matter (6.1). Taking into account the Newtonian form, as well, according to which there are two derivatives acting on the potential, one may attempt to generalize these equations by involving the symmetric Ricci and scalar curvature tensors. This is so because either these tensors are appropriate second rank tensors (Ricci) involving two derivatives of the metric tensor, or (in the case of the scalar) appropriate second rank tensors can be constructed from them by multiplication with the metric. Indeed, by choosing the following *linear* combination of second-rank covariant tensors: $\mathcal{O}_{\mu\nu} = AR_{\mu\nu} + Bg_{\mu\nu}R + \Lambda g_{\mu\nu}$, where A, B, Λ are constants, and taking into account the covariant constancy of the metric (5.33), and the Bianchi identities (5.59), we observe that one is forced to choose the ratio of the constants $A/B = -1/2$.

From this we are led to the following generic form for the equations of Einstein's gravitation and its interaction with matter:

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + g_{\mu\nu}\Lambda = \kappa T_{\mu\nu} \quad (6.5)$$

where $G_{\mu\nu}$ defined the so-called *Einstein tensor*, of rank $\binom{0}{2}$. The constant Λ is in fact *arbitrary*, and is called the *Cosmological Constant*, to which we shall return when we do cosmology. The constant κ can be determined by requiring that the equations have the correct Newtonian limit, i.e. reproduce in some appropriate limit of weak gravitational fields, and low velocities, the Newtonian equations (6.3),(6.4). A precise determination of the constant κ will be given in the next chapter.

6.2.3 Einstein's equations as field equations from an action

The equations (6.5) can be derived from a generally covariant action, which has as a *symmetry* (i.e. the *action* remains *invariant* under) the general coordinate transformations

$$x^\mu \rightarrow x'^\mu(x^\nu) . \quad (6.6)$$

The action is called the Einstein-Hilbert action, in honour of the German mathematician D. Hilbert who first proposed it, and assumes the form:

$$S_G + S_{matter} = \frac{1}{\kappa} \int d^4x \sqrt{-g} (R - 2\Lambda) + \int d^4x \sqrt{-g} \mathcal{L}_{matter} \quad (6.7)$$

where $g \equiv \text{Det}(g_{\mu\nu})$ denotes the determinant of the gravitational field $g_{\mu\nu}(x)$. The $-g$ notation denotes positive definite quantities, so that the square root is well defined. Notice that d^4x is not invariant under general coordinate transformations, and the appropriate infinitesimal proper volume element, which guarantees invariance under the transformations (6.6) necessitates multiplication by $\sqrt{-g}$.

The quantity S_{matter} (\mathcal{L}_{matter}) denotes the ‘‘Matter’’ action (Lagrangian), which includes the quantum fields in the theory that are not of gravitational nature, i.e. this part could be the lagrangian of the Standard Model of elementary particle physics, but placed in the gravitational background, that is contracting indices with $g_{\mu\nu}$ and replacing ordinary derivatives of the flat-space-time formalism by gravitational covariant derivatives (5.30), (5.31) *etc.* Thus the matter action depends on the gravitational field and its derivatives (through the Christoffel symbol entering the gravitational covariant derivatives). For example, if one considers as matter Lagrangian the Maxwell Lagrangian of Electromagnetism, the resulting form will read:

$$\mathcal{L}_{Maxwell} = -\frac{1}{4}F_{\mu\nu}F_{\rho\sigma}g^{\mu\rho}g^{\nu\sigma} , \quad (6.8)$$

with $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ the (antisymmetric) Maxwell field-strength tensor of the electromagnetic potential A_μ . Notice that, in view of the symmetry of the Christoffel symbol $\Gamma_{\mu\nu}^\alpha = \Gamma_{\nu\mu}^\alpha$, the covariant derivatives acting on the electromagnetic potential in the antisymmetric $F_{\mu\nu}$ above give their position to ordinary derivatives. Fermion (spinor) fields $\psi(x)$ are included by considering the appropriate quantum electrodynamics action ‘‘covariantised’’, i.e. replacing ordinary derivatives of flat space time by the covariant derivatives in the presence of the gravitational field, following (5.30),(5.31) *etc.* The resulting lagrangian reads:

$$\mathcal{L}_{QED-fermions} = \bar{\psi}\gamma^\mu (\partial_\mu + \Gamma_{\mu\nu}^\nu - eA_\mu) \psi \quad (6.9)$$

where e is the electric charge and we work in unit of $c = 1$. The Clifford algebra in curved space-time satisfied by the Dirac γ^μ matrices is given by:

$$\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu} .$$

To derive Einstein's equations (6.5) from the action (6.7) we consider infinitesimal variations of (6.7) with respect to arbitrary variations of the gravitational field, $\delta g_{\mu\nu}$, satisfying Hamilton's principle of least action (c.f. Appendix A for a review of the relevant basic concepts), i.e.

$$0 = \delta(S_G + S_{matter}) = \int d^4x \delta g^{\mu\nu} \left(\frac{\delta(\sqrt{-g}R)}{\delta g^{\mu\nu}} + \frac{\delta(\sqrt{-g}\mathcal{L}_{matter})}{\delta g^{\mu\nu}} \right) \quad (6.10)$$

where the notation δ denotes here variation with respect to the gravitational field. Assuming that the gravitational and matter field configurations are such that asymptotically in space and time there are no contributions to the action, that is that the space-time boundary terms vanish, we have:

$$\begin{aligned} \delta\sqrt{-g} &= -\frac{1}{2\sqrt{-g}}\delta g = -\frac{1}{2\sqrt{-g}}g g_{\mu\nu}\delta g^{\mu\nu} = -\frac{1}{2}\sqrt{-g}g_{\mu\nu}\delta g^{\mu\nu}, \\ \delta R &= \delta g^{\mu\nu}R_{\mu\nu} + g^{\mu\nu}\delta R_{\mu\nu} \end{aligned} \quad (6.11)$$

It is straightforward to see that the variation of the Ricci tensor yields total covariant derivative terms:

$$\delta R_{\mu\nu} \equiv \delta R_{\mu\alpha\nu}^\alpha = (\delta\Gamma_{\nu\mu}^\rho)_{;\rho} - (\delta\Gamma_{\rho\mu}^\rho)_{;\nu} \quad (6.12)$$

Hence,

$$\delta R = R_{\mu\nu}\delta g^{\mu\nu} + (g^{\mu\nu}\delta\Gamma_{\nu\mu}^\sigma - g^{\mu\sigma}\delta\Gamma_{\rho\mu}^\rho)_{;\sigma} \quad (6.13)$$

The last term on the right-hand side of (6.13) yields zero contributions to the variation of the action, since, being a total derivative, it could only contribute surface terms proportional to the *variation* $\delta g_{\mu\nu}$ *at infinity*, which *vanishes* by assumption, as stated previously.

Thus, the variation of the action with respect to the gravitational field $g_{\mu\nu}$ yields the local equations (since the variations $\delta g_{\mu\nu}$ are arbitrary):

$$\frac{1}{\kappa} \left(R_{\mu\nu} - \frac{1}{2}g_{\mu\nu} + \Lambda g_{\mu\nu} \right) = -\frac{2}{\sqrt{-g}} \frac{\delta(\sqrt{-g}\mathcal{L}_{matter})}{\delta g^{\mu\nu}} \equiv T_{\mu\nu} \quad (6.14)$$

which yields the equations (6.5), with the appropriate definition of the stress-energy tensor of matter, $T_{\mu\nu}$, in terms of the gravitational-field variations of the covariant matter Lagrangian.

In the next subsection we proceed to determine the constant κ in terms of the Gravitational (Newton) constant G_N , by requiring agreement of the general relativity theory with the Newtonian dynamics in the non-relativistic weak-gravitational-field limit (Newtonian limit), with $\Lambda = 0$. *Locally*, the cosmological constant term is negligible, and this allows linearization about the flat space time, which is important in yielding the Newtonian dynamics as a limiting case. It must be stressed that the existence of a Newtonian limit locally provides a non-trivial consistency check of Einstein's theory of General Relativity. We remark already at this stage that in the global (cosmological) case with a non-zero cosmological constant $\Lambda > 0$, one cannot linearise about flat Minkowski space time, because the resulting de Sitter space is not asymptotically flat, as we shall discuss later on, in section 9.3.

6.2.4 Weak gravitational fields, the Newtonian limit and the final form of Einstein's equations: determining the constant κ

Considering *weak gravitational fields*, in the case $\Lambda = 0$, means that the spacetime metric differs only marginally from the flat Minkowski metric $\eta_{\mu\nu}$ (3.14), i.e.

$$\boxed{g_{\mu\nu} \simeq \eta_{\mu\nu} + h_{\mu\nu}} \quad (6.15)$$

where $|h_{\mu\nu}| \ll 1$, and hence quadratic, and higher order, terms in h are ignored from the pertinent expressions. The quantity $h_{\mu\nu}$ is called a (weak) *perturbation* of the flat Minkowski metric. Such

fields may be the gravitational fields generated by a distribution of matter which is *far away* from the region of space on which the weak field is measured. It is in this limit that Einstein's theory reduces formally to that of Newton, but as we shall see this reduction is only formal, since there are important conceptual differences between the two theories. When substituting the approximation (6.15) in the relevant expressions, indices are raised and lowered with the Minkowski metric $\eta_{\mu\nu}$ alone (this is an *approximation*).

When writing down Einstein's equations (6.5) with $\Lambda = 0$ in the case of weak fields, one should take into account that any trace of $h_{\mu\nu}$ in the matter part disappears. In this sense, it is sufficient to replace the covariant conservation law (6.1) by the flat space time conservation law (3.35). Any covariant derivative part would correspond to higher order terms in the perturbation $h_{\mu\nu}$.

With these in mind, we now proceed to analyse the form of the $\Lambda = 0$ equations (6.5) in the weak-field limit. First we need an approximate expression for the Riemann Curvature tensor to first order in $h_{\mu\nu}$ which follows from the definition (5.48) and the weak field approximation (6.15):

$$R_{\alpha\beta\mu\nu} \equiv g_{\alpha\rho} R^{\rho}{}_{\beta\mu\nu} = \frac{1}{2} (h_{\alpha\nu,\beta\mu} + h_{\beta\mu,\alpha\nu} - h_{\alpha\mu,\beta\nu} - h_{\beta\nu,\alpha\mu}). \quad (6.16)$$

In this approximation the inverse of the metric is given by

$$g^{\mu\nu} = \eta^{\mu\nu} - h^{\mu\nu} \quad (6.17)$$

this is because, ignoring terms of order h^2 ,

$$g_{\mu\nu} g^{\nu\lambda} = \delta_{\mu}^{\lambda} + \mathcal{O}(h^2).$$

In writing down the Einstein's tensor $G_{\mu\nu}$, upon the linearising approximation (6.15), one arrives at an expression that formally has more terms than in the original Einstein's equations, and looks a bit awkward. To remedie this we define the quantity

$$\bar{h}_{\mu\nu} \equiv h_{\mu\nu} - \frac{1}{2} \eta_{\mu\nu} h^{\alpha}{}_{\alpha}. \quad (6.18)$$

from which it follows that:

$$\bar{h}^{\alpha}{}_{\alpha} = -h^{\alpha}{}_{\alpha}, \quad h_{\mu\nu} = \bar{h}_{\mu\nu} - \frac{1}{2} \eta_{\mu\nu} \bar{h}^{\alpha}{}_{\alpha}. \quad (6.19)$$

In terms of $\bar{h}_{\mu\nu}$, and to leading (first) order in \bar{h} , the einstein tensor $G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R$ is written as:

$$G_{\mu\nu} = \frac{1}{2} \left(-\bar{h}^{\alpha}{}_{\mu\nu,\alpha} - \eta_{\mu\nu} \bar{h}^{\alpha\beta}{}_{,\alpha\beta} + \bar{h}^{\alpha}{}_{\mu\alpha,\nu} + \bar{h}^{\alpha}{}_{\nu\alpha,\mu} \right) \quad (6.20)$$

It can be shown that $h_{\mu\nu}$, and thus, $\bar{h}_{\mu\nu}$ transform as tensors in flat spacetime under Lorentz coordinate transformations $x^{\alpha} \rightarrow \Lambda^{\alpha}_{\beta} x^{\beta}$.

Exercise 6.1 *The Lorentz transformation in flat spacetimes, $x^{\mu} \rightarrow \Lambda^{\mu}_{\nu} x^{\nu}$ is designed in such a way that $\Lambda^{\mu}_{\alpha} \Lambda^{\nu}_{\beta} \eta^{\alpha\beta} = \eta^{\mu\nu}$, where $\eta^{\mu\nu}$ is the Minkowski metric. Using this property show that $h_{\mu\nu}$ defined in (6.15) transforms as a rank $\binom{0}{2}$ tensor under Lorentz transformations.*

Another important property of (6.15) is that its form is preserved under a small change in coordinates (infinitesimal general coordinate transformations, which are more general than Lorentz), $x^{\alpha} \rightarrow x^{\alpha} + \xi^{\alpha}(x^{\beta})$, $|\xi^{\alpha}| \ll 1$:

$$h_{\mu\nu} \rightarrow h_{\mu\nu} - \xi_{\mu,\nu} - \xi_{\nu,\mu}, \quad \xi_{\mu} = \eta_{\mu\nu} \xi^{\nu} \quad (6.21)$$

These transformations are called *gauge transformations*, since they bear a strong resemblance to the gauge transformations of electromagnetism. Although a detailed use of such transformations

will not be part of the undergraduate course, however it should be mentioned that their implementation simplifies enormously the gravitational field equations in many circumstances. An example of this will be discussed briefly below, when we analyse gravitational wave propagation.

With the help of the above gauge transformations, it can be shown that one can choose a local coordinate system (i.e. choose the vectors ξ^α), in which (we shall not prove this):

$$\boxed{\bar{h}'_{\nu,\mu}{}^\mu = 0.} \quad (6.22)$$

This is called the *Lorentz gauge*, again due to an analogy with electromagnetism.

What is important to realize is that this gauge is actually a *class* of gauges. Indeed, suppose one has chosen a coordinate system $x^\mu \rightarrow x^\alpha + \eta^\alpha(x)$, in which $\bar{h}_{,\nu}^{(\text{old})\mu\nu} = 0$. Under a further change of coordinates $x^\mu + \eta^\mu \rightarrow x^\mu + \eta^\mu + \xi(x)^\mu$ one has:

$$\bar{h}_{\mu\nu}^{(\text{new})} = \bar{h}_{\mu\nu}^{(\text{old})} - \xi_{\mu,\nu} - \xi_{\nu,\mu} + \eta_{\mu\nu}\xi_{,\alpha}^\alpha \quad (6.23)$$

from which

$$\bar{h}_{,\nu}^{(\text{new})\mu\nu} = \bar{h}_{,\nu}^{(\text{old})\mu\nu} - \xi^{\mu,\nu}{}_{,\nu} \quad (6.24)$$

But since, $\bar{h}_{,\nu}^{(\text{old})\mu\nu} = 0$, one observes that $\bar{h}_{,\nu}^{(\text{new})\mu\nu} = 0$, provided we choose a (non unique) vector ξ^α such that

$$\xi^{\mu,\alpha}{}_{,\alpha} = 0. \quad (6.25)$$

Thus, the Lorentz gauge is a class of gauges (coordinate systems).

In *this class of coordinate systems* the Einstein tensor $G_{\mu\nu}$ (6.20) becomes:

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = -\frac{1}{2}\bar{h}_{\mu\nu,\alpha}{}^\alpha \equiv -\frac{1}{2}\partial^\kappa\partial_\kappa\bar{h}_{\mu\nu} \equiv -\frac{1}{2}\square\bar{h}_{\mu\nu}, \quad (6.26)$$

where in all the above formulæ we have dropped terms of order h^2 and higher. The symbol \square is called the *d'Alembertian*, and as you will recall from formula (3.17) of your tensor calculus notes, it is defined by

$$\square \equiv \partial_\mu\partial^\mu = -\frac{\partial^2}{\partial t^2} + \nabla^2. \quad (6.27)$$

In this coordinate system Einstein's equations (6.5) reduce to

$$\boxed{\square\bar{h}_{\mu\nu} = -2\kappa T_{\mu\nu}.} \quad (6.28)$$

Exercise 6.2 Prove equations (6.26) and (6.28) using (6.22) without proof.

In the Newtonian limit, the gravitational field is assumed weak enough so that it can only result in very low velocities, $|\vec{v}| \ll 1$ (in units where $c = 1$), which implies that in this limit the components of the stress-energy tensor of matter $T^{\mu\nu}$ have the following hierarchy:

$$|T^{00}| \gg |T^{0i}| \gg |T^{ij}|. \quad (6.29)$$

This follows from the physical interpretation of the various components of the tensor T mentioned in previous sections. In particular, since T^{0i} is the momentum density, it is proportional to the velocity in the Newtonian (non-relativistic) limit, while the energy density $T^{00} \equiv \rho$ is essentially independent (or at least contains parts that are independent) of the (non-relativistic) velocity. This implies the first inequality above. Similarly, the components T^{ij} are proportional to the second power of the (non-relativistic) velocity, and hence they are negligible compared with T^{0i} in this limit.

From equation (6.28) this implies that in the Newtonian limit the dominant component of $\bar{h}_{\mu\nu}$ is \bar{h}_{00} . For fields that change only because the sources are moving with (non-relativistic) velocity v , $\partial/\partial t$ is of the same order as $v\partial/\partial x$ and hence $\square \simeq \nabla^2 + \mathcal{O}(v^2\nabla^2)$. Therefore the Newtonian limit is described by the 00-component of Einstein's equations, which now reduces to

$$\nabla^2 \bar{h}_{00} = -2\kappa\rho. \quad (6.30)$$

To determine κ we must compare (6.30) with (6.4). It is obvious from such a comparison that

$$\bar{h}_{00} \propto -\Phi, \quad \text{and} \quad \kappa \propto G_N.$$

There is one more ingredient that is necessary to take into account in our comparison and this comes from the first of Newton's equations (6.3). To this end, we first notice that the condition (6.19) implies that to the order in h (or \bar{h}) we are working

$$\bar{h}_\alpha^\alpha = -h_\alpha^\alpha = -\bar{h}^{00}, \quad (6.31)$$

since all the other components of $\bar{h}_{\mu\nu}$ are negligible. From (6.19) this implies:

$$h_{00} \simeq \frac{1}{2}\bar{h}_{00}, \quad h_{xx} \simeq h_{yy} \simeq h_{zz} \simeq \frac{1}{2}\bar{h}_{00} \quad (6.32)$$

Recalling that the non-diagonal components $h_{\mu\nu} = \bar{h}_{\mu\nu}$ (6.19), we observe that such non-diagonal terms are not dominant. Hence, from (6.32) one arrives at the following spacetime, describing the non-relativistic, weak-field (Newtonian) limit of Einstein's equations:

$$ds^2 = -(1 - \frac{1}{2}\bar{h}_{00})dt^2 + (1 + \frac{1}{2}\bar{h}_{00})(dx^2 + dy^2 + dz^2) \quad (6.33)$$

Exercise 6.3 Compute the Christoffel symbols, and the associated geodesics of the metric (6.33). Show that the geodesics reduce to the Newtonian equation (6.3) upon the identification:

$$\bar{h}_{00} = -4\Phi \quad (6.34)$$

where Φ is the Newtonian gravitational potential.

Exercise 6.4 Show that the non-vanishing components of the Riemann tensor in the Newtonian limit of General Relativity are given in terms of the Newtonian potential Φ :

$$R_{0j0}^i = -\frac{\partial^2 \Phi}{\partial x^i \partial x^j} \quad (6.35)$$

where $i, j = 1, 2, 3$ are spatial indices.

From this, and taking into account (6.30), it follows that we must identify

$$\kappa = 8\pi G_N \quad (6.36)$$

where G_N is Newton's gravitational constant.

The system of equations (6.15), (6.30), (6.22) and (6.36) constitutes what is called *the weak-field or Linearized theory of Gravitation*. We shall see that this theory is important in that it allows us to get useful information from distant gravitational sources, for instance one can predict *gravitational waves* this way. This will be done in the next subsection.

Having determined κ through the linearised-gravitation approach for the $\Lambda = 0$ (local) case, we next assume – following Newton and Einstein – that κ is a *universal constant* and hence its value can be applied to the general (global) $\Lambda \neq 0$ case, that characterises the whole Universe. We therefore arrive at the following final form of Einstein's equations:

$$\boxed{G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + g_{\mu\nu}\Lambda = 8\pi G_N T_{\mu\nu}} \quad (6.37)$$

where $T_{\mu\nu}$ is the stress-energy tensor of matter which is thus responsible for curving spacetime. Note that the equations (6.39) have been written with the indices down. They can also be written in precisely the same form but with the indices up. The stress tensor is *covariantly* constant, as we have discussed before:

$$\boxed{T^{\mu\nu}{}_{;\nu} = 0} \quad (6.38)$$

As mentioned previously, this equation follows from Einstein's equations by taking the covariant derivative on both sides, on account of general properties of the Riemann tensor. It is therefore *not an independent* equation. The equations (6.37) (and the conservation equation (6.38), but not as an independent equation) constitute the *fundamental equations* for Einstein's Theory of Gravitation, and describe the dynamical interplay/link between matter and curvature of spacetime in an elegant geometric formalism. The interplay between geometry of spacetime and matter dynamics is the important conceptual contribution of Einstein compared with Newtonian gravitation. The rest of this course will deal with solutions (approximate or exact) of these equations.

In their original form, the cosmological constant $\Lambda = 0$, and this is what we shall assume in the following sections until we study cosmology, when we shall come back to the issue of $\Lambda \neq 0$. In this special case, of *zero cosmological constant*, Einstein's equations read:

$$\boxed{G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi G_N T_{\mu\nu}} \quad (6.39)$$

In the next subsection we shall discuss a time dependent solution of these equations in the limit of weak fields, in the case $\Lambda = 0$. This should be contrasted with the Newtonian case, which, as we have just seen, was approximately static.

Exercise 6.5 The problem of Tides in the Newtonian limit of General Relativity: *Give a rough estimate of the height of spring and neap tides using the above-described Newtonian limit (6.33), (6.34) of Einstein's equations. Consider for simplicity an element of the Ocean water on the equator. Assume the following magnitudes: mass of Moon, $M_{\text{moon}} = 7.35 \times 10^{22}$ Kg, mass of Sun $M_{\odot} = 1.99 \times 10^{30}$ Kg, and Mass of Earth $M_{\oplus} = 5.97 \times 10^{24}$ Kg, the (mean) distance of Sun from Earth $R_{\odot} = 1.49 \times 10^{11}$ m, and of Moon from Earth $R_{\text{moon}} = 3.84 \times 10^8$ m, and the radius of Earth $r_{\oplus} = 6.37 \times 10^6$ m.*

Solution:

Spring tides occur when the Sun, Moon and Earth are *in* the same line, whilst neap tides occur when the Sun and Moon are at right angles relative to Earth (see fig. 17).

An element on the Equator at high tide is *in equilibrium*, which means that the gravitational acceleration due to Earth's gravitational potential at distance $r = r_{\oplus} + h$, $h \ll r_{\oplus}$, $g(r) = -M/r^2$ (in units $G_N = 1$), should compensate the tidal accelerations appearing in the geodesic deviation equation (5.47) due to the effects of Moon and Sun.

Notice that for the purposes of this problem, to leading order in $h \ll r_{\oplus}$, and taking into account that since $r_{\oplus} \ll R_{\odot}, R_{\text{moon}}$, we may treat the radius of the earth r_{\oplus} as the "separation" between neighboring geodesics (dashed-dotted lines in figure 17a, one geodesic passing through the center of Earth, and the others pertaining to the motion of particles in the ocean element. Then we are free to use the formula for infinitesimal geodesic deviation (5.47) to describe the pertinent 'tidal acceleration', according to our discussion in section 5.

In the arrangement of figure 17, the tidal accelerations, to leading order in h , which we restrict ourselves here, are then given by (5.47), where in the Newtonian limit the geodesic parameter is identified with the (universal) Newtonian time t :

$$\begin{aligned} \text{tidal accel.} &: \simeq r_{\oplus} R_{1010}^{\odot} + r_{\oplus} R_{1010}^{\text{moon}} , \\ \text{tidal accel. for anelement at } 90^\circ \text{ in longitude (low tide)} &: \simeq r_{\oplus} R_{1010}^{\odot} + r_{\oplus} R_{1010}^{\text{moon}} , \end{aligned} \quad (6.40)$$

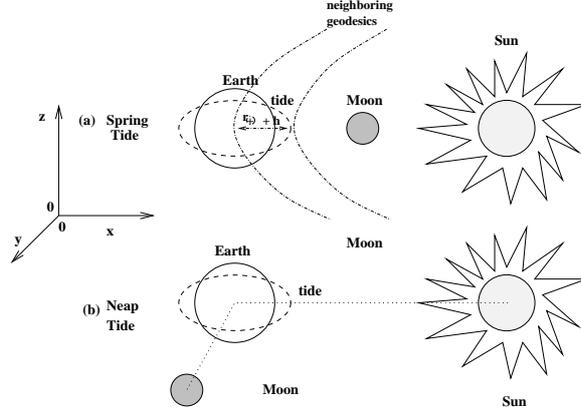


Figure 17: Arrangements of Sun and Moon relative to the Earth, for (a) spring and (b) neap tides. The radius of Earth $r_{\oplus} \ll R_{\odot}, R_{\text{moon}}$ so it may be treated as the separation of ‘neighboring geodesics’ to a good approximation.

The equilibrium conditions then read:

$$0 = r(r_{\oplus} + h) + r_{\oplus}R_{1010}^{\odot} + r_{\oplus}R_{1010}^{\text{moon}} \quad (6.41)$$

An ocean element at an angle 90° away experiences low tide, i.e. the equilibrium condition reads in this case:

$$0 = r(r_{\oplus} - h) + r_{\oplus}R_{2020}^{\odot} + r_{\oplus}R_{2020}^{\text{moon}} \quad (6.42)$$

The Riemann curvature tensors in (6.41), (6.42) are given by (6.35) in the Newtonian limit.

Subtracting (6.41) from (6.42) we obtain, to linear order in h :

$$0 = 2hg'(r_{\oplus}) + r_{\oplus} (R_{1010}^{\odot} - R_{2020}^{\odot} + R_{1010}^{\text{moon}} - R_{2020}^{\text{moon}}) \quad (6.43)$$

From (6.35) we have:

$$\begin{aligned} R_{2020}^{\odot}(y_{\odot} = z_{\odot} = 0, x_{\odot} = R_{\odot}) &= \frac{M_{\odot}}{R_{\odot}^3} , \\ R_{2020}^{\odot}(x_{\odot} = z_{\odot} = 0, y_{\odot} = R_{\odot}) &= -2\frac{M_{\odot}}{R_{\odot}^3} , \\ R_{1010}^{\odot}(y_{\odot} = z_{\odot} = 0, x_{\odot} = R_{\odot}) &= -2\frac{M_{\odot}}{R_{\odot}^3} , \\ R_{1010}^{\odot}(x_{\odot} = z_{\odot} = 0, y_{\odot} = R_{\odot}) &= \frac{M_{\odot}}{R_{\odot}^3} . \end{aligned} \quad (6.44)$$

Similar expressions occur for the Moon (i.e. replace in the above formula the quantities M_{\odot} and R_{\odot} by $M_{\text{moon}}, R_{\text{moon}}$).

Spring tides occur when the Sun and Moon are on the same line as the Earth, say the x-axis. From the equilibrium condition then (6.43), by substituting the appropriate expressions from (6.44) we obtain:

$$h_{\text{spring tides}} = \frac{3}{4} \left(\frac{M_{\odot}}{R_{\odot}^3} + \frac{M_{\text{moon}}}{R_{\text{moon}}^3} \right) \frac{r_{\oplus}^4}{M_{\oplus}} \simeq 0.39 \text{ m} .$$

Neap tides occur when the Sun is, say, on the x axis and the moon on the y-axis. Following a similar procedure as for the spring tides, one obtains in this case:

$$h_{\text{neap tides}} = \frac{3}{4} \left(\frac{M_{\text{moon}}}{R_{\text{moon}}^3} - \frac{M_{\odot}}{R_{\odot}^3} \right) \frac{r_{\oplus}^4}{M_{\oplus}} \simeq 0.15 \text{ m} .$$

In the actual situation the tidal effects are considerably larger due to hydrodynamical effects in the volume of water, which have been ignored in our simplified situation. Nevertheless this exercise shows how one can use the Newtonian limit of General Relativity to calculate (in conceptually and technically novel ways) effects that are known from the Mechanics courses.

6.3 Gravitational Waves

6.3.1 Why gravitational Waves?

Gravitational waves are one of the most important predictions of Einstein's general theory of relativity, which at present lacks experimental confirmation. The gravitational waves can be derived at present formally only in the case of weak gravitational fields, which may characterize the field far from a gravitating object. This is, of course, not an exact solution of the non-linear Einstein's equations (6.39), and this is one of the reasons why the existence of gravitational waves is still debated by some. For this purpose terrestrial and satellite experiments are currently being designed which hope to arrive at the required sensitivity to detect the weak gravitational waves believed to be produced during extreme astrophysical events (e.g. the collapse of stars). Notice that such gravitational waves are extremely weak due to the fact that their sources lie at enormous distances from the point of observation (the Earth), and this makes all the experiments for the gravitational waves extremely difficult. The recent advances in technology, however, provide some optimistic signs that the existence of gravitational waves can be confirmed (or otherwise) within the foreseeable future.

Since we are dealing with weak gravitational phenomena, as explained above, it will be sufficient to consider again the weak field postulate (6.15), or *linearized gravitation*, and its consequence (6.28), the linearized Einstein's equations. We shall examine below first properties of gravitational waves as they propagate in empty space, and subsequently an approximate mechanism by which gravitational waves are generated.

The weak field Einstein' equations in a slightly curved spacetime (6.28) read:

$$\left(-\frac{\partial^2}{\partial t^2} + \nabla^2 \right) \bar{h}_{\mu\nu} = -16\pi T_{\mu\nu} \quad (6.45)$$

where the symbol $\nabla^2 = \delta^{ij} \frac{\partial}{\partial x^i} \frac{\partial}{\partial x^j}$ denotes the usual Laplacian in Euclidean three-dimensional space. Note that this equation is simply a wave equation, with whose solutions we should already be familiar from the wave mechanics course.

6.3.2 Wave propagation in empty space

Consider first this equation in *empty space*, i.e. when $T^{\mu\nu}$ vanishes everywhere:

$$\left(-\frac{\partial^2}{\partial t^2} + \nabla^2 \right) \bar{h}_{\mu\nu} = 0. \quad (6.46)$$

From your wave mechanics courses you will recall that the *plane-wave* solution of this equation has the general form

$$\bar{h}_{\mu\nu} = A_{\mu\nu} \exp\{ik_{\kappa}x^{\kappa}\} \quad (6.47)$$

where $A_{\mu\nu}$ is some constant tensor of rank $\binom{0}{2}$ and k is the wave-four-vector $k^{\mu} = (\omega, \vec{k})$ where ω is the frequency of the wave and \vec{k} is the wave-three-vector. From the wave equation (6.46) and the Lorentz condition (6.22) we obtain

$$\eta_{\mu\nu} k^{\mu} k^{\nu} = 0, \quad k^{\mu} A_{\mu\nu} = 0. \quad (6.48)$$

Exercise 6.6 Prove the relations (6.48).

The first of relations (6.48) gives the dispersion relation of the plane wave, which in components implies that

$$\omega^2 = |\vec{k}|^2 \quad (6.49)$$

and therefore that the phase velocity as well as the group velocity is that of light, which in our units is unity.

Exercise 6.7 Prove that the phase and group velocities of the gravitational waves are both unity (in units where the velocity of light is unity).

The second of the conditions (6.48) implies that the gravitational waves are transverse (i.e. the oscillations are orthogonal to the direction of motion). We now mention for completeness that the possibility of choosing the Lorentz class of gauges (6.22),(6.23) and (6.25), allows further simplification in the form of $A_{\mu\nu}$. We shall state them without proof:

$$A_\alpha^\alpha = A_{\alpha\beta}u^\beta = 0 \quad (6.50)$$

where u^μ is an arbitrary but fixed four velocity, i.e. any constant timelike vector (recall that $u^\mu u_\mu = -1$).

The conditions (6.48),(6.50) constitute what is called *transverse traceless (TT) gauge*, the word traceless referring to the vanishing of $A_\alpha^\alpha = 0$, because, if one views A_α^β as a matrix, the quantity A_α^α is its trace, i.e. the sum of its diagonal elements (caution: this is valid *only* if you consider the mixed index object A_μ^ν as a matrix, because in that case the indices are contracted with the Kronecker δ_μ^ν to form A_μ^μ). In this gauge (symbolically denoted by *TT*) one has then

$$\bar{h}_{\mu\nu}^{TT} = h_{\mu\nu}^{TT} \quad (6.51)$$

One can also choose $u^\mu = \delta_0^\mu$ in which case

$$A_{\mu 0} = 0 \quad (6.52)$$

for all μ . For more details on the effects of the Lorentz gauge on the form of the gravitational wave, see textbooks (e.g. B. Schutz, *A first course in General relativity*, Cambridge Univ. Press 1985).

6.3.3 The effects of gravitational waves on free particles

As you may recall from your wave-mechanics course, any wave is a superposition of plane waves. For illustration purposes, as well as for simplicity, consider a wave that travels along the z direction in space. In this case there are only two independent components of $h_{\mu\nu}$, h_{xx} and $h_{xy} = h_{yx}$. Let a particle be initially *at rest*, in a chosen background Lorentz frame, when it encounters the gravitational wave. Choose the *TT* gauge referred to this frame, i.e. choose u^μ in (6.50) to be the initial four-velocity of the particle $u_0^\mu = (1, \vec{0})$. The free particle follows a geodesic equation (5.19) with respect to the time like parameter τ :

$$\frac{d}{d\tau}u^\alpha + \Gamma_{\mu\nu}^\alpha u^\mu u^\nu = 0 \quad (6.53)$$

Since the particle is initially at rest ($u_0^\mu = (1, \vec{0})$), the initial value (denoted by a subscript 0) of its (proper) acceleration $du^\mu/d\tau$ is:

$$\left(\frac{d}{d\tau}u^\alpha\right)_0 = -\Gamma_{00}^\alpha = -\frac{1}{2}\eta^{\alpha\beta}(h_{\beta,0} + h_{0,\beta} - h_{00,\beta}) \quad (6.54)$$

But since in the TT gauge $h_{\beta 0} = 0$ due to (6.51) and (6.52), one observes that initially the acceleration vanishes. This will be also true a moment later and then by similar arguments a moment later etc., i.e. the particle *remains at rest for ever in the TT gauge, under the influence of the gravitational wave*. This result simply means that by working in the TT gauge we have chosen our coordinate frame in such a way so that the particle always is at rest under the influence of the gravitational wave. The situation is analogous to going to the momentarily comoving reference frame (MCRF) of a fluid. This does not carry any observer independent information, as it is simply a coordinate choice.

To really study the effects of a gravitational wave on the free particles we should consider two particles, conveniently placing one at the origin of our (local) coordinate system, and the other at $x = \varepsilon, y = z = 0$. Both particles are initially at rest. Let us now calculate the proper distance between them under the influence of a wave propagating along the z direction. The presence of the wave amounts to disturbing the space time locally, so that the flat metric $\eta_{\mu\nu}$ is distorted to $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$, for weak waves which is the case assumed here, as explained at the beginning of this section. The proper distance is then given by:

$$\begin{aligned} \Delta\ell &\equiv \int |ds|^{1/2} = \int |g_{\mu\nu} dx^\mu dx^\nu|^{1/2} = \\ &\int_0^\varepsilon |g_{xx}|^{1/2} dx \simeq |g_{xx}(x=0)|^{1/2} \varepsilon \simeq [1 + \frac{1}{2} h_{xx}^{TT}(x=0)] \varepsilon \end{aligned} \quad (6.55)$$

This gives a non-trivial *observer independent* effect, given that $h_{xx}^{TT} \neq 0$ and is in general time dependent. Thus, *under the influence of a gravitational wave, the proper distance between two test particles changes with time*.

This can be seen in an alternative, but completely equivalent, way, by taking into account the effects of curvature of space time induced by the presence of the gravitational wave. As we have discussed previously, when we defined the Riemann curvature tensor, the latter appears in the equation for geodesic deviation (5.47) between two neighboring geodesics in a curved spacetime. Let v^α be the vector quantifying the geodesic deviation between the two particles. Then, from (5.47), we have (we consider the deviation equation with respect to a timelike affine parameter (proper time) τ):

$$\frac{D^2}{D\tau^2} v^\alpha = R_{\mu\nu\beta}^\alpha u^\mu u^\nu v^\beta \quad (6.56)$$

To lowest order in $h_{\mu\nu}$, and taking into account that the Riemann curvature tensor is already of order $\mathcal{O}(h_{\alpha\beta})$, one may set $u^\mu = (1, \vec{0})$. Then, equation (6.56), reads:

$$\frac{D^2}{D\tau^2} v^\alpha = \frac{\partial^2}{\partial t^2} v^\alpha = \varepsilon R^\alpha_{00x} = -\varepsilon R^\alpha_{0x0} \quad (6.57)$$

Exercise 6.8 *By implementing the TT gauge (6.22),(6.50), and the definition of the Riemann Curvature tensor (6.16), in the linearized (weak-field) approximation for gravity, show that in the TT gauge the Riemann curvature tensor has the following form:*

$$\begin{aligned} R^x_{0x0} &= R_{x0x0} = -\frac{1}{2} h_{xx,00}^{TT} \\ R^y_{0x0} &= R_{y0x0} = -\frac{1}{2} h_{xy,00}^{TT} \\ R^y_{0y0} &= R_{y0y0} = -\frac{1}{2} h_{yy,00}^{TT} = -R^x_{0x0} \end{aligned} \quad (6.58)$$

and all the other independent components vanish.

From (6.58),(6.57) we then observe that if the two particles have initially a separation ε in the x direction, then, after a time t under the influence of the gravitational wave, they will have a

separation vector which obeys:

$$\begin{aligned}\frac{\partial^2}{\partial t^2}v^x &= \frac{1}{2}\varepsilon\frac{\partial^2}{\partial t^2}h_{xx}^{TT}, \\ \frac{\partial^2}{\partial t^2}v^y &= \frac{1}{2}\varepsilon\frac{\partial^2}{\partial t^2}h_{xy}^{TT}\end{aligned}\tag{6.59}$$

This is consistent with the result (6.55).

In a similar manner, if the initial separation of the particles was in the y direction one has:

$$\begin{aligned}\frac{\partial^2}{\partial t^2}v^y &= \frac{1}{2}\varepsilon\frac{\partial^2}{\partial t^2}h_{yy}^{TT} = -\frac{1}{2}\varepsilon\frac{\partial^2}{\partial t^2}h_{xx}^{TT}, \\ \frac{\partial^2}{\partial t^2}v^x &= \frac{1}{2}\varepsilon\frac{\partial^2}{\partial t^2}h_{xy}^{TT}\end{aligned}\tag{6.60}$$

6.3.4 Polarization of Gravitational Waves

The equations (6.59),(6.60) are helpful in describing the polarization of the gravitational wave. To this end, consider a circle of particles in the xy plane, as in figure 18. The particles are initially at rest.

There are two cases of waves we shall consider:

- (i) $h_{xx}^{TT} \neq 0$, $h_{xy}^{TT} = 0$. In terms of their proper distance relative to the particle in the centre, the particles will be moved during their encounter with the wave, in the way shown in figure 18(b): say, first *in*, and then, since the oscillating h_{xx}^{TT} changes sign, *out*.
- (ii) $h_{xx}^{TT} = h_{yy}^{TT} = 0$, $h_{xy}^{TT} \neq 0$. In this case the circle is distorted as in figure 18(c).

Since h_{xy}^{TT} , h_{xx}^{TT} are independent, the two cases (b) and (c) in the figure 18 provide a pictorial representation of the two different *linear polarizations of the gravitational wave*. The fact that the two states are rotated (with respect to one another) by 45° is actually an important consequence of the fact that gravity is generated by a symmetric rank two *tensor*. We shall not prove this, but the reader should remember that this situation, i.e. the fact that the two polarizations differ by 45° , is *different* from the case of electromagnetic waves, whose polarizations differ by 90° . This latter property has to do with the fact that the electromagnetic waves are generated by a *vector field* (the electromagnetic potential A_μ).

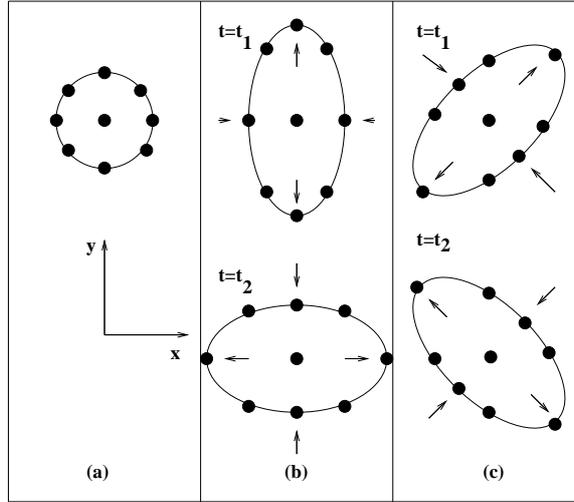
With these comments we close our brief discussion on the propagation of gravitational waves in empty space, and their effects on test particles. These will help us understand how gravitational wave detectors work, which will be the topic of a subsequent section.

6.3.5 An approximate analysis of wave generation

Our object now is to solve the gravitational wave equation (6.45) for the non-empty space case. Consider a distant gravitational source which occupies a bounded region of *space*, \mathcal{D} , far away from the point of observation. Due to the linearized approximation, the conservation law of the stress tensor (6.38) acquires the flat spacetime form as in (3.35). This is so because the observer lies very far from the gravitational source, and hence any effect of the gravitational field on the matter part of Einstein's equations is negligible. We shall assume for definiteness that $T^{\mu\nu}$ is non-vanishing only in the *interior* of the region \mathcal{D} , i.e. $T^{\mu\nu}$ vanishes outside and on the boundary (which we denote by $\partial\mathcal{D}$) of \mathcal{D} . For our purposes in this course we shall consider only the simplified but instructive case in which the (distant) source of the gravitational field has a stress-energy tensor of oscillatory form

$$T_{\mu\nu} = \mathbf{Re}\left\{S_{\mu\nu}(\vec{x})e^{-i\Omega t}\right\},\tag{6.61}$$

with $S_{\mu\nu}(\vec{x})$ being a function only of the spatial coordinates x^i and \mathbf{Re} denotes the real part. Due to the boundedness assumption above, we have the condition that $S_{\mu\nu} \neq 0$ only in a bounded region



(a) circle of particles before the passage of the wave
(b) distortion caused by + polarized wave
(c) distortion caused by x polarized wave

Figure 18: On the effects of the two polarizations of a gravitational wave on a circle of particles initially at rest. The wave propagates in the z direction (perpendicular to the page), whilst the particles lie on the xy plane (parallel to the page). The two polarizations are differing by 45° .

of space, \mathcal{D} , which is assumed spherical with a radius ϵ very small compared with the wavelength $2\pi/\Omega$ of the gravitational wave of frequency Ω .

We shall do the analysis by means of a number of exercises, which are quite straightforward but also instructive, and will help the student to assimilate the pertinent material. As we shall show below the expression for the gravitational wave in the non-empty space case considered here will be of the form

$$\bar{h}_{\mu\nu} \simeq \frac{A_{\mu\nu}}{r} \exp(i\Omega(r-t)) + \mathcal{O}(r^{-2}) \quad (6.62)$$

where $A_{\mu\nu}$ is to be determined, and r is the (large) spherical polar radial coordinate, whose origin is chosen to be the location of the source of the gravitational wave. The reader is invited to compare this expression with the corresponding one for plane waves in empty space (6.47).

Exercise 6.9 Show that a solution of (6.45) has the form $\bar{h}_{\mu\nu} = \text{Re}(B_{\mu\nu}(x^i)e^{-i\Omega t})$, where Re denotes the real part (which for convenience can be taken at the end of the computations), and $B_{\mu\nu}$ satisfies the equation:

$$(\nabla^2 + \Omega^2)B_{\mu\nu} = -16\pi S_{\mu\nu} \quad (6.63)$$

Since we want waves emitted by the source, one may assume the following outgoing-wave form of $B_{\mu\nu}$:

$$B_{\mu\nu} = \frac{A_{\mu\nu}}{r} e^{i\Omega r} \quad (6.64)$$

where $A_{\mu\nu}$ are constants.

We now integrate (6.63) over the three-space. Let us first look at the term $\int d^3x \Omega^2 B_{\mu\nu}$. Recalling our assumption that the source is non zero only over a sphere of radius $\epsilon \ll 2\pi/\Omega$, this

integral can be bounded:

$$\int d^3x \Omega^2 B_{\mu\nu} \leq \Omega^2 |B_{\mu\nu}|_{\max} \frac{4\pi\epsilon^3}{3} \quad (6.65)$$

where $|B_{\mu\nu}|_{\max}$ is the maximum value $B_{\mu\nu}$ takes inside the source. Thus, from (6.65) we observe that this term is *negligible* compared to the other terms of the integral, i.e.

$$\int_{\mathcal{D}} d^3x \nabla^2 B_{\mu\nu} = -16\pi \int_{\mathcal{D}} d^3x S_{\mu\nu} \quad (6.66)$$

Then, by virtue of Gauss's theorem, the left-hand-side of the above equation can be written as

$$\begin{aligned} \int_{\mathcal{D}} d^3x \nabla^2 B_{\mu\nu} &= \int_{\mathcal{D}} d^3x \nabla^2 \left[\frac{A_{\mu\nu}}{r} e^{i\Omega r} \right] \\ &= A_{\mu\nu} \oint dS n_i \nabla_i \left[\frac{e^{i\Omega r}}{r} \right] \end{aligned} \quad (6.67)$$

Since the region \mathcal{D} is taken to be a small sphere with radius $\epsilon \rightarrow 0$ it follows that

$$\int_{\mathcal{D}} d^3x \nabla^2 B_{\mu\nu} = 4\pi\epsilon^2 \left(\frac{d}{dr} \left[A_{\mu\nu} \frac{e^{i\Omega r}}{r} \right] \right) \Big|_{r=\epsilon} \simeq -4\pi A_{\mu\nu}. \quad (6.68)$$

Defining

$$J_{\mu\nu} = \int_{\mathcal{D}} d^3x S_{\mu\nu} \quad (6.69)$$

and taking into account the relation (6.67), and the definition of $\bar{h}_{\mu\nu}$ one obtains:

$$\begin{aligned} A_{\mu\nu} &= 4J_{\mu\nu}, \\ \bar{h}_{\mu\nu} &= 4\text{Re} \left(\frac{J_{\mu\nu}}{r} e^{i\Omega(r-t)} \right) \end{aligned}$$

This gives the expression for the gravitational wave generated by the (distant) source. We stress once again that the above analysis is approximate, since we keep only dominant terms as r^{-1} becomes small, where the weak-field approximation applies.

One can simplify the expressions for $A_{\mu\nu}$, $\bar{h}_{\mu\nu}$ considerably. From the definition (6.69) one can show that

$$-i\Omega J^{\mu 0} e^{-i\Omega t} = \int_{\mathcal{D}} d^3x T^{\mu 0}_{,0}. \quad (6.70)$$

Exercise 6.10 Prove (6.70).

Then, using the equation $T^{\mu\nu}_{, \nu} = 0$, then, and Gauss's theorem, it is straightforward to deduce:

$$i\Omega J^{\mu 0} e^{-i\Omega t} = \oint T^{\mu j} n_j dS, \quad (6.71)$$

where n_j is a vector normal to a surface bounding the volume \mathcal{D} completely containing the source of the gravitational waves. From this it follows directly that $J^{\mu 0} = 0$, given that $T_{\mu\nu}$ vanishes on the boundary $\partial\mathcal{D}$ by assumption. Hence

$$\boxed{\bar{h}_{\mu 0} = 0} \quad (6.72)$$

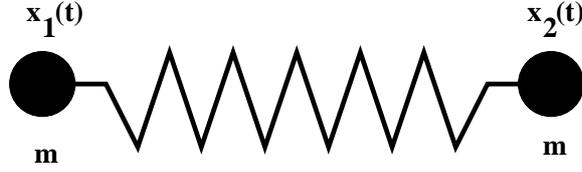


Figure 19: A prototype of a gravitational wave detector: a spring with two identical masses attached at its ends.

By making use of the tensor virial theorem (3.46), i.e. $\frac{\partial^2}{\partial t^2} \int_{\mathcal{D}} d^3x T^{00} x^i x^j = 2 \int_{\mathcal{D}} d^3x T^{ij}$, one can express J^{ij} (6.69) in terms of T^{00} (which for a source in slow motion is approximately the energy density ρ , as we have seen previously):

$$J^{ij} \exp(-i\Omega t) = \frac{\partial^2}{\partial t^2} \mathcal{I}^{ij}, \quad \mathcal{I}^{ij} \equiv \int d^3x T^{00} x^i x^j = D^{ij} \exp(-i\Omega t) \quad (6.73)$$

where the integral \mathcal{I}^{ij} is often referred to as the *quadrupole moment tensor of the mass distribution*.

From the last expression we may then write (taking the real part for completeness):

$$\boxed{\bar{h}_{ij} = -2\text{Re} \left(\Omega^2 \frac{D_{ij}}{r} \exp(i\Omega(r-t)) \right)} \quad (6.74)$$

The above formula neglects not only terms of order r^{-2} but also terms of order r^{-1} which are not dominant in the slow motion approximation. In particular, the terms \bar{h}_{ij}^j are of higher order and this guarantees that the gauge condition (6.22) is satisfied, to leading order in r^{-1} and Ω , by (6.72), (6.74). Because of (6.74) this approximation is often called the *quadrupole approximation* for gravitational radiation.

6.3.6 Detection of Gravitational Waves

Nearly all astrophysical phenomena are believed to emit gravitational waves, and the most violent ones give off radiation in large amounts. Gravitational waves are very important in principle, since they can carry information that electromagnetic waves cannot give to us. For example, gravitational waves produced in a supernova explosion come to us carrying important information about the nature of the explosion, whilst electromagnetic waves are scattered a countless number of times due to the dense material surrounding the explosion, and thus lose important information.

However, from the practical view point, gravitational waves are extremely difficult to detect experimentally, due to their weak nature, i.e. the fact that the amplitudes of the metric perturbations that can be expected from distant sources are extremely small. This is the main reason why they have not been detected so far.

Below we shall describe briefly the principle of the simplest kind of gravitational-wave detectors, and we shall illustrate, by means of an example, the weakness of the possible ‘signal’ of gravitational waves. This will help the reader understand the difficulty of the detection of gravitational waves from distant astrophysical phenomena, where the associated signals will be much more suppressed than those involved in this specific example.

An idealization of a detector of gravitational waves is depicted in figure 19. It consists of a spring with two identical masses m attached at its ends. The spring has constant k , damping constant (due to friction) ν and unstretched length ℓ_0 , in the TT gauge coordinate frame. The system lies on the x axis of the TT frame. The masses are at coordinate positions $x_1(t)$ and $x_2(t)$.

From any Mechanics course, we recall that in flat space, i.e. in the absence of the wave, the dynamical equations of the system are the following (as usual the ∂ denotes ordinary partial

differentiation with respect to time $t = x_0$ in the TT frame, with $,_{00}$ the second derivative):

$$\begin{aligned} mx_{1,00} &= -k(x_1 - x_2 + \ell_0) - \nu(x_1 - x_2)_{,0} \\ mx_{2,00} &= -k(x_2 - x_1 - \ell_0) - \nu(x_2 - x_1)_{,0} \end{aligned} \quad (6.75)$$

Defining the relative displacement $\xi = x_2 - x_1 - \ell_0$, and $\omega_0^2 = 2k/m$, $\gamma = \nu/m$, one can combine the above equations into one for ξ :

$$\xi_{,00} + 2\gamma\xi_{,0} + \omega_0^2\xi = 0 \quad (6.76)$$

which is nothing other than the equation of a damped harmonic oscillator.

Let us now find how the above equation is modified in the (slightly) curved space induced by the gravitational wave. As we have seen previously, the presence of the wave distorts the proper distance between the two masses. To find the appropriate form, we first note that in the TT coordinate frame, as we have discussed previously, a free particle tends to remain at rest. Thus, if a local frame is initially at rest at, say x_1 , then when the wave arrives it will continue to be at rest. Let its coordinates (after the wave has arrived) be x'^μ . We assume that the only motion is due to the wave, i.e. the coordinates x' differ from x only by terms of order $\mathcal{O}(h_{\alpha\beta})$. Similarly for the displacement vector $\xi = \mathcal{O}(\ell_0 h_{\alpha\beta}) \ll \ell_0$. Thus the masses' velocities are small, and the dynamics of the system is described well enough by the non-relativistic Newtonian first law connecting the acceleration with the force F^j in the TT coordinate frame:

$$mx'^j_{,00} = F'^j \quad \implies \quad mx^j_{,00} = F^j + \mathcal{O}(|h_{\mu\nu}|^2) \quad (6.77)$$

Since the only non-gravitational force acting on each mass is that due to the spring, and all the motions are slow, the spring will exert a force proportional to the *proper* extension, the latter measured using the metric. If the proper length of the spring is ℓ , and if we assume that the gravitational wave travels in the z direction, then the proper length is (cf. (6.55)):

$$\ell(t) = \int_{x_1(t)}^{x_2(t)} dx \left[1 + h_{xx}{}^{TT}(t) \right]^{1/2} \quad (6.78)$$

From the equation in the right-hand-side of the arrow in (6.77) we have:

$$\begin{aligned} mx_{1,00} &= -k(\ell_0 - \ell) - \nu(\ell_0 - \ell)_{,0}, \\ mx_{2,00} &= -k(\ell - \ell_0) - \nu(\ell - \ell_0)_{,0}. \end{aligned} \quad (6.79)$$

Defining ω_0 and γ as before, and $\xi = \ell - \ell_0 = x_2 - x_1 - \ell_0 + \frac{1}{2}h_{xx}{}^{TT}(x_2 - x_1) + \mathcal{O}(|h_{\alpha\beta}|^2)$, we can solve for $x_2 - x_1$ to leading order in $h_{\mu\nu}{}^{TT}$ to obtain:

$$x_2 - x_1 = \ell_0 + \xi - \frac{1}{2}h_{xx}{}^{TT}\ell_0 + \mathcal{O}(|h_{\alpha\beta}|^2), \quad (6.80)$$

Substituting in (6.79) we can arrive at the extension of (6.76) in the presence of the wave:

$$\xi_{,00} + 2\gamma\xi_{,0} + \omega_0^2\xi = \frac{1}{2}\ell_0 h_{xx,00}{}^{TT} \quad (6.81)$$

to first order in $h_{xx}{}^{TT}$. Comparing with (6.76) we observe that this equation has the form of a *forced* damped harmonic oscillator, the 'force' being provided by the distortion of the space time, manifested through the non-zero $h_{xx}{}^{TT}$, under the action of the gravitational wave. Equation (6.81) is the fundamental equation that governs the response of the detector to the gravitational wave.

Let us provide now some simple estimates on the amplitude of the oscillations encountered in a gravitational wave. Consider an oscillatory wave:

$$h_{xx}{}^{TT} = A \cos \Omega t \quad (6.82)$$

The steady solution for ξ in this case is:

$$\xi = R \cos(\Omega t + \phi), \quad (6.83)$$

with

$$R = \frac{1}{2} \frac{\ell_0 \Omega^2 A}{\left((\omega_0 - \Omega)^2 + 4\Omega^2 \gamma^2 \right)^{1/2}}, \quad \tan \phi = \frac{2\gamma \Omega}{\omega_0^2 - \Omega^2} \quad (6.84)$$

The energy E of oscillation of the detector, to lowest order in h_{xx}^{TT} is:

$$E = \frac{1}{2} m (x_{1,0})^2 + \frac{1}{2} m (x_{2,0})^2 + \frac{1}{2} k \xi^2 \quad (6.85)$$

For a detector who is initially at rest (before the wave has arrived), we have $x_{1,0} = -x_{2,0} = -\frac{1}{2} \xi_{,0}$, so that

$$E = \frac{1}{4} m \left[(\xi_{,0})^2 + \omega_0^2 \xi^2 \right] = \frac{1}{4} m R^2 \left(\Omega^2 \sin^2(\Omega t + \phi) + \omega_0^2 \cos^2(\Omega t + \phi) \right) \quad (6.86)$$

One is interested in taking the mean value $\langle E \rangle$ over a period Ω . From the above equation this yields:

$$\langle E \rangle = \frac{1}{8} m R^2 (\omega_0^2 + \Omega^2) \quad (6.87)$$

In most experiments one is interested in detecting a known (astrophysical) source, whose frequency is known. In this case one adjusts the detector in such a way that $\omega_0 = \Omega$ (resonant detectors). In this case the amplitude of the response will be

$$R_{\text{resonant}} = \frac{1}{4} \ell_0 A \frac{\Omega}{\gamma} \quad (6.88)$$

and the energy of the vibration

$$\langle E \rangle_{\text{resonant}} = \frac{1}{64} m \ell_0^2 \Omega^2 A^2 \left(\frac{\Omega}{\gamma} \right)^2 = \frac{1}{16} m \ell_0^2 \Omega^2 A^2 Q^2 \quad (6.89)$$

where the last equality on the right-hand-side stems from the definition of the *quality factor* Q of an oscillator $Q \equiv \frac{\omega_0}{2\gamma}$.

In practice, the oldest type of gravitational wave detector is the *resonant oscillator* pioneered by J. Weber (U. of Maryland, 1961). The detector consisted of a massive long aluminum cylindrical bar. In such a device the rôle of the spring is played by the elasticity of the bar when it is stretched along its axis. When the waves hit the bar broadside they excite longitudinal modes of vibration. The aluminum bars of the Weber detector had a mass 1.4×10^3 Kg, length $\ell_0 = 1.5$ m, resonant frequency $\omega_0 = 10^4$ s⁻¹, and $Q \sim 10^5$. According to the analysis above, (6.88),(6.89), then, this means that for such detectors, a strong resonant gravitational wave of amplitude $A = 10^{-20}$ will excite the bar to an energy of order of 10^{-20} J. The resonant amplitude will be then $R_{\text{resonant}} = 10^{-15}$ m, i.e. roughly the diameter of an atomic nucleus ! For realistic situations, gravitational waves generated by extreme astrophysical phenomena at distant sources, will be much more suppressed than this, and will last for too short a time to bring the bar to its full resonant amplitude.

This simple example demonstrated the difficulty in the detection of the gravitational wave, and partially explains probably why they have not been detected so far. Even the slightest noise, such as thermal (finite temperature) effects, could jeopardize the detection of the gravitational wave. The modern advances in *cryogenics* will certainly prove helpful in providing the new generation of gravitational wave detectors with the necessary thermal isolation.

Because the above analysis is only based on a linearized theory of gravity, this makes people wonder whether gravitational waves do really exist. At present, an exact solution of Einstein's equations at the source (e.g. knowledge of the stress tensor *etc*) is still lacking. This makes the detection of the wave very important, and for this purpose new, more accurate experiments, are currently being designed, which take advantage of the modern technological advances in order to increase the experimental sensitivity. The most interesting and promising experiments are those of (laser) interferometric type, in which the presence of the wave will be detected through interference patterns, a technique that leads to sensitivities down to 10^{-18} m or smaller, for the next generation of such interferometers, currently under construction. More details can be found in the web (e.g. <http://www.geo600.uni-hannover.de/>, or <http://www.ncsa.uiuc.edu/Cyberia/NumRel/LIGO.html>, <http://www.ncsa.uiuc.edu/Cyberia/NumRel/GravWaves.html> *etc.*).

6.3.7 Discussion

As mentioned above, the fact that gravitational waves have been studied only as approximate solutions to Einstein's field equations, might make the reader have some doubts about their real existence. However, if such a phenomenon turns out to be verified in nature it will be extremely important in yielding invaluable information on distant (and extreme) astrophysical phenomena, such as collapse of stars, supernova explosions *etc.*, which could not be retrieved otherwise. The reason is simple. The gravitational wave carry energy (and thus information) away from the source, part of which then is transmitted onto the detector, as we have just seen. In some cases the gravitational wave is the only way (if observed) of transmitting accurate information about the distant source that produced the wave, given that electromagnetic waves are usually scattered a very large number of times in the neighborhood of the source, especially if the latter is an exploding supernova or other celestial object. In this way the electromagnetic wave lose an important part of the information. The interested reader may find more extensive analysis on the energy carried away by gravitational waves in relevant textbooks (e.g. B. Schutz, *A first course in General relativity*, Cambridge Univ. Press 1985).

7 An exact Solution to Einstein's Equations: The Schwarzschild Spacetime

There are not many exact solutions to Einstein's equations (6.39) known to date. However, the few that are known are very important in providing us sufficient quantitative support towards our quest for understanding the Universe as a whole. In the rest of the course we shall examine two such solutions. In this section we shall discuss an exact solution to Einstein's equations in the exterior of a *massive* body with *spherical symmetry*, the so-called *Schwarzschild solution*, whilst in the next section we shall examine a *Cosmological Solution*, the so-called Friedmann-Robertson-Walker Spacetime, describing our (observable) Universe as a whole. The material and terminology used in this section follows closely the presentation in the book of **E. Taylor** and **J.A. Wheeler**, *Exploring Black Holes* (Addison Wesley Longham 2000), which the student is strongly advised to read. Many more physically interesting topics and examples are discussed there.

7.1 The Schwarzschild Metric and Birkhoff's Theorem

Consider a Massive body of mass M , with spherical symmetry, embedded in a three dimensional space. Take the origin of a coordinate frame to be located at the centre of the spherical body. This spherically symmetric configuration is actually a property of Matter under the influence of gravitation. If one ignores rotation of the spherical body (which we shall not take into account for the purposes of this course), then matter has the tendency, by virtue of gravitation, to agglomerate into *spherical centres of attraction*. Nicolaus Copernicus was the first to suggest that Earth was not the only such centre of gravitational attraction, but there were multiple centres of gravity, provided by the Sun and the other planets as well.

Karl Schwarzschild was the first person to propose (in 1915, soon after Einstein wrote down his famous equations) a *spherically symmetric exact* solution to the equations, with *zero cosmological constant* $\Lambda = 0$. This solution describes pretty well the form of the spacetime geometry in the *exterior* of a spherical body. The closer the distribution of matter to exact spherical symmetry, the better the spacetime geometry around the body resembles that of Schwarzschild.

There is a very important fact the reader should remember about this solution: Schwarzschild's solution describes the space time **external** to **any** isolated spherically symmetric body of the Universe.

Below we shall only give the form of the metric in the Schwarzschild geometry. We shall not prove that this is a solution to Einstein's equations (6.39), with $\Lambda = 0$, for the purposes of this course. The Schwarzschild space time reads:

$$ds^2 = - \left(1 - \frac{2M}{r}\right) dt^2 + \frac{dr^2}{\left(1 - \frac{2M}{r}\right)} + r^2 (d\theta^2 + \sin^2\theta d\phi^2) \quad (7.1)$$

in spherical polar coordinates (r, θ, ϕ) , for the space-like infinitesimal proper distance. The corresponding formula for the time-like (proper time τ) interval is:

$$d\tau^2 = \left(1 - \frac{2M}{r}\right) dt^2 - \frac{dr^2}{\left(1 - \frac{2M}{r}\right)} - r^2 (d\theta^2 + \sin^2\theta d\phi^2) \quad (7.2)$$

In the above formulæ the coordinate time t is the *far away time* measured on clocks *far away* from the centre of attraction, where the influence of gravity can be neglected. We shall come back to this issue later on.

At the moment we note that, as we have repeatedly said in this course, we are working in geometrized units where $G_N = c = 1$, which implies that masses are measured in units of length. In the above formulæ (7.1),(7.2), M is the mass of the spherical body. If one wishes to reinstate the units of G_N , and c , then this is done by simply replacing t by $x^0 = ct$ and M by $G_N M/c^2$, where now t is measured in seconds and M in normal units of mass, e.g. Kg.

It is important to remember that the Mass M is a constant of integration in Einstein's equations. Thus, the Schwarzschild metric characterizes regions of space where there may be simply spherical symmetry, without the presence of a massive body at the origin of the spherical shells.

Indeed, the generality of the Schwarzschild metric is encoded in the following theorem, due to Birkhoff (1928), which we shall state and use below without proof. This is a very important theorem that the reader should remember, because it leads, as we shall see, to important results concerning physical applications of the Schwarzschild metric in many situations.

Birkhoff's theorem (1928): The Schwarzschild solution (7.1) is the **ONLY** spherically symmetric, asymptotically flat (i.e. tends to flat Minkowski space time in the limit $M/r \ll 1$) solution to Einstein's VACUUM (empty space $T_{\mu\nu} = 0$) field equations (6.39), even if one drops the initial assumption that the metric is static, i.e. if one starts from the general case where the components of the metric tensor $g_{\mu\nu}$ are assumed functions of both r, t .

The above theorem means that even a *radially* pulsating or collapsing star will have a *static exterior* of constant mass. One conclusion one can draw from this is the following:

There are no gravitational waves from pulsating spherical systems. This is because the latter are time dependent solutions, as we have seen, whilst, by virtue of Birkhoff's theorem, the exterior geometry of spherical objects is described by the static Schwarzschild solution (7.1), even if the object itself is non static.

This last statement was obtained without any calculations, by simply recalling the basic theorems of Schwarzschild and Birkhoff on the structure of the exterior geometry of spherical bodies. One can indeed confirm this result on the absence of gravitational waves by performing computations in the linearized theory. The absence of gravitational waves from a pulsating celestial object finds a nice analogy in electromagnetism, where there is no ‘monopole’ electromagnetic radiation either.

Using Birkhoff’s theorem, as well as the fact that the mass is an integration constant in Einstein’s equations, one can arrive at another important conclusion:

There are no gravitational forces acting on test particles in the interior of a hollow self-gravitating sphere.

Indeed, Birkhoff’s theorem states that a spherically symmetric *vacuum* (i.e. empty space) gravitational field is always static, and is always in the Schwarzschild solution (7.1), where it should be remembered that the mass M is an integration constant in the solution of Einstein’s equations (6.39).

The space time defined by the interior of a self-gravitating hollow sphere is a spherically-symmetric vacuum configuration, and hence we must have the Schwarzschild solution (7.1), according to Birkhoff’s theorem. But here the point $r = 0$ is *regular*, not a singularity.

From (7.1), and taking into account that the mass M is an integration constant, as mentioned above, the condition of regularity at $r = 0$ is achieved by choosing $M = 0$ in the interior of the hollow sphere. Thus, the space time inside the hollow sphere becomes the ordinary flat Minkowski space time, in which there are no gravitational forces, and hence a test particle experiences *no gravitational forces* inside a self-gravitating hollow sphere.

7.2 The Schwarzschild Metric and Black Holes: Horizons

The Schwarzschild space time applies to the description of the exterior geometry of ordinary celestial objects, such as Planets, stars *etc*, but also to *Black Holes*, which are objects resulting for instance from the collapse of ordinary matter, e.g. stars. Such spherically symmetric objects have a very strong gravitational attraction, so strong that nothing, not even light can escape from them. In our discussion below we shall not often distinguish these cases, except when is necessary. Regarding black holes we should mention that the point $r = 2M$, at which the components of the metric (7.1) become *degenerate or singular* is called the *event horizon* (or simply horizon) of the Black Hole.

As we shall see in a subsequent exercise, in the next subsection, the spacetime curvature, which provides an invariant characterization of the geometry, is *perfectly regular* at that point, the true singularity in the case of a black hole occurring at $r = 0$, i.e. in the interior for which the Schwarzschild solution (7.1) is not strictly valid. This reflects the fact that the Schwarzschild coordinates are not *adequate* to describe the entire spacetime in both the exterior and the interior of a Black Hole. In fact, if one assumes the form of the Schwarzschild metric (7.1) as valid intact *inside* the horizon $r < 2M$, then one sees that the r coordinate becomes now *timelike*, whilst the t coordinate becomes *spacelike*. This is an important feature of the horizon interior which is captured correctly by the Schwarzschild solution (7.1). The basic issue with the Schwarzschild coordinates is to find appropriate transformations that yield a metric which has *smooth behaviour* at $r = 2M$ and can extend the exterior solution (7.1) to the interior of the horizon; such extensions do not necessarily have the form (7.1). In this course we shall not be dealing with such Black Hole interior, but we mention that one can indeed find an appropriate set of coordinates which accomplish this task.

Before closing this subsection, it is worth noting that *every* spherically symmetric body (e.g. stars, planets, billiard balls, etc.) has a Schwarzschild horizon at (in natural units) $r = 2M$. A black hole is simply an object so dense that all its constituent matter is contained *within* this horizon. On the other hand stars have horizons which are within the stellar envelope, i.e. inside the star. For instance, the sun has a mass (in geometrized units) of $M_{\odot} = 1.477$ km and

hence Schwarzschild radius 2.954 km, which is obviously inside the sun, which has mean radius 6.960×10^8 m. The Earth has a mass of $M_{\oplus} = .444$ cm, and therefore a Schwarzschild radius of .888 cm and an actual mean radius of 6.371×10^3 km. Why, then, does neither the sun nor the Earth collapse to a black hole inside their Schwarzschild radii? The answer is simply that to create an object as dense as a black hole requires the total mass present to exceed a critical value, which is roughly six solar masses. We shall not cover the process of gravitational collapse in a star in this course, but it useful to know the existence and approximate value of the critical mass.

7.3 Coordinate Systems in Schwarzschild Geometry

7.3.1 Concepts and basic notions in the Schwarzschild Geometry

In what follows we shall consider the simplified situation in which the ϕ angular parameter of the Schwarzschild space time (7.1) is fixed, i.e. $d\phi = 0$, but the angle θ is now varied in the range $[0, 2\pi]$ (instead of $[0, \pi]$ of the spherically symmetric case (7.1)). This is a special case, but sufficient for our purposes in this course, as it captures all the non-trivial, and physically important, features of the Schwarzschild geometry.

In this case the space-like Schwarzschild metric (7.1) becomes:

$$ds^2 = - \left(1 - \frac{2M}{r}\right) dt^2 + \frac{dr^2}{\left(1 - \frac{2M}{r}\right)} + r^2 d\theta^2 \quad (7.3)$$

The space part is described in terms of polar coordinates in this special case.

On the other hand the time-like (proper time) separation is:

$$d\tau^2 = \left(1 - \frac{2M}{r}\right) dt^2 - \frac{dr^2}{\left(1 - \frac{2M}{r}\right)} - r^2 d\theta^2 \quad (7.4)$$

Below we shall use (7.3),(7.4) as our guides in the discussion of the properties of the Schwarzschild spacetime outside massive spherical bodies.

The first important notion of the geometry, already mentioned, is that it yields the flat Minkowski spacetime *far away* (i.e. for $M/r \ll 1$) from the centre of gravitational attraction, as follows directly from (7.3).

Exercise 7.1 Show that, in the limit $M/r \ll 1$, the geometry (7.3) reduces to that of flat Minkowski spacetime $ds^2 = -dt^2 + dx^2 + dy^2$, where x, y , are appropriate Cartesian coordinates.

Second notion is that of neighboring, concentric spherical shells, surrounding the massive body. Consider an observer living on a spherical shell at r surrounding the body, and being concentric with it. Suppose that the shell observer measures radial distances by means of a rod spanning the radial separation dr between his shell and a neighboring one. Suppose also that the observer sets off two firecrackers, one at each end of the rod, at the same (far away) time $dt = 0$. Such explosions constitute the two events whose separation is described by the metric (7.3). The two events have zero azimuthal separation $d\theta = 0$, and, hence, their *proper distance* is obtained from (7.3) as:

$$ds = dr_{\text{shell}} = \frac{dr}{\left(1 - \frac{2M}{r}\right)^{1/2}} \quad (7.5)$$

This proper distance is the distance dr_{shell} the shell observer measures directly. As a result of the factor $\left(1 - \frac{2M}{r}\right)^{1/2}$ in the denominator of (7.5), which is less than unity, one observes that dr_{shell} is **greater** than dr . The change of this factor from place to place implies non-trivial spacetime curvature. This can be confirmed by direct computation of the curvature scalar in the two-dimensional Schwarzschild geometry, which we leave as an exercise (exercise 7.2).

On the other hand, consider now a situation in which the two events are the sequential ticks of a clock bolted to the shell. In this case the radial and azimuthal separations between the events are zero ($dr = 0$, $d\theta = 0$). Hence from (7.4) one has in this case that the *proper time* separation between the events, which is the time dt_{shell} measured by the clock bolted to the shell, is given by:

$$d\tau = dt_{\text{shell}} = \left(1 - \frac{2M}{r}\right)^{1/2} dt \quad (7.6)$$

Here dt corresponds to the lapse of far away time, measured by an observer (clock) far way from the centre of the gravitational attraction. We observe that $dt_{\text{shell}} < dt$.

This phenomenon is related to the *Gravitational Redshift*. This is because the period of light in the Schwarzschild geometry (caused by the presence of a Massive body, and hence non-trivial gravitational field) increases as the light climbs away from the centre of the attraction, i.e. a far away observer sees the period of light being longer (dt) than it was at the point of emission dt_{shell} (the latter assumed closer to the gravitational centre of attraction). Notice that the period of light corresponds to two events, like in the example above where the events were identified with the two ticks of the clock.

Exercise 7.2 Using the definition of the Riemann Curvature tensor, compute the curvature scalar associated with the two-dimensional version (7.3) of the Schwarzschild spacetime (i.e. $d\theta = 0$). What do you observe for the value of the curvature scalar at $r = 2M$ where the metric seems not to be well defined? Are there any points in spacetime at which the curvature scalar diverges?

Note on Vacuum Solutions to Einstein's Equations

In empty space Einstein's equations in d spacetime dimensions read

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 0. \quad (7.7)$$

By “tracing” these equations, i.e. by contracting with $g^{\mu\nu}$ we obtain

$$\frac{(d-2)}{2}R = 0. \quad (7.8)$$

Therefore for any dimension $d > 2$ this implies that $R = 0$ and hence from (7.7) the metric is so-called *Ricci flat*, i.e. $R_{\mu\nu} = 0$.

Notice that this does *not* necessarily imply that the spacetime is flat, since there might still be non-vanishing components of the Riemann curvature tensor; in these circumstances, a non-zero and invariant characterization of the geometry is given by not by the single scalar curvature, but by invariants constructed from higher powers of the Riemann curvature tensor.

In two-dimensional vacuum spacetimes, equations (7.7) are in fact valid *identically*, and hence as we see from equation (7.8) (by setting $d = 2$) R may be non-trivial in the vacuum. In fact this is exactly what happens in the two-dimensional Schwarzschild metric in exercise 7.2.

Note on the Schwarzschild Solution in Four Dimensions

In view of the above discussion, in four spacetime dimensions, the Schwarzschild metric, which is an *exact* solution to the four-dimensional vacuum field equations, is *Ricci flat*. However, this does not imply that the metric is not curved, for there are non-zero components of the Riemann curvature tensor. A non-zero and invariant measure of the curvature is in this case given by the “square” of the Riemann tensor, $R^{\alpha\beta\mu\nu}R_{\alpha\beta\mu\nu} = 48M^2/r^6$. We observe from this that the point $r = 0$ is a true singularity of the geometry. As we have mentioned previously, the Schwarzschild solution is strictly not valid in the form (7.1) in the interior of the horizon, $r < 2M$. For our purposes below, especially when we discuss plunging toward a black hole, we shall use the Schwarzschild metric even in the interior regions to compute the observer independent proper time taken for

a plunging observer to reach the singularity starting from the horizon. In fact as mentioned previously one can show that there are extensions of the Schwarzschild solution, i.e. appropriate coordinate transformations, that provide us with a correct choice of coordinates to discuss the physics at and inside the horizon. However for our purposes here we shall continue using (rather formally) the metric (7.1) to get qualitative information even inside the horizon.

7.3.2 Three Coordinate systems in the Schwarzschild Space Time

In the Schwarzschild geometry (7.1), which is valid *only* in the *exterior* of a spherical massive body, or in general in the *vacuum* (empty space), e.g. the interior of a hollow sphere, there are **three** coordinate systems involved.

1. *Free Float (inertial) frame, valid Locally.*

As we have already mentioned, in General Relativity one can choose, within a sufficiently small coordinate patch of spacetime (*locally*), a free-float (inertial) coordinate system, in which Special Relativity is valid. This carries over, of course, to the specific case of Schwarzschild spacetime. The terminology “sufficiently small” used here means simply that all the effects of tidal accelerations (due to the non-uniformity of the gravitational field) are negligible, within the accuracy of the measurements performed by the local observer.

2. *Spherical Shell Observer*

This coordinate system is valid locally on a *spherical shell* surrounding the massive body, and being *concentric* with it. An example of a spherical shell is that of the surface of Earth, the latter viewed as a massive body. If one ignores, to a first approximation the rotation of Earth, the metric outside Earth is of Schwarzschild form, given that the spacetime is empty outside Earth (of course this statement is valid sufficiently near Earth, so that one does not encounter another celestial object, e.g. Moon, Sun, other planets *etc.*).

Special Relativity is sufficient for the shell observer, at least locally in space and time. Let us verify this statement by looking at the Schwarzschild metric (7.3), or (7.4). Using (7.5) and (7.6) it is straightforward to rewrite the Schwarzschild metric in terms of the shell observer quantities dt_{shell} , dr_{shell} . The result for the time-like metric (7.4) is:

$$d\tau^2 = dt_{\text{shell}}^2 - dr_{\text{shell}}^2 - r^2 d\theta^2 \quad (7.9)$$

Take into account now that $r d\theta$ is *defined* so that this quantity is the directly measured distance along the surface of the shell (see figure 20). Hence the right-hand-side of the last equation contains only coordinate increments measured directly by the shell observer. The form of (7.9) is formally similar to that of flat space time. A shell observer, for instance, measures the speed of light to be unity (cf. (7.9)). But the shell observer is not a free-float system. Every experiment taking place on the shell is influenced by the “gravitational force” in the sense that the shell observer is not an inertial observer.

3. Bookkeeper coordinates r, θ, t : The Schwarzschild Bookkeeper.

A free float observer makes observations that span only a little patch of spacetime (local observations). In contrast, the coordinates r, θ, t called Schwarzschild coordinates, provide a *global description* of space time, since, for instance, they describe two events that can be separated at large distances in spacetime, e.g. even lying on opposite ends of a black hole. The Schwarzschild observer, however, is a *bookkeeper*, a “top-level accountant” who does not make measurements for him/herself, but simply compiles the results of measurements done by local free-float and shell observers and provided to him/her.

Let us give a precise definition of this observer: One can construct an imaginary lattice of spherical shells each characterized by r , the angle θ and covered with clocks that measure the far away time t . This Schwarzschild lattice can in principle start near the horizon ($r = 2M$) and extend outwards indefinitely from an isolated massive body. The Schwarzschild

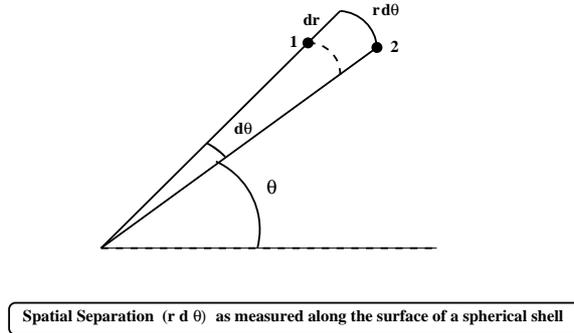


Figure 20: Spatial separation $r d\theta$ as measured directly on a surface of a spherical shell.

coordinates of any event outside the horizon can then be read directly using this lattice. This collection of shells and clocks can be collectively called a *Schwarzschild Observer*.

A ‘report’ by this observer on the results of the measurements performed by the shell and free float observers is done as follows: imagine that an orbiting satellite emits two sequential flashes as it flies past two shells concentric, say, to a black hole (or to any other spherical body in that matter). The local shell observers measure directly dt_{shell} and dr_{shell} using clocks bolted to their shell and rods respectively, and then they convert these measurements into results for dt and dr using (7.6) and (7.5) respectively. The local observers also can measure the change in azimuthal angle $d\theta$ in the plane of orbit (cf. figure 20).

Then, the bookkeeper Schwarzschild observer prepares a *table* as follows: the bookkeeper knows r, θ, t at the beginning of this increment of time. To these he/she adds increments dr and $d\theta$ for each lapse of far-away time dt reported by the local shell observer. The result of this act is a *table*, a diagram which is called a **Schwarzschild Map**, which traces the satellite through spacetime as expressed in coordinates r, θ, t .

7.3.3 Some Physical Applications of Schwarzschild Spacetime

Below we shall illustrate how one can use the above results in order to make important predictions on the effects of General Relativity by means of instructive exercises with their model solutions. In particular, we shall study the issue of gravitational redshift, and the behaviour of clocks in gravitational fields, treating Earth as a spherical non-rotating body of Mass M .

Exercise 7.3 Gravitational Redshift: *Consider a spherically-symmetric non-rotating body of mass M , and let two concentric shells surrounding the body be located at $r_1 = 4M$ and $r_2 = 8M$, where r denotes the radial spherical polar coordinate, and we work in a system of units in which the Mass is measured in units of length. Let light be emitted from the shell r_1 and absorbed at the shell r_2 . Show that the period of this light is increased by a factor 1.22 as a consequence of the gravitational red-shift.*

Solution: According to Birkhoff’s theorem, in this case, the space-time outside the body (vacuum solution) is of Schwarzschild form. It will be sufficient for our purposes to consider (7.3) or (7.4).

For a shell observer, living at a shell of radius r , the period of light may be thought of as corresponding to the temporal separation between two events, dt_{shell} as measured by clocks bolted to the shell where the observer lives. This time may be related to the period of light dt measured by a remote observer as follows: First set $dr = d\theta = d\phi = 0$ in the expression for the Schwarzschild metric, since the events occur at the same place. This leaves dt_{shell} as the proper

time $d\tau$ (recall that the proper time is defined as $d\tau^2 = -ds^2$, i.e. from the time-like form (7.4) of the Schwarzschild metric). From (7.6) we have:

$$dt_{shell} = \left(1 - \frac{2M}{r}\right)^{1/2} dt \quad (7.10)$$

To find the period of light measured by a second shell observer at r_2 , along the radial coordinate of the Schwarzschild space time, one should use equation (7.10) *twice*, once for each shell, and make the remote lapse time dt equal in both cases.

$$\frac{dt_{shell_1}}{\left(1 - \frac{2M}{r_1}\right)^{1/2}} = dt = \frac{dt_{shell_2}}{\left(1 - \frac{2M}{r_2}\right)^{1/2}},$$

from which for $r_1 = r_2/2 = 4M$ one obtains:

$$\frac{dt_{shell_2}}{dt_{shell_1}} = \frac{\left(1 - \frac{1}{4}\right)^{1/2}}{\left(1 - \frac{1}{2}\right)^{1/2}} = \frac{0.866}{0.707} = 1.22$$

Thus the period of light is *increased* (redshifted) by the factor 1.22 as it climbs from $r_1 = 4M$ to $r_2 = 8M$. This is sufficient to shift yellow light to deep red.

Exercise 7.4 Clocks in Gravitational Fields: *An aircraft is flying back and forth for 15 hours at an altitude $h = 9000$ m. The plane carries atomic clocks that are compared by laser pulse with identical clocks on the ground. Assume that the plane flew very slowly so that it can be considered ‘almost on station’ at the altitude h above the Earth’s surface. Moreover, consider h very small compared to the radius of earth $r = 6.4 \times 10^6$ m, so that $h+r \simeq r$ to a good approximation. Treating the Earth as a spherical non-rotating body of Mass (in units of length) $M = 4.4 \times 10^{-3}$ m, show that, as a consequence of General-Relativistic effects, during the $t_{shell}=15$ hour flight, the plane’s clock gained approximately*

$$dt_{shell} \simeq \left(\frac{Mh}{r^2}\right) t_{shell} \simeq 52.2 \times 10^{-9} \text{ s,}$$

as compared with the ground clocks.

Solution: Treating Earth as a non-rotating spherical body of Mass M and radius r , we have that the geometry in the exterior is well-described by the Schwarzschild vacuum solution to Einstein’s equations. Assuming that the plane flies very slowly back and forth, one may practically assume that the airplane is “almost at station” (i.e. does not move with respect to an observer on the ground) at the altitude h . Hence one may ignore any special-relativistic effects, associated with the Lorentz time-stretching γ factor, due to plane’s velocity, and concentrate exclusively on the general-relativistic effects of the Schwarzschild geometry.

Call the clock at the surface of the Earth the shell clock. Let t_{shell} be the time the airplane has been ‘on station’ at the altitude h , i.e. $t_{shell} = 15 \text{ hours} = 15 \times 3600 \text{ sec}$, and t the corresponding far away time, measured by an observer at infinity from the Earth’s centre of attraction.

From the formulæ of the Schwarzschild geometry, these two times are related as follows: first we recall the formula for the corresponding differentials (7.6):

$$dt_{shell} = \left(1 - \frac{2M}{r}\right)^{1/2} dt \quad (7.11)$$

and then one may integrate over time, both sides of this equation, assuming M and r independent of t , in which case one obtains for t_{shell} in terms of t :

$$t_{shell} = \left(1 - \frac{2M}{r}\right)^{1/2} t \quad (7.12)$$

The plane's altitude h is identified with the radial distance dr_{shell} between spherical shells surrounding earth, as measured by a shell observer (recall that $h \ll r$, which allows us to identify h with a differential). This is nothing other than the proper distance dr_{shell} in the Schwarzschild geometry for simultaneous events for a far-away observer $dt = 0$, which also have zero angular separation. Hence from (7.5) one has:

$$dr_{shell} = \frac{dr}{\left(1 - \frac{2M}{r}\right)^{1/2}} \quad (7.13)$$

Take the derivative of the expression (7.12) with respect to r , and approximate $r + h \simeq r$, and use (7.11),(7.13) to convert the resulting dr to dr_{shell} . We then have:

$$dt_{shell} = \frac{M}{r^2} \frac{dr_{shell} t_{shell}}{r} \left(1 - \frac{2M}{r}\right)^{-1/2}$$

Numerically, for the values of M and r given in the problem, the term $2M/r \simeq 10^{-9} \ll 1$, so one may neglect it in front of 1 in the last factor of the right-hand side of the equation above, and get:

$$dt_{shell} \simeq \left(\frac{Mh}{r^2}\right) t_{shell}$$

The term dt_{shell} gives the difference in readings between the airborne clocks and the earthbound clocks. The fact that the altitude h is small compared to the Earth's radius allows the identification of the required difference in readings with the differential dt_{shell} . Substituting the values given in the problem, one then finds that the plane's clock *gained*, during the 15 hour flight, approximately $dt_{shell} \simeq 52.2 \times 10^{-9}$ sec, compared with the ground clock.

The above exercise shows therefore that clocks tick differently, depending on the *altitude*. This is an *important effect of General Relativity: clocks run slow in the presence of gravitation, as compared with those far away from the centre of gravitational attraction*. In a similar spirit to the gravitational redshift, this is indeed a consequence of the less-than-unity factor $\left(1 - \frac{2M}{r}\right)^{1/2}$ in front of dt in the Schwarzschild metric (7.1).

Historical Note:

The above exercise is actually a true experiment taken place on November 22, 1975 over Chesapeake Bay in USA. This was one of the most accurate experiments verifying the predictions of general relativity as regards its effects on clocks in gravitational fields (in the case of the experiment the gravitational field of Earth).

7.4 Plunging towards a centre of Gravitational Attraction, e.g. a Black Hole

7.4.1 The principal of Extremal Aging and Conserved Energy in the curved Schwarzschild Geometry

Consider the case of a stone plunging radially (i.e. $d\theta = 0$) towards the centre of gravitational attraction, as shown in figure 21. The stone emits three flashes. All flashes are fixed in position and the first and last are also fixed in time. The time t at which the stone passes through the intermediate dot in the figure is then allowed to vary until the wristwatch (proper) time is *maximized*, as in the corresponding case of flat space time examined previously in the course. This is dictated by the *Principle of Extremal Aging*, which is valid intact in curved spacetimes as well.

For simplicity we replace the differentials $d\tau$ and dt in the expression for the time-like form of the Schwarzschild metric (7.4) by τ and t respectively, with the understanding that these separations are small. In this case we have:

$$\tau^2 = \left(1 - \frac{2M}{r}\right) t^2 + \text{terms without } t \quad (7.14)$$

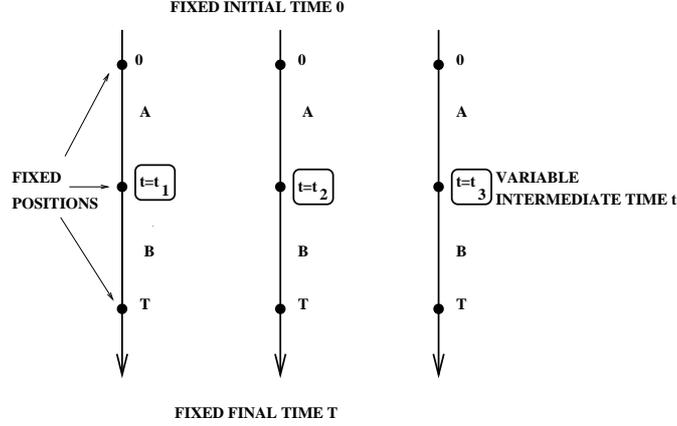


Figure 21: Deriving the expression for the energy in the Schwarzschild geometry from the principal of extremal aging. A stone is plunging radially towards the centre of gravitational attraction, as it emits three flashes. All flashes are fixed in position and the first and last are also fixed in time. The time the stone passes through the intermediate dot is then allowed to vary until the wristwatch (proper) time is maximized, as in the corresponding case of flat space time examined previously in the course.

We apply successively the above formula for the two segments A and B depicted in figure 21.

For the first segment we have (by analogy with the flat spacetime case):

$$\tau_A = \left[\left(1 - \frac{2M}{r_A}\right)t^2 + \text{terms without } t \right]^{1/2} \quad (7.15)$$

Take the derivative with respect to t :

$$\frac{d\tau_A}{dt} = \frac{\left(1 - \frac{2M}{r_A}\right)t}{\left[\left(1 - \frac{2M}{r_A}\right)t^2 + \text{terms without } t\right]^{1/2}} = \left(1 - \frac{2M}{r_A}\right) \frac{t}{\tau_A} \quad (7.16)$$

In a similar manner for segment B one finds:

$$\tau_B = \left[\left(1 - \frac{2M}{r_B}\right)(T - t)^2 + \text{terms without } t \right]^{1/2} \quad (7.17)$$

Take the derivative with respect to t :

$$\frac{d\tau_B}{dt} = - \frac{\left(1 - \frac{2M}{r_B}\right)(T - t)}{\left[\left(1 - \frac{2M}{r_B}\right)(T - t)^2 + \text{terms without } t\right]^{1/2}} = - \left(1 - \frac{2M}{r_B}\right) \frac{T - t}{\tau_B} \quad (7.18)$$

The total wristwatch time is given by the sum $\tau = \tau_A + \tau_B$. The principle of extremal aging implies that this time must be an *extremum* (maximum actually in this case), which means:

$$\frac{d\tau}{dt} = 0 = \frac{d\tau_A}{dt} + \frac{d\tau_B}{dt} = \left(1 - \frac{2M}{r_A}\right) \frac{t}{\tau_A} - \left(1 - \frac{2M}{r_B}\right) \frac{T - t}{\tau_B} \quad (7.19)$$

Setting $t = t_A$ and $T - t = t_B$ we observe from the last equality on the right-hand-side of the above equation that

$$\left(1 - \frac{2M}{r_A}\right) \frac{t_A}{\tau_A} = \left(1 - \frac{2M}{r_B}\right) \frac{t_B}{\tau_B} \quad (7.20)$$

The procedure can be repeated for an arbitrary partition of the segment of figure 21, which implies that the following quantity is *conserved* in any segment of the stone's path (we return to differential notation for the segment's coordinate- and proper-time separations, $dt, d\tau$):

$$\boxed{\frac{E}{m} = \left(1 - \frac{2M}{r}\right) \frac{dt}{d\tau}} \quad (7.21)$$

The conserved quantity E/m has been identified with the energy per unit mass in the curved Schwarzschild geometry. This identification follows from the fact that for large r (i.e. far from the gravitational attraction) the quantity (7.21) becomes identical to the special relativity form of the energy (3.10). This latter statement is consistent with the fact that the Schwarzschild spacetime (7.1) reduces to the flat Minkowski space time for large $r \gg M$.

7.4.2 Plunging towards a centre of gravitational attraction

The existence of a conserved energy during plunging is important, and allows to study the falling of a particle towards the centre of a gravitational attraction. We examine such a plunging in the following exercise, where we use the formulæ developed above, together with the energy conservation, in order to compute the wristwatch time elapsed from the point of crossing the event horizon of a black hole till the moment of crunch (i.e. when the falling particle reaches the singularity).

Exercise 7.5 Plunging towards a Black Hole: *Starting from rest at a great distance an observer is plunging straight (i.e. radially) towards a non-rotating black hole of mass equivalent to eight solar masses, $M = 8M_{\odot}$. The observer sets his wristwatch to noon as he determines (by one means or another) that he is crossing the horizon. Determine how much time (in seconds) is left, according to the wristwatch of the observer, until the instant of crunch (i.e. when he approaches the singularity). Assume without proof the formula for the energy in the Schwarzschild geometry (7.21), involving proper (wristwatch) and far-away times.*

Solution: The geometry of the non-rotating Black Hole is that of Schwarzschild. The radial fall is sufficiently described, therefore, by the following line element ($d\theta = 0$):

$$ds^2 = -d\tau^2 = -\left(1 - \frac{2M}{r}\right) dt^2 + \left(1 - \frac{2M}{r}\right)^{-1} dr^2 \quad (7.22)$$

where τ is the proper (wristwatch) time of the plunging observer, t is the far away time, and we work in the usual system of units with $c = G_N = 1$.

During the plunge towards the singularity of the black hole, starting from rest at infinity, energy E is conserved as discussed above. If m is the mass of the plunging particle/observer, then the conserved energy E in the Schwarzschild geometry is given by (cf. (7.21)):

$$\frac{E}{m} = \left(1 - \frac{2M}{r}\right) \frac{dt}{d\tau} \quad (7.23)$$

where τ is the proper time of the Schwarzschild geometry, i.e. the wristwatch time of the plunging observer, and t is the far away time.

Because the particle/observer starts at rest at infinity ($r \rightarrow \infty$) one has that $E/m = 1$ in units of $c = 1$. Thus

$$\left(1 - \frac{2M}{r}\right) \frac{dt}{d\tau} = 1 \quad (7.24)$$

From (7.22),(7.24) we have:

$$\left(1 - \frac{2M}{r}\right)^2 dt^2 = d\tau^2 = \left(1 - \frac{2M}{r}\right) dt^2 - \frac{dr^2}{\left(1 - \frac{2M}{r}\right)} \quad (7.25)$$

Dividing through by dt^2 we can solve for dr/dt (starting from rest at $r = \infty$):

$$\frac{dr}{dt} = - \left(1 - \frac{2M}{r}\right) \left(\frac{2M}{r}\right)^{1/2} \quad (7.26)$$

We are interested in finding the correlation between r -coordinate and wristwatch time τ .

From (7.24) and (7.26) we have

$$\frac{dr}{d\tau} = \frac{dr}{dt} \frac{dt}{d\tau} = - \left(\frac{2M}{r}\right)^{1/2} \quad (7.27)$$

from which

$$d\tau = - \frac{r^{1/2} dr}{(2M)^{1/2}} \quad (7.28)$$

Since we are interested in finding the wristwatch time left from the moment of crossing the horizon ($r = 2M$) to the instant of crunch, i.e. when the plunging observer reaches the singularity at $r = 0$, we should integrate the right-hand-side of (7.28) over r , from $r = 2M$ to $r = 0$:

$$\tau = - \int_{2M}^0 \frac{r^{1/2} dr}{(2M)^{1/2}} = \frac{2}{3} \frac{(2M)^{3/2}}{(2M)^{1/2}} = \frac{4}{3} M \quad (7.29)$$

This formula expresses the result in meters. To convert the result in seconds one must divide by the value of the speed of light *in vacuo* $c = 3 \times 10^8$ m/sec, which so far, in the special system of units chosen, has been taken to be unity. In seconds, and when the mass M is expressed in units of the solar mass, this time corresponds to $\tau = 6.57 \times 10^{-6} M/M_\odot = 5.26 \times 10^{-5}$ s. Notice from (7.29) that the bigger the mass of the Black Hole M is, the longer the wristwatch time interval until the instant of crunch will be.

Exercise 7.6 (i) Show that a stone falling radially into a black hole from zero initial velocity at spatial infinity moves with the speed of light as it crosses the event horizon ($r = 2M$) as measured by nearby shell observers.

(ii) Imagine now that the stone is thrown radially into a black hole but with a non-trivial initial velocity $v_{\text{far}} < 1$ at a great distance. Show that, at distance r from the centre of the black hole of Mass M , the radial velocity observed by shell observers is:

$$\frac{dr_{\text{shell}}}{dt_{\text{shell}}} = - \left[1 - \frac{1}{\gamma_{\text{far}}^2} \left(1 - \frac{2M}{r}\right)\right]^{1/2}. \quad (7.30)$$

(iii) Thus show that it is impossible to make the final observed speed as the stone crosses the event horizon, as measured by nearby shell observers, greater than the speed of light in vacuo, which thus remains the ultimate (upper bound) velocity.

7.5 Effective Potential and Orbits in Schwarzschild Spacetimes

In this chapter we shall examine the orbits of satellites around Massive bodies by means of the Schwarzschild Geometry. A Massive body will curve the exterior spacetime according to (7.1), and in this curved geometry a satellite will follow *geodesics* according to Einstein's view of orbits. During orbit the satellite floats freely, and no force is exerted on it. One can invoke the principle of extremal aging in order to show, in a similar spirit to the conservation of energy examined above, that during the orbit the total angular momentum L will remain conserved:

$$\boxed{r^2 \frac{d\theta}{d\tau} = \text{constant} = \frac{L}{m}} \quad (7.31)$$

We shall not prove this here, but the reader is asked to remember this conservation law, which retains a similar form to that in Newtonian Mechanics, with the important difference that now the universal Newtonian time is replaced by the observer independent proper time τ .

As mentioned, one way of describing the orbit is to write down the geodesic equation (5.19) appropriate for the Schwarzschild metric. In fact we leave as an exercise to the reader to show that the geodesic equations lead automatically to the conservation of *both* the energy (7.21) and angular momentum (7.31). This is consistent with the principle of extremal aging, given that, as discussed in the relevant chapter, the geodesic equations are obtained from extremization of the proper-time interval.

Exercise 7.7 *Starting from the three-dimensional time-like version of Schwarzschild spacetime (7.4), and applying a suitable variational principle (i.e. the “Lagrange equation method”), treating the proper time τ as the affine parameter, write down the corresponding geodesic equations for a particle of mass $m > 0$ in a Schwarzschild spacetime, and from them determine the Christoffel symbols for the metric (7.4). Show then that the geodesic equations for the time t and angular coordinate θ lead automatically to the conservation of energy (7.21) and angular momentum (7.31) respectively. Express, then, the third equation (for the r coordinate) in terms of the conserved Energy per unit mass, E/m , and angular momentum per unit mass, L/m , and show that it acquires the form (Hint: divide (7.4) through by $d\tau$ and argue that this is equivalent to the r equation):*

$$\left(\frac{dr}{d\tau}\right)^2 = \left(\frac{E}{m}\right)^2 - \left(1 - \frac{2M}{r}\right) \left[1 + \frac{(L/m)^2}{r^2}\right]. \quad (7.32)$$

In what follows we shall follow an alternative method, that of the *effective potential*, which is the method we followed in the Newtonian analysis in the beginning of the course.

To find the form of the effective potential we start from the time-like form of the Schwarzschild metric (7.4). Using the law of conservation of energy (7.21) we may solve for dt in terms of the conserved energy E :

$$dt = \frac{E/m}{1 - \frac{2M}{r}} d\tau \quad (7.33)$$

In a similar way we may use the law of conservation of angular momentum (7.31), and solve for $d\theta$ in terms of L :

$$d\theta = \frac{L/m}{r^2} d\tau \quad (7.34)$$

Substituting these expressions into (7.4) and solving for dr^2 we obtain:

$$dr^2 = \left[\left(\frac{E}{m}\right)^2 - \left(1 - \frac{2M}{r}\right) \left\{ 1 + \left(\frac{L}{mr}\right)^2 \right\} \right] d\tau^2 \quad (7.35)$$

Dividing through by $d\tau^2$, then, one may obtain an expression for the square of the radial velocity $dr/d\tau$ as registered in the satellite’s wristwatch time:

$$\left(\frac{dr}{d\tau}\right)^2 = \left(\frac{E}{m}\right)^2 - \left(1 - \frac{2M}{r}\right) \left[1 + \frac{(L/m)^2}{r^2}\right] \quad (7.36)$$

The *square* of the *effective potential* V_{eff}^2 is then **defined** by writing:

$$\left(\frac{dr}{d\tau}\right)^2 = \left(\frac{E}{m}\right)^2 - \left(\frac{V_{\text{eff}}}{m}\right)^2 \quad (7.37)$$

i.e. it is defined by what we have to take away from the *square* of the total energy to get the square of the radial velocity. The reader should contrast this general relativistic situation with the

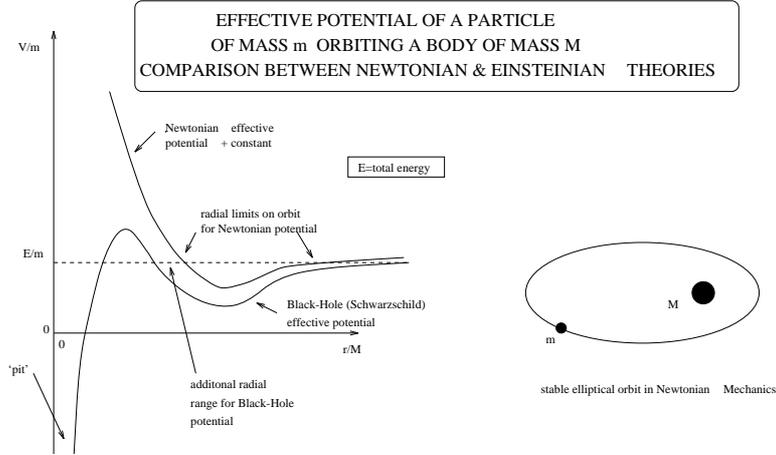


Figure 22: The Schwarzschild Effective Potential (2.29) as compared with the Newtonian potential (2.19).

corresponding Newtonian case, where one does not have squares, but simply the energy and the effective potential appearing in the formula (2.18) for the square of the Newtonian radial velocity.

Thus, in the Schwarzschild Spacetime, the effective potential per unit mass, appropriate for a satellite that orbits around a massive body, including a Black Hole, responsible for the appearance of the Schwarzschild spacetime in the exterior geometry, is given by:

$$\left(\frac{V_{\text{eff}}(r)}{m} \right)^2 \equiv \left(1 - \frac{2M}{r} \right) \left[1 + \frac{(L/m)^2}{r^2} \right] \quad (7.38)$$

Hence, as we observe from (7.38), in general relativity, which is the most appropriate theory to describe motion around a Black hole or in general a massive body, the important difference, as compared to the case of, say, the elliptical Newtonian orbits about the Sun, is the fact that, in addition to the attractive potential of gravity at great distances, and the repulsive effects of angular momentum at intermediate distances, which also characterize the Newtonian theory (2.19), Einstein's theory adds at even smaller distances a *pit* in the effective potential (7.38) (c.f. fig. 22). This pit captures a particle that comes too close, which does not happen in Newtonian theory, establishes a critical distance of closest approach for this black-hole capture process, and for a particle that approaches the critical point without crossing it lengthens the turnaround time as compared with Newtonian expectations. This lengthening of the turnaround time makes the time for radial motion longer than the period of one revolution, thereby causing the major axis of an otherwise elliptical orbit to rotate (*precession of the perihelion*), and deflects a fast particle through larger angles than a Newtonian theory would predict.

An approximate computation of the perihelion precession has already been done in *Exercise 2.1*. There, we have defined the "effective potential per unit mass" in equation (2.29) by means of the symbol U_{eff}/m . At that time we did not have a feeling of how such a term appears. Now we are well equipped to understand, from (7.38), that this symbol U_{eff}/m is actually the square of the effective potential $(V_{\text{eff}}(r)/m)^2$ of the general relativistic formalism.

However, as we have seen in the analysis of *Exercise 2.1*, one can still use the Newtonian treatment in finding (approximately) the perihelion precession, i.e. by manipulating directly the symbol U_{eff} and comparing it with the corresponding Newtonian effective potential. It should be stressed that because the relativistic effective potential (7.38) (or (2.29)) has a *constant* term on the right-hand-side, which is lacking in the corresponding Newtonian expression (2.19), one should add this constant in the latter expression when comparing it with (7.38). The Schwarzschild effective

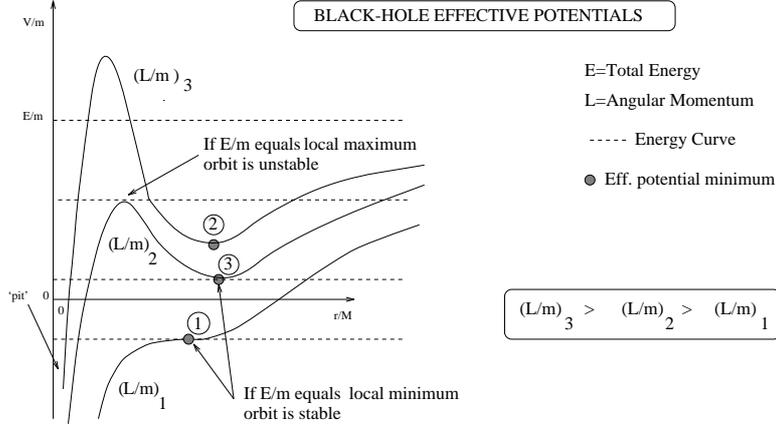


Figure 23: Various Schwarzschild-Spacetime Effective Potentials.

potential (7.38) (or equivalently the ‘symbol’ (2.29)) is plotted in the figures 22-24, and compared with the corresponding Newtonian potential (2.19).

From the form of the effective potential, and its generic features involving capture of a satellite, one may sketch the various types of orbits that are encountered in satellite motion in the Schwarzschild Geometry. This is done *qualitatively* in the figures. The reader is invited to remember these features and how they differ from the Newtonian case.

As in the Newtonian case, there are *circular orbits* in the curved Schwarzschild geometry. To get a stable circular orbit the particle’s energy must lie at the minimum of the effective potential (cf. dots in figure 23). The circular orbit is unstable if the particle’s energy is equal to the maximum of the effective potential (c.f. peaks of the barriers in fig. 23). A satellite in an unstable orbit about, say, a Black Hole, will leave the orbit under the slightest perturbation. It may then be captured by the Black Hole as indicated in figure 24.

7.6 Motion of Light in Schwarzschild Geometry

7.6.1 Null Geodesics for light

Light follows by definition *null geodesics in a curved spacetime*. Formally, the latter are defined by setting $ds^2 = 0$ in the expression of the infinitesimal proper distance in terms of the metric

$$ds_{\text{light}}^2 = -d\tau_{\text{light}}^2 = 0 = g_{\mu\nu} dx^\mu dx^\nu \quad (7.39)$$

In the context of Schwarzschild metric (7.1), this formula can be applied to give expressions for the radial and tangential motion of light.

For radial motion we set $d\theta = 0$ in (7.1) (or (7.2) and then from (7.39) we obtain:

$$\frac{dr^2}{1 - \frac{2M}{r}} = \left(1 - \frac{2M}{r}\right) dt^2 \quad (7.40)$$

from which one obtains for the radial velocity in Schwarzschild coordinates:

$$\frac{dr}{dt} = \pm \left(1 - \frac{2M}{r}\right) \quad (7.41)$$

where the two signs refer to the inward (-) or outward (+) radial direction of motion, with respect to the centre of gravitational attraction.

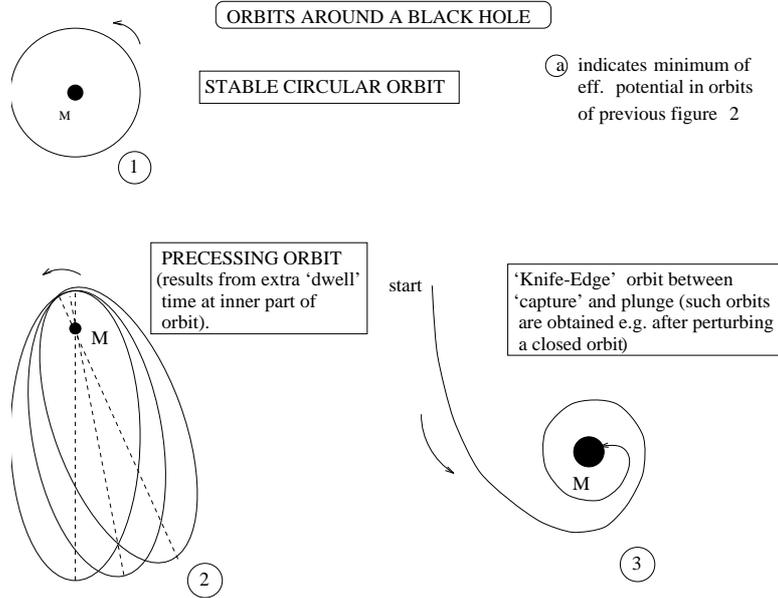


Figure 24: Various Satellite Orbits in Schwarzschild Spacetime corresponding to the Effective Potentials of fig. 23.

Exercise 7.8 Carry out a similar analysis for tangential motion. Note that $r d\theta$ is the tangential displacement, and hence $r d\theta/dt$ is the tangential Schwarzschild bookkeeper velocity. Show that:

$$r \frac{d\theta}{dt} = \pm \left(1 - \frac{2M}{r}\right)^{1/2} \quad (7.42)$$

The above results (7.41),(7.42) presents us with a puzzle. So far we have learned that the velocity of light *in vacuo*, $c = 1$ (in our units), is the ultimate upper bound of velocities, and actually is an invariant independent of observers. From the above expressions we now observe that the speed of light differs from unity near a Black Hole (or in general in a Schwarzschild Spacetime), and actually *vanishes* at the horizon ($dr/dt|_{\text{horizon}} = d\theta/dt|_{\text{horizon}} = 0$)! What is going on?

The answer is *simple*. There is no paradox in the results (7.41),(7.42) as regards the basic notion of relativity that the speed of light is an invariant. The reason is that the above formulæ (7.41),(7.42) involve bookkeeper coordinates, and as we have explained previously such coordinates cannot be used for direct measurement. They are simply accounting entries of the Schwarzschild bookkeeper. *No nearby observer* will measure the slowed speed of light.

This can be confirmed mathematically by looking at the velocity of light as measured by a shell observer. Recall the locally flat form (7.9) of the spacetime a shell observer perceives. For light we have:

$$d\tau^2 = 0 = dt_{\text{shell}}^2 - dr_{\text{shell}}^2 - r^2 d\theta^2 = dt_{\text{shell}}^2 - ds_{\text{shell}}^2 \quad (7.43)$$

and hence:

$$\frac{ds_{\text{shell}}}{dt_{\text{shell}}} = \pm 1 \quad (7.44)$$

where again the two signs refer to the direction of motion. Thus a shell observer measures (even at the horizon) the special-relativistic unit value of the speed of light. This result is valid for *all* shell observers, i.e. is an invariant, as dictated by relativity.

We next proceed to write down the equations of motion describing the orbit of light in Schwarzschild geometry. We recall that for a particle of mass m one has the following geodesic equations (which have been derived in *Exercise 7.5*):

$$\begin{aligned}\left(\frac{dr}{d\tau}\right)^2 &= \left(\frac{E}{m}\right)^2 - \left(1 - \frac{2M}{r}\right) \left[1 + \frac{(L/m)^2}{r^2}\right], \\ \frac{d\theta}{d\tau} &= \frac{(L/m)}{r^2}, \\ \frac{d\tau}{dt} &= \frac{\left(1 - \frac{2M}{r}\right)}{(E/m)}.\end{aligned}\tag{7.45}$$

For light one has the complication that $d\tau = 0$, since the light follows a null geodesic (7.39). To write down the correct equations for light, then, one must eliminate the proper time $d\tau$ from (7.45). To this end first multiply through the first two of the equations (7.45) by $d\tau/dt$, using the third equation. The result is:

$$\begin{aligned}\left(\frac{dr}{dt}\right)^2 &= \left(\frac{dr}{d\tau}\right)^2 \left(\frac{d\tau}{dt}\right)^2 = \left(1 - \frac{2M}{r}\right)^2 - \left(1 - \frac{2M}{r}\right)^3 \left[\frac{m^2}{E^2} + \frac{1}{r^2} \left(\frac{L}{E}\right)^2\right], \\ \frac{d\theta}{dt} &= \frac{d\theta}{d\tau} \frac{d\tau}{dt} = \left(\frac{L}{E}\right) \frac{\left(1 - \frac{2M}{r}\right)}{r^2}\end{aligned}\tag{7.46}$$

Then set $m = 0$, since light has zero invariant mass, which gives:

$$\begin{aligned}\frac{dr}{dt} &= \pm \left(1 - \frac{2M}{r}\right) \left[1 - \left(1 - \frac{2M}{r}\right) \frac{b_{\text{light}}^2}{r^2}\right]^{1/2}, \\ r \frac{d\theta}{dt} &= \pm \frac{b_{\text{light}}}{r} \left(1 - \frac{2M}{r}\right), \\ b_{\text{light}} &= \text{impact parameter for light} \equiv \lim_{m \rightarrow 0} \frac{\text{angular momentum}}{\text{linear momentum}} = \frac{L}{E}\end{aligned}\tag{7.47}$$

The equations (7.47) are thus the geodesic equations (in terms of Schwarzschild bookkeeper coordinates) describing orbits of light in the (exterior) neighbourhood of a non-rotating Black Hole, or in general a Massive non-rotating Body.

Exercise 7.9 Define the light speed reckoned by the Schwarzschild bookkeeper as: Light speed by bookkeeper = $\left[\left(\frac{dr}{dt}\right)^2 + \left(r \frac{d\theta}{dt}\right)^2\right]^{1/2}$. Using the equations (7.47) above, find an expression of this velocity in terms of the impact parameter b_{light} and the Mass of the Black Hole M . Show that at great distances $M/r \rightarrow 0$ this velocity approaches unity, whilst vanishes at the horizon $2M/r \rightarrow 1$. Explain why this is not in contradiction with the relativity principle that the speed of light is an invariant (unity in our set of units).

Exercise 7.10 Using the shell observer formulae (7.5), (7.6), prove that a shell observer measures the following radial and tangential velocities of light:

$$\begin{aligned}\frac{dr_{\text{shell}}}{dt_{\text{shell}}} &= \pm \left[1 - \left(1 - \frac{2M}{r}\right) \frac{b_{\text{light}}^2}{r^2}\right]^{1/2}, \\ r \left(\frac{d\theta}{dt_{\text{shell}}}\right) &= \pm \left(1 - \frac{2M}{r}\right)^{1/2} \frac{b_{\text{light}}}{r}\end{aligned}\tag{7.48}$$

Using these shell quantities we can now define the notion of the *effective potential for light*. Recalling that the square of the effective potential in the Schwarzschild geometry (7.38)

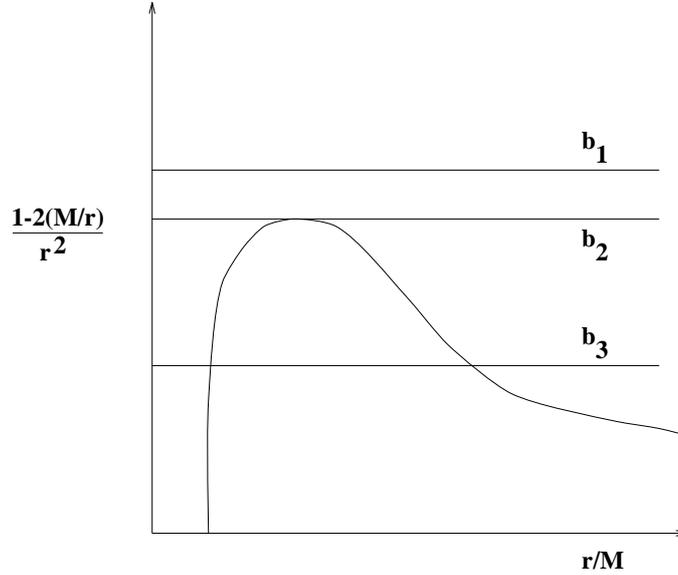


Figure 25: Square of the Effective potential for light for various impact parameters b_i , $i \in \{1, 2, 3\}$ (denoted by the horizontal solid lines). There is no minimum in this potential, therefore there are no stable circular orbits for light.

is defined by what one has to subtract from the square of the total energy in order to obtain the square of the radial velocity, we observe that one may apply this definition to the shell radial velocity $dr_{\text{shell}}/dt_{\text{shell}}$ given in (7.48). We rewrite the first of these equations as:

$$\frac{1}{b_{\text{shell}}^2} \left(\frac{dr_{\text{shell}}}{dt_{\text{shell}}} \right)^2 = \frac{1}{b_{\text{shell}}^2} - \frac{\left(1 - \frac{2M}{r}\right)}{r^2} \quad (7.49)$$

The left hand side of this equation is in some (admittedly strange!) sense a measure of the radial velocity of photon (viewed as a ‘particle’). The first term on the right-hand-side depends, through the impact parameter b_{light} , on the choice of orbit, but not on the Schwarzschild geometry. Therefore may be viewed as a constant of motion. The second term does not depend on the choice of orbit but does depend on the geometry. Hence it behaves like the square of an effective potential, and indeed this is what we shall take as our *definition* of the square of the effective potential for light:

$$\boxed{\left(\begin{array}{c} \text{effective potential} \\ \text{for light} \end{array} \right)^2 = \frac{1 - \frac{2M}{r}}{r^2}} \quad (7.50)$$

It is important to notice that this expression is actually independent of the energy of light or its impact parameter. Therefore it applies to the light of all wavelengths. Only one effective potential is needed to analyze the motion of light of any frequency in a Schwarzschild Geometry. A plot of the square of the effective potential (7.50) is given in figure 25. Since there is no no minimum in this potential, there *are no stable circular orbits for light*.

7.6.2 Bending of Light Near Massive Bodies

We are now in position to discuss the precise shape of the trajectory of light in the neighbourhood of a massive body, such as Sun *etc*. We shall do so by means of a series of instructive exercises, which concern manipulations of the equations (7.47) (c.f. figure 26).

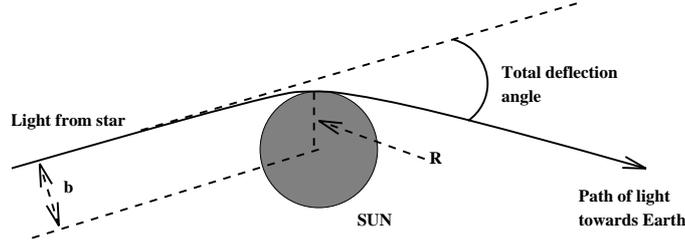


Figure 26: Deflection of a light beam as it grazes the surface of a spherically-symmetric Massive body (e.g. the Sun), of Mass M .

Exercise 7.11 Calculate the deflection angle $\Delta\theta$ of a light beam as it just grazes the surface of a spherical and non-rotating celestial object with Mass M . Perform an approximate calculation for the specific case in which the massive celestial body is the Sun, whose Mass (in meters) is $M_{\odot} = 1477$ meters, and its radius $R = 7 \times 10^8$ meters.

Solution: The light beam obeys the equations of motion (7.47), which can be rewritten as:

$$\begin{aligned} \left(\frac{dr}{dt}\right)^2 &= \left(1 - \frac{2M}{r}\right)^2 - \left(1 - \frac{2M}{r}\right)^3 \frac{b_{\text{light}}^2}{r^2}, \\ \left(\frac{d\theta}{dt}\right)^2 &= \frac{b_{\text{light}}^2}{r^4} \left(1 - \frac{2M}{r}\right)^2, \end{aligned} \quad (7.51)$$

where b_{light} is the impact parameter, which from now on we shall call b for brevity.

Divide these equations, so as to obtain an expression for $(d\theta/dr)$:

$$\left(\frac{d\theta}{dr}\right)^2 = \frac{b^2/r^4}{1 - \left(1 - \frac{2M}{r}\right) \frac{b^2}{r^2}}, \quad (7.52)$$

from which we obtain:

$$d\theta = \frac{dr}{r^2 \left[\frac{1}{b^2} - \frac{1}{r^2} \left(1 - \frac{2M}{r}\right) \right]^{1/2}} \quad (7.53)$$

In principle, in order to find the total deflection angle we should integrate this relation from $r = \infty$ to $r = R$ where R is the radius of the Massive body of mass M . The total deflection is *twice this result* (cf. figure 26). Unfortunately, the right-hand-side of equation (7.53) does not exist in integral tables, so we need to make some physically meaningful *approximations* to get a correct estimate of the deflection angle.

Physical Approximations for computing deflection of light by the Sun:

The first step is standard in orbital mechanics. Change variable from r to $u \equiv R/r$. Then $dr = -r^2 du/R$ and hence the integral from $r = R$ to $r = \infty$ (outwards) now becomes an integral from $u = 1$ to $u = 0$ respectively.

With these in mind, then, the total deflection is (remember we multiply the integrated result (7.53) by 2, in order to get both 'legs' of light trajectory (cf. figure 26)):

$$\theta_{\text{total}} = 2 \int d\theta = -2 \int_1^0 du \frac{1}{\left[\frac{R^2}{b^2} - u^2 + 2 \frac{M}{R} u^3 \right]^{1/2}} \quad (7.54)$$

There is an important point to notice here. The impact parameter b is a function of both M, R , because the impact parameter depends on the orbit, or better characterizes the orbit, and hence if

we wish to consider the orbit of figure 26, where the light grazes the surface of the Massive Body, then we must *choose* an appropriate b .

The question now arises as to how one can we compute b . This can be done by going to the shell observer coordinates (the shell now is the surface of the Body of mass M .), and in particular the relation (7.48) for the radial shell velocity $dr_{\text{shell}}/dt_{\text{shell}}$, which we rewrite as:

$$\frac{1}{b^2} \left(\frac{dr_{\text{shell}}}{dt_{\text{shell}}} \right)^2 = \frac{1}{b^2} - \frac{\left(1 - \frac{2M}{r}\right)}{r^2} \quad (7.55)$$

As the light grazes the surface of the Massive Body, it only moves tangentially, and hence the radial velocity should vanish. Setting $r = R$ in the above relation and requiring that the left-hand-side vanishes, yields:

$$\frac{R^2}{b^2} = 1 - \frac{2M}{R} \quad (7.56)$$

which thus gives b in terms of M, R . Substituting into (7.54) we have:

$$\theta_{\text{total}} = -2 \int_1^0 \frac{du}{\left[1 - u^2 - \frac{2M}{R}(1 - u^3)\right]^{1/2}} = -2 \int_1^0 \frac{(1 - u^2)^{-1/2} du}{\left[1 - \frac{2M}{R} \frac{(1 - u^3)}{1 - u^2}\right]^{1/2}} \quad (7.57)$$

For the case of Sun the quantity $2M/R = 2M_{\odot}/R \simeq 4 \times 10^{-6} \ll 1$, hence one may use the standard approximation (*binomial expansion*):

$$(1 + x)^n \simeq 1 + nx, \quad \text{provided } |x| \ll 1, |nx| \ll 1. \quad (7.58)$$

Applying this approximation to equation (7.57) we have:

$$\theta_{\text{total}} \simeq 2 \int_1^0 \left[\frac{-du}{(1 - u^2)^{1/2}} - \frac{M}{R} \frac{du}{(1 - u^2)^{3/2}} + \frac{M}{R} \frac{u^3 du}{(1 - u^2)^{3/2}} \right] \quad (7.59)$$

We use the following (indefinite) integrals (which in case they are needed will be provided):

$$\int \frac{-du}{(1 - u^2)^{1/2}} = -\arcsin u + \text{const} \quad (7.60)$$

and

$$\int \left[\frac{du}{(1 - u^2)^{3/2}} - \frac{u^3}{(1 - u^2)^{3/2}} \right] = \frac{u}{(1 - u^2)^{1/2}} - (1 - u^2)^{1/2} - \frac{1}{(1 - u^2)^{1/2}} \quad (7.61)$$

From which (7.59) becomes:

$$\theta_{\text{total}} \simeq \pi + \frac{4M}{R} \quad (7.62)$$

Exercise 7.12 Discuss carefully the lower limit of integration $u = 1$ in arriving at the result (7.62).

The term π is the limit of θ_{total} as $M \rightarrow 0$, i.e. it describes the path of light (straight line) in flat spacetimes. Thus the required *deflection* of light $\Delta\theta$ is the difference of θ_{total} from this value π :

$$\boxed{\Delta\theta \simeq \frac{4M}{R}} \quad (7.63)$$

which is the famous formula describing deflection of light by the Sun, that gave Einstein instant fame.

Exercise 7.13 For the case of Sun, $M = M_{\odot} = 1477$ meters, and $R = 7 \times 10^8$ meters, express the result for $\Delta\theta$ (7.63) in radians.

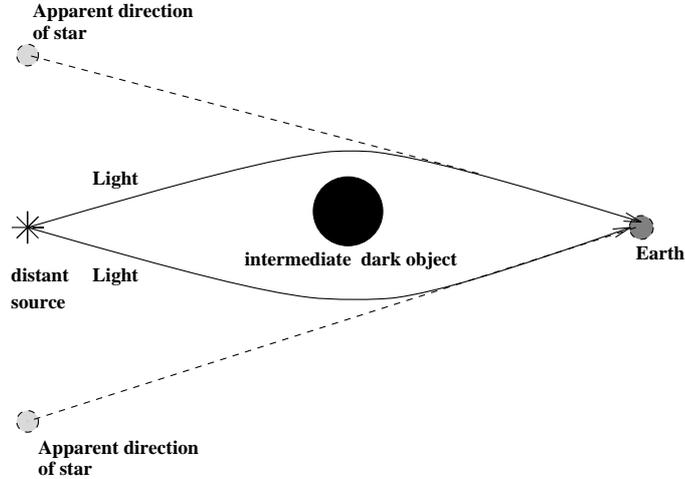


Figure 27: The principle of gravitational lensing. The lensing is produced as a result of deflection of light beams from a distant astrophysical source as they pass near an intermediate massive dark object (e.g. a cluster of galaxies).

7.6.3 Gravitational lensing

An important consequence of this formula is *gravitational lensing*, whose geometric construction is depicted in figure 27. The gravitational lensing is caused when light from a distant astrophysical source is deflected by an intermediate dark object (for instance a cluster of galaxies *etc.*). In the previous calculation of the deflection of light by the Sun we used an orbit that grazed the surface of the Sun, because this had the dominant effect for our purposes there, as we can see from (7.55).

However, for the situation depicted in figure 27 we should not consider only orbits that graze the surface of the intermediate dark object. From equation (7.55) we observe that, for large distances $r \gg 2M$, which is the case of distant astrophysical sources, one has simply that $b \simeq R$, in which case equation (7.55) becomes:

$$\Delta\theta_{\text{lensing}} \simeq \frac{4M}{b} \quad (7.64)$$

The gravitational Lensing has become now an important technique in providing us information about the existence of intermediate cluster of galaxies *etc.*, and other “dark” celestial objects, whose detection would be impossible otherwise.

8 An Introduction to Cosmology

8.1 What is Cosmology?

The solar system and the dynamics of its constituents, and even the dynamics of galaxies are adequately described by Newtonian gravitational theory. However when we wish to discuss the Universe as a whole, i.e. discuss physics on scales far larger than that of clusters of galaxies, then general relativity becomes important. To understand the above statements recall that a measure of whether Newtonian gravity is sufficient is provided by ratios of the form M/R where M is some typical mass scale and R some typical distance scale. Newtonian mechanics is applicable to the cases where $M/R \ll 1$ while general relativity is expected to become important $M/R \geq 1$. To get a typical idea of the numbers involved in this ratio notice that in a galaxy, which contains, say,

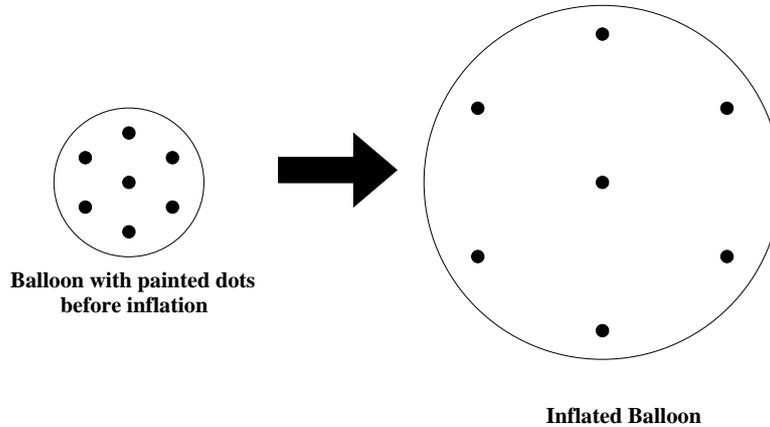


Figure 28: The balloon model for understanding Hubble’s law: paint dots on a balloon, inflate it, and then you observe that as it grows all relative distances between marked points grow at a rate proportional to their magnitudes.

10^{11} stars in a radius of about 15 kpc ($= 4.5 \times 10^{20}$ m) and hence the ratio M/R in this case is 10^{-6} .

The branch of general relativity which examines the Universe as a whole is called *cosmology*. There are two basic assumptions (which are supported by observations) underlying cosmological models:

1. Isotropic: the Universe looks the same in every direction (at least on sufficiently large scales).
2. Homogeneity: the Universe is isotropic about every point. This implies that there is a uniformity in the composition of the Universe about every point, i.e. the Universe is characterized by a uniform energy density, uniform distribution of galactic types, with uniform chemical and stellar composition etc. Further, this implies that there is no special point in the Universe.

An important observation (due to E. Hubble in 1929) is that distant galaxies are measured to recede from each other with speeds proportional to their separation. The stipulation “distant” galaxies is necessary to remove effects of local clustering (as in the local group); for example, our nearest large neighbour, the Andromeda galaxy, is not receding from us at the rate predicted by Hubble’s law. Hubble’s law in mathematical terms states:

$$v = Hd, \tag{8.1}$$

where v is the recession speed of the galaxy and d is the distance from the observer. The constant (which is truly only a parameter, as it may vary in time, as we shall see later) H is called “Hubble’s constant” and has the approximate numerical value $71 \pm 6 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (1 Mpc = 10^3 kpc) today. Hubble’s law may be understood schematically by the model of the balloon (see figure 28). Paint dots on a balloon and then inflate it. As it grows, the distance on the surface of the balloon between any two points grows at a rate proportional to that distance.

Observational cosmology deals necessarily with the history of the Universe based on astrophysical observations. Theoretical cosmology, on the other hand, studies the past and attempts to make predictions for the future. But basically, Cosmology is a study of our *past*. If we are to make some large-scale model of the Universe we must make some assumptions about regions that we have no way of observing, because they are too distant to be seen by our telescope. In the Universe, if the latter is assumed to have a *finite Age*, which seems to be the case with our Universe, there are two kinds of inaccessible regions.

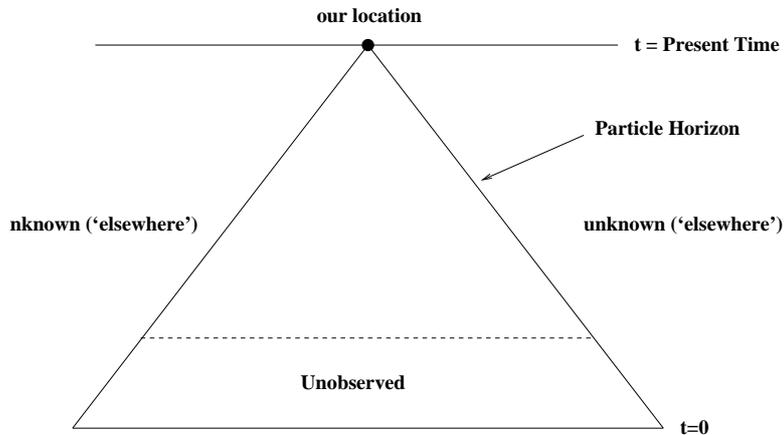


Figure 29: Schematic spacetime diagram showing the past history (past lightcone) of our Universe. The unknown regions have not had the time to send us information if we assume finite age of the Universe. The unobserved regions are obscured by intervening matter. Every moment more and more of the unknown regions enter our particle horizon.

1. The first is the region which is so distant that in a Universe of finite Age no information (traveling on a null geodesic) could reach us, no matter how early this information began traveling. These regions are termed ‘unknown’ (or ‘elsewhere’) in the spacetime diagram depicted in figure 29, which is a kind of past light cone for our Universe (recall the relevant discussion on figure 4. Such unknown regions have no influence when we study our past, because they cannot affect the interior of our past light cone. On the other hand this past light cone is a kind of ‘horizon’, called ‘particle horizon’, since every moment more and more of these unknown regions enter our light cone, and hence they *can affect* our future. For instance, we have evidence from observations, as mentioned earlier, that our Universe is pretty much homogeneous and isotropic on large scales. But, if tomorrow, some of the unknown regions enter the past light cone, and reveal some inhomogeneity on a large scale, then we will certainly have to revise our model for the Universe. It is in this sense that Cosmology is really a retrospective study, since it really tries to help us understand our past, and the predictions that it makes are based on models that have been constructed in order to give agreement with things that occurred in the past. We cannot really know whether the homogeneous and isotropic large-scale properties of the part of the universe lying inside our past light cone are shared by the unknown regions outside this light cone. If, however, such inhomogeneous regions existed, then we would have presented with a philosophical puzzle, as to why the Universe until now was observed to be homogeneous³. It is to avoid such kind of difficult and rather puzzling issues that many scientists believe that homogeneity and isotropy also characterize the unknown regions. This is called the *cosmological principle*.
2. The second kind of unknown regions are the part of the interior of the light cone underneath the dashed line in figure 29. This region includes matter (galaxies *etc.*) that is so distant that our instruments (telescopes) cannot get any information about them. Such regions may be reached in the future by improving our means of detection, e.g. if we manage to detect gravitational waves, then we might be able of obtaining information on such distant sources, given that all other means of obtaining information (e.g. electromagnetic waves) cannot work. However, such regions are not *in principle unknowable*, and in fact the study of cosmology may help us understand what happened in regions of which direct observation is not possible at present. Recall, that we really know nothing about very early cosmology,

³It should be remarked that there are some scientists who believe that the Universe at a very early stage was not homogeneous and/or isotropic. We shall not deal with such questions in the context of this course.

and as mentioned previously, there are even models of the very early universe, suggesting that homogeneity and isotropy were not valid at such early stages. At any rate, in the limited purposes of our undergraduate course we shall not be dealing with such early cosmological models.

8.2 General Relativistic Cosmological Models

8.2.1 The Robertson–Walker Spacetime

The cosmological principle stated above asserts that the three-dimensional space is a space of maximal symmetry, that is a space with constant curvature at a given time (but the curvature will in general change with the time). The most general four-dimensional metric which satisfies these criteria is the *Robertson–Walker* metric. To understand the underlying geometry let us first proceed with the construction of the Robertson-Walker metric in a toy Universe, living in two-spatial-dimensions, where we can visualise things. As a first step, towards the construction, the toy Universe is represented as a shaded two-dimensional surface, *embedded* in a “fictitious” three dimensional sphere (Euclidean 3-sphere) of Radius R , as in fig. 30.

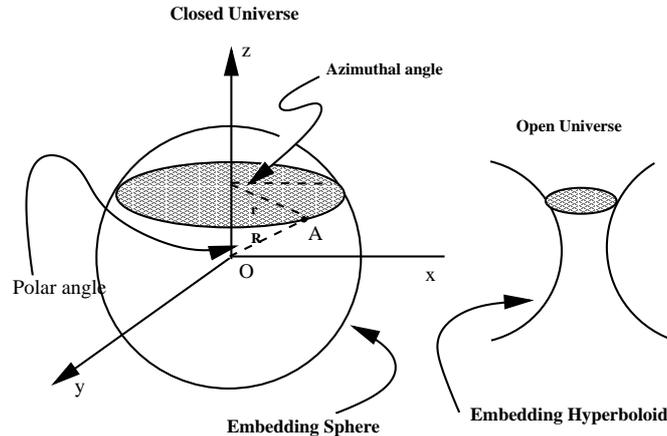


Figure 30: Left: Embedding of a toy Universe living in two-spatial-dimensions (two-sphere) in a fictitious three dimensional Euclidean sphere (three-sphere). The point A of the two-sphere, corresponding to polar coordinates (\hat{r}, θ) (on the shaded plane), or (θ, ϕ) polar-azimuthal (respectively) angles of spherical coordinates, denotes a point of the toy Universe. This is the first step to understand the geometrical construction of a Robertson-Walker space time. The radius R of the three-sphere is eventually made to depend on the cosmic time, which results in the full Robertson-Walker Universe. This construction corresponds to a closed Universe. Right: To obtain an open (hyperbolic) toy Universe we replace the real radius R by a purely imaginary number iR , corresponding to an embedding in a three-dimensional hyperboloid. The spatially flat Universe is obtained as the limit $R \rightarrow \infty$. These constructions can be generalised straightforwardly (but cannot be visualised) to four space-time dimensions.

Consider a point A on this toy Universe, with Cartesian coordinates (x, y, z) on the fictitious embedding sphere. In terms of polar coordinates \hat{r}, θ on the x_3 -plane ($x_3^2 = R^2 - \hat{r}^2$, depicted as shaded in fig. 30) we write ⁴:

$$x = \hat{r} \cos \theta, \quad y = \hat{r} \sin \theta \quad (8.2)$$

⁴Note that in our construction the angle θ coincides with the azimuthal angle of spherical coordinates.

with the equation for the 3-sphere:

$$x^2 + y^2 + z^2 = R^2, \quad z = \pm \sqrt{R^2 - \hat{r}^2}, \quad dz = \mp \frac{d\hat{r}}{\sqrt{R^2 - \hat{r}^2}} \quad (8.3)$$

The differential line element for the 3-sphere then reads:

$$d\ell_3^2 = dx^2 + dy^2 + dz^2 = d\hat{r}^2 + \hat{r}^2 d\theta^2 + \frac{d\hat{r}^2}{R^2 - \hat{r}^2} \quad (8.4)$$

Passing into dimensionless quantities, $r \equiv \frac{\hat{r}}{R}$ we can write:

$$d\ell_3^2 = R^2 \left(\frac{dr^2}{1 - r^2} + r^2 d\theta^2 \right) \quad (8.5)$$

We can construct a (2+1)-dimensional space-time for a description of our toy *homogeneous* and *isotropic* Universe, by adding the cosmic time t , with as Minkowski signature, and making the radius R cosmic-time dependent, $R \rightarrow R(t) \equiv a(t)R_0$, where R_0 can be taken to be the size of the Universe today, and $a(t) \equiv R(t)/R_0$ is the so-called scale factor. The pertinent space-time invariant line element of the (2+1)-dimensional Universe reads:

$$ds_{\text{closed}}^2 = -dt^2 + d\ell_3^2 = -dt^2 + a(t)^2 R_0^2 \left(\frac{dr^2}{1 - r^2} + r^2 d\theta^2 \right) \quad (8.6)$$

From now on we shall work in units where $R_0 = 1$ for convenience.

As the Universe expands or contracts, the coordinates (\hat{r}, θ) remain unchanged, they are “*co-moving*”. Also note that the physical distance between two co-moving points in the space of a homogeneous and isotropic Universe scales with R , hence the name *scale factor*.

Above we used planar polar coordinates (on the x_3 -plane) for the description of a point on the two-sphere. Equivalently, we can represent the point A by using the angular spherical polar coordinates (c.f. fig. 30), comprising of polar (ϕ) and azimuthal (θ) angles (in our notation). In such a case, $x = R \sin\phi \cos\theta$, $y = R \sin\phi \sin\theta$ and $z = R \cos\phi$. The spatial infinitesimal element of the embedding three-sphere then becomes in that case: $d\ell_3^2 = R^2 (d\phi^2 + \sin^2\phi d\theta^2)$. It is customary to revert to the usual notation of spherical polar coordinates $\phi \rightarrow \theta$ for the polar, and $\theta \rightarrow \phi$ for the azimuthal angle, hence in terms of such coordinates:

$$d\ell_3^2 = R^2 (d\theta^2 + \sin^2\theta d\phi^2) \quad (8.7)$$

The above construction corresponds actually to a *closed* toy Universe, since as can be seen by the embedding of fig. 30, for real (positive) radius of the embedding sphere, $R > 0$, the two-dimensional *space* of the Universe is bounded (at any *given moment* of the cosmic time) by the finite surface of the three sphere. However, we may now consider an embedding of the toy two-space-dimensional Universe in a *hyperbolic* 3-surface (c.f. right panel of fig. 30), in such a way so that there is no natural end of space, since the hyperbolic surface does not end. This Universe is called *open*, and is obtained formally from the above construction by making the radius R purely imaginary, $R \rightarrow iR$. From (8.6), then, we obtain in such a case

$$ds_{\text{open}}^2 = -dt^2 + a(t)^2 R_0^2 \left(\frac{dr^2}{1 + r^2} + r^2 d\theta^2 \right) \quad (8.8)$$

In such a case, one can also use the usual spherical polar angular coordinate system for a description of a point in the toy hyperbolic Universe, with the spatial-infinitesimal element for the embedding three-surface being given by:

$$d\ell_3^2 = R^2 (d\theta^2 + \sinh^2\theta d\phi^2) \quad (8.9)$$

Finally, in the limit $R \rightarrow \infty$ one obtains a spatially flat Universe. In such a case (8.6) yields

$$ds_{\text{flat}}^2 = -dt^2 + a(t)^2 R_0^2 (dr^2 + r^2 d\theta^2) \quad (8.10)$$

where now r is the usual polar coordinate. Notice, though, that in the flat Universe case, we still leave the scale factor in front of the spatial part of the metric element (8.10), after taking the infinite-radius-embedding limit.

The above construction can be straightforwardly generalised (but cannot be visualised) to four space-time dimensions; one can start from a three spatial-dimension Universe, embedded in a fictitious four-dimensional Euclidean sphere. Using the appropriate spherical polar coordinates, and following the above construction, we can arrive at the four space-time dimensional Robertson-Walker metric

$$\boxed{ds^2 = -dt^2 + a^2(t) \left\{ \frac{dr^2}{1 - kr^2} + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2 \right\}} \quad (8.11)$$

In the above, $a(t)$ is the the *scale factor* of the Universe, which is a measure of the size of the Universe at a given time in the coordinate system of (8.11), and k is a constant. By a rescaling of the coordinate r it can be shown (c.f. two-dimensional example above) that k may take only one of the three values $\{-1, 0, +1\}$. These values characterize three types of Universe, which we now discuss.

- $k = +1$ First it is convenient to redefine the coordinate system such that $r \mapsto \chi(r)$ as follows:

$$d\chi^2 = \frac{dr^2}{1 - r^2} \quad \Rightarrow \quad r = \sin \chi. \quad (8.12)$$

This is the higher-dimensional analogue of the two-dimensional angular coordinate system, discussed previously, which corresponds to the angular coordinate system of a four-dimensional embedding sphere, (χ, θ, ϕ) . In this angular system, the Robertson-Walker metric (8.11) for the spatial coordinates (i.e. fixing the time at, say, $t = t_0$) is

$$d\ell^2 = a^2(t_0) \left\{ d\chi^2 + \sin^2 \chi (d\theta^2 + \sin^2 \theta d\phi^2) \right\}. \quad (8.13)$$

It can be shown that this is actually the metric of a three-sphere of radius $a(t_0)$ embedded in a four-dimensional Euclidean space (c.f. two-dimensional example above). This model describes a *closed* or spherical Robertson-Walker spacetime. For a Universe with $a(t)$ a monotonically increasing function of t this corresponds to the balloon picture depicted in figure 28.

- $k = 0$ In this case at any moment in time ($t = t_0$) the spatial part of the Robertson-Walker metric (8.11) reduces to that of a *flat* Euclidean space

$$d\ell^2 = d\bar{r}^2 + \bar{r}^2 (d\theta^2 + \sin^2 \theta d\phi^2) \quad (8.14)$$

where $\bar{r} = a(t_0)r$ is a rescaled radial coordinate. This is the *flat* Robertson-Walker Universe, which is obviously homogeneous and isotropic.

- $k = -1$ In this case we transform the radial coordinate (in analogy with the $k = +1$ case) $r \mapsto \xi(r)$ such that

$$d\xi^2 = \frac{dr^2}{1 + r^2}, \quad \Rightarrow \quad r = \sinh \xi. \quad (8.15)$$

Hence the spatial part of the metric (8.11) at time $t = t_0$ becomes

$$d\ell^2 = a^2(t_0) \left\{ d\xi^2 + \sinh^2 \xi (d\theta^2 + \sin^2 \theta d\phi^2) \right\}. \quad (8.16)$$

This is the *hyperbolic* or *open* Robertson-Walker Universe. In this Universe we observe that as the proper radial coordinate ξ increases away from the origin, the circumferences of the spheres increase as $\sinh^2 \xi > \xi^2$ (for positive ξ). Thus the circumferences increase more rapidly than in flat space; this space is not realizable (in contrast with the $k = +1$ case) as a three-dimensional hypersurface embedded in a Euclidean four-dimensional space (i.e. we cannot draw it). Since the circumferences grow unboundedly with ξ this universe is open, i.e. there is no natural end to the space.

8.2.2 The Geometry of the Robertson-Walker Universe

It will be convenient in what follows to re-write the space-time metric (8.11), corresponding to the Robertson-Walker Universe, as follows:

$$ds^2 = -dt^2 + h_{ij} dx^i dx^j, \quad i, j = 1, 2, 3. \quad (8.17)$$

The coordinates x^i represent either the ‘‘polar coordinates’’ (r, θ, ϕ) or the spherical angular coordinates (χ, θ, ϕ) of the embedding four-surface. In this form, the Christoffel symbols are particularly easy to compute, as a result of the symmetry of the metric. The non-trivial components are:

$$\Gamma_{ij}^0 = \frac{\dot{a}}{a} h_{ij}, \quad \Gamma_{0j}^i = \frac{\dot{a}}{a} \delta_j^i, \quad \Gamma_{jk}^i = \frac{1}{2} h^{il} (h_{lj, k} + h_{lk, j} - h_{jk, l}). \quad (8.18)$$

The three-dimensional spatial metric $ds_3^2 = h_{ij} dx^i dx^j$ is *maximally symmetric*, with the corresponding components of the three-dimensional-space Riemann tensor being given by:

$$\begin{aligned} {}^3R_{ijkl} &= \frac{k}{a^2(t)} (h_{ik} h_{jl} - h_{il} h_{kj}), \\ {}^3R_{ij} &= \frac{2k}{a^2(t)} h_{ij}, \quad {}^3R = \frac{6k}{a^2(t)}, \end{aligned} \quad (8.19)$$

thereby clarifying the rôle of the parameter k as being characteristic for the spatial curvature of the Universe. For $k = +1$ the (positive) spatial curvature is that of a three-sphere of radius $a(t)$ (we remind the reader that we are working in units of $R_0 = 1$). For $k = -1$ the spatial curvature is negative, as corresponding to a three-hyperboloid, while for $k = 0$ the space is flat, having a vanishing three-Riemann tensor.

For future use we also give the four-dimensional space-time Ricci and scalar-curvature tensors:

$$\begin{aligned} R_{00} &= -3 \frac{\ddot{a}}{a}, \quad R_{ij} = \left(\frac{\ddot{a}}{a} + 2 \frac{(\dot{a})^2}{a^2} + 2 \frac{k}{a^2} \right) h_{ij}, \\ R &= 6 \left(\frac{\ddot{a}}{a} + \frac{(\dot{a})^2}{a^2} + \frac{k}{a^2} \right). \end{aligned} \quad (8.20)$$

This completes our brief discussion on the geometrical characteristics of the Robertson-Walker space-time. In what follows we shall make frequent use of these results.

8.2.3 The Hubble Law

We shall now derive Hubble’s law (8.1) in this spacetime. First it is convenient to rewrite the three cases above in a unified notation:

$$ds^2 = -dt^2 + a^2(t) \left\{ d\chi^2 + f^2(\chi) (d\theta^2 + \sin^2 \theta d\phi^2) \right\}, \quad (8.21)$$

where

$$f(\chi) = \begin{cases} \sin \chi & \text{for } k = +1, \\ \chi & \text{for } k = 0, \\ \sinh \chi & \text{for } k = -1. \end{cases} \quad (8.22)$$

The coordinate χ is related to the distance d from, say, a star or galaxy at rest with respect to the coordinate system (8.21), as follows from (8.21) by looking only at radial situations (i.e. setting $dt = d\theta = d\phi = 0$), and integrating $d = \int |ds| = \int_0^\chi a(t) d\chi$:

$$d = a(t) \chi. \quad (8.23)$$

From this we obtain

$$\frac{\partial d}{\partial t} = \frac{\dot{a}}{a}d, \quad \dot{a} \equiv \frac{\partial a}{\partial t}. \quad (8.24)$$

From the last equation we observe that the rate of increase of the distance d (i.e. the recession speed of the object) is proportional to the distance itself (d) which is Hubble's law (8.1). The Hubble parameter is *not* a constant in general, for it is given by

$$H(t) = \frac{\dot{a}(t)}{a(t)}. \quad (8.25)$$

Note that the scale factor a is a slowly-varying function of t and we find that the galaxies we can observe are sufficiently close (i.e. not far in the past) for the Hubble parameter to be nearly constant (with the value given above) for the observations we can make.

The present-day Hubble constant is measured today to have the value:

$$H_0 = 100h \text{Km sec}^{-1} \text{ Mpc} = 2.1332 h 10^{-42} \text{ GeV}, \quad \text{with } 0 \leq h \leq 1. \quad (8.26)$$

$$\text{WMAP 2006 measurement, } h_{\text{WMAP}} \sim 0.71, \quad (8.27)$$

where h is known as the reduced Hubble constant.

8.3 Motion of light in Robertson–Walker spacetimes: the cosmological redshift

Light follows *null geodesics* in the Robertson–Walker spacetime (8.11). We can choose a coordinate system in which light travels radially (i.e. $\theta = \phi = \text{constant}$ in (8.21)):

$$ds^2 = 0 = g_{\mu\nu}dx^\mu dx^\nu = -dt^2 + a^2(t)d\chi^2. \quad (8.28)$$

We define the four-momentum of the photon as $p_\mu = g_{\mu\nu}dx^\nu/d\lambda$ where λ is an affine parameter of the null radial geodesic. Notice that the momentum is defined as a covariant vector, as follows from the Lagrangian definition,

$$p_\mu = \frac{\partial L}{\partial \dot{x}^\mu}, \quad \dot{x}^\mu \equiv \frac{dx^\mu}{d\lambda}, \quad (8.29)$$

and L is the Lagrangian, which in this case is taken to be $L = \frac{1}{2}g_{\mu\nu} \frac{dx^\mu}{d\lambda} \frac{dx^\nu}{d\lambda}$. In the case of a homogeneous Universe the radial component of the four-momentum, p_χ is *constant* along the geodesics, as follows from Lagrange's equations (i.e. the radial geodesics equations of the spacetime (8.28)):

$$\frac{dp_\chi}{d\lambda} = \frac{\partial(g_{\chi\chi}p^\chi)}{\partial \chi} = 0. \quad (8.30)$$

since $g_{\mu\nu}$ is independent of the radial coordinate χ . Therefore we may normalize $p_\chi = -1$ (the minus sign is due to the fact that the direction of the photon is towards the observer). Thus, from the nullness of p_μ , $g^{\mu\nu}p_\mu p_\nu = 0$, we have $p_0^2 = 1/a^2(t)$, from which

$$p_\mu = \left(-\frac{1}{a(t)}, -1, 0, 0\right). \quad (8.31)$$

where the minus sign in the temporal component is taken because by definition $p^\mu = (E, \vec{p})$, the energy of the photon $E > 0$, and $p_\mu = g_{\mu\nu}p^\nu$, with $g_{00} = -1$ in our sign convention for the metric.

Suppose now that the photon is emitted from a source which is at rest with respect to the cosmological frame (8.21) and received by an observer also at rest in the same frame. In general

if an observer moves through the spacetime with four-velocity u^μ the energy (equivalently the frequency ν) of the photon as measured by this observer is given in an invariant way (cf. exercise 3.9) by

$$\nu = -g_{\mu\nu}p^\mu u^\nu. \quad (8.32)$$

The frequencies measured in the rest system of the source and by the observer are related by

$$\frac{\nu_{\text{source}}}{\nu_{\text{obs}}} = \frac{(g_{\alpha\beta}p^\alpha u^\beta)_{\text{source}}}{(g_{\alpha\beta}p^\alpha u^\beta)_{\text{obs}}}. \quad (8.33)$$

Applying this formula to sources and receivers at rest with respect to the cosmological rest frame, i.e. $u^\mu = (1, \vec{0})$, $p_\mu = g_{\mu\nu}p^\nu = (-\frac{1}{a(t)}, -1, 0, 0)$ we have (setting the emission time to t_{emit} and the observation time to t_{obs})

$$\frac{\nu_{\text{source}}}{\nu_{\text{obs}}} = \frac{a(t_{\text{obs}})}{a(t_{\text{emit}})}. \quad (8.34)$$

Alternatively,

$$\nu a = \text{constant}. \quad (8.35)$$

The *redshift* z is defined as the relative change in wavelengths $\lambda = \nu^{-1}$,

$$z = \frac{\lambda_{\text{obs}} - \lambda_{\text{emit}}}{\lambda_{\text{emit}}} = \frac{a(t_{\text{obs}})}{a(t_{\text{emit}})} - 1. \quad (8.36)$$

or, in a more familiar notation, if the scale factor of the current era when the observations are performed is denoted by $a(t_{\text{obs}}) \equiv a_0$, and the emission took place at a time in the past $t < t_{\text{today}} = t_{\text{obs}}$ of the expanding Universe, at an era with scale factor $a(t) < a_0$:

$$\frac{a(t)}{a_0} = \frac{1}{1+z}, \quad z > 0. \quad (8.37)$$

This formula gives the *cosmological redshift*, which we see occurs because between the time of emission and observation the Universe will in general change its scale factor a . This latter effect is due to the spacetime curvature of the Robertson-Walker Universe. The reader is invited to compare this result with the result on the gravitational redshift in the Schwarzschild curved geometry of exercise 7.3. Both results are due to the curved geometry.

Notice from (8.33) that in general, if the motion of the source is taken into account there will be, in addition to contributions from the curved geometry, also *Doppler* (Special Relativistic) contributions to the redshift, as a result of the spatial velocity \vec{v} of the source (cf. equation (3.26) in exercise 3.10).

8.4 Dynamics of Robertson–Walker Spacetimes: a Non-vacuum Global Solution to Einstein’s Equations in the presence of a Cosmic Fluid

To understand the dynamics of the metric (8.11) we notice that it is a solution of Einstein’s equations (6.39) upon modelling the Universe as a *perfect fluid* with energy density $\rho(t)$ and pressure $p(t)$, i.e. its stress-energy tensor is given by (6.2):

$$T^{\mu\nu} = pg^{\mu\nu} + (p + \rho)u^\mu u^\nu \quad (8.38)$$

which satisfies the covariant conservation equation (6.1):

$$\boxed{T^{\mu\nu}{}_{;\nu} = 0} \quad (8.39)$$

This latter equation does not give independent information from Einstein's equations, since as we have seen in section 6, Einstein's equations automatically imply (8.39) due to the Bianchi identities (5.59) of the curvature tensor. We shall see explicitly this fact in exercise 8.3 below.

Below we shall study the most important dynamical properties of the Robertson–Walker space-time by means of a series of instructive exercises, which involve mathematical manipulations of the pertinent Einstein's equations. The reader should keep in mind that the Robertson–Walker metric (8.11) is *not a vacuum solution* of Einstein's equations.

Upon the assumption of a perfect-fluid stress-energy tensor (8.38), Einstein's equations for a Robertson–Walker (RW) universe (8.11), with scale factor $a(t)$, assume the form:

$$\begin{aligned} -3\frac{\dot{a}^2}{a^2} - 3\frac{k}{a^2} + \Lambda &= -8\pi G_N \rho , \\ -2\frac{\ddot{a}}{a} - \frac{\dot{a}^2}{a^2} - \frac{k}{a^2} + \Lambda &= 8\pi G_N p, \end{aligned} \quad (8.40)$$

where k is the usual characteristic parameter of the RW cosmology, G_N is Newton's constant, Λ is the cosmological constant, ρ is the energy density, and p is the pressure.

Exercise 8.1 *Show that from the equations (8.40) one can deduce the following:*

$$\dot{\rho} + 3(\rho + p)\frac{\dot{a}}{a} = 0 , \quad (8.41)$$

$$\frac{\ddot{a}}{a} + \frac{4\pi G_N}{3}(\rho + 3p) = \frac{\Lambda}{3} . \quad (8.42)$$

Solution

To obtain equation (8.41) one should first differentiate the first of the RW equations (8.40) with respect to time. One then gets:

$$-8\pi G_N \dot{\rho} = -6\left(\frac{\dot{a}}{a}\right) \left[\frac{\ddot{a}}{a} - \left(\frac{\dot{a}}{a}\right)^2 \right] + 6k\frac{\dot{a}}{a^3} . \quad (8.43)$$

Subtracting the equations (8.40), one can then solve for the quantity $\frac{\ddot{a}}{a} - \left(\frac{\dot{a}}{a}\right)^2$, to obtain:

$$\frac{\ddot{a}}{a} - \left(\frac{\dot{a}}{a}\right)^2 = \frac{k}{a^2} - 4\pi G_N(\rho + p) .$$

To show the required result we then need to substitute the last expression into the expression (8.43) for $\dot{\rho}$. The required result, then, follows immediately.

To get equation (8.42) multiply the second of eqs. (8.40) by 3 and subtract these two equations, the result follows immediately.

There are two cases of fluids in Cosmology which are of particular interest. The first is the so-called *matter dominated era*, which is essentially 'dust', characterized by $p = 0$, and which is the present epoch. The second is the early-universe radiation dominated era, characterized by $p = \frac{1}{3}\rho$.

Exercise 8.2 *Show that:*

$$\frac{d}{da}(\rho a^3) + 3pa^2 = 0, \quad (8.44)$$

and from this show that for a matter dominated universe ('dust')

$$\rho_{dust} \propto a^{-3} ,$$

and for a radiation dominated universe one has

$$\rho_{rad} \propto a^{-4} .$$

Solution

We start from $\frac{d}{da}(\rho a^3) = \frac{d\rho}{da}a^3 + 3a^2\rho = \frac{d\rho}{dt} \frac{dt}{da}a^3 + 3a^2\rho$. Taking into account that $(\dot{a}(t))^{-1} = dt/da$, since a is only a function of a single variable, t , and using (8.41) we may write the last equation as:

$$\frac{d}{da}(\rho a^3) = -3(\rho + p)a^2 + 3\rho a^2 = -3pa^2, \quad (8.45)$$

which yields the required result

$$\frac{d}{da}(\rho a^3) + 3pa^2 = 0. \quad (8.46)$$

The case of dust (which is the case today) is characterized by

$$\boxed{\text{dust (matter dominated era): } p = 0.} \quad (8.47)$$

The required result then follows directly from (8.46), since in that case the right-hand-side is zero, thereby giving

$$\rho_{dust} = \frac{\text{const}}{a^3} \quad (8.48)$$

In the case of pure radiation

$$\boxed{\text{radiation dominated era: } p = \frac{1}{3}\rho} \quad (8.49)$$

Thus from (8.46) we have

$$\frac{d\rho}{da}a^3 + 4a^2\rho = 0$$

which can be re-written as:

$$\frac{d\rho}{\rho} = -4\frac{da}{a}$$

which can be integrated straightforwardly to give the required result $\rho_{rad} = \text{const}'/a^4$.

As can be directly seen from (8.44), by using the chain rule of differentiation $\frac{d}{da(t)} = \frac{dt}{da(t)} \frac{d}{dt}$ and taking into account that $da(t)/dt = \left(dt/da(t)\right)^{-1} \neq 0$ (in fact is positive for an expanding universe), one may rewrite (8.44) as:

$$\frac{d}{dt}(\rho a^3(t)) = -p \frac{d}{dt}(a^3(t)) \quad (8.50)$$

The above result (8.50) is easily interpreted physically: the term $a^3(t)$ is proportional to the volume V of any fluid element, so the left hand side of (8.51) is the rate of change of the total energy of the fluid, while the right-hand-side is the work the fluid does as it expands $-pdV$.

Exercise 8.3 For a Robertson Walker spacetime (8.11) accept without proof that the Christoffel symbol components $\Gamma_{00}^0 = \Gamma_{j0}^0 = 0$, j a spatial index, and

$$\Gamma_{\alpha\nu}^\alpha = \left(\ln |g|^{1/2}\right)_{,\nu}$$

(where repeated indices denote summation as usual), where g is the determinant of the diagonal Robertson Walker metric (i.e. the product of its diagonal elements).

Using the equation (8.39) for a Robertson–Walker Universe (8.11) show that in the case of a perfect fluid universe (8.38), with $\rho = \rho(t)$, and $p = p(t)$, the following equation emerges:

$$\frac{d}{dt}(\rho a^3(t)) = -p \frac{d}{dt}(a^3(t)) \quad (8.51)$$

which is equivalent to (8.44), as can be directly seen from (8.44) by using the chain rule of differentiation $\frac{d}{da(t)} = \frac{dt}{da(t)} \frac{d}{dt}$ and taking into account that $da(t)/dt = \left(dt/da(t)\right)^{-1} \neq 0$ (in fact is positive for an expanding universe). This verifies that (8.39) does not yield independent information compared with Einstein's equations, as expected.

8.5 The Friedmann Equation: computing the time evolution of the scale factor in model cosmologies

We now come to an important issue, that of determining the way our Universe expands with time. To this end we need to solve Einstein's equations (8.40), for some model cosmologies in which the equation of state $p = f(\rho)$ is given, and find the time dependence of the scale factor $a(t)$. We shall do so in what follows by means of a series of instructive exercises.

Exercise 8.4 (i) Rewrite the first of (8.40) for $\Lambda = 0$ as:

$$\dot{a}^2 = \frac{8\pi G_N}{3} \rho a^2 - k \quad (8.52)$$

which is known as the Friedmann equation. Consider the asymptotic limit $a(t) \rightarrow 0$, for small t , in the case $\Lambda = 0$, in (8.52) and show, that in this case one obtains the following approximate equation for 'dust',

$$\dot{a}^2(t) \simeq (\text{const}) \times \frac{8\pi G_N}{3a(t)}.$$

From this deduce the form of $a(t)$ as a function of time, for small t , assuming an expanding universe for small t .

(ii) Under which condition for k is the expression for $a(t)$ as a function of (small) t , obtained in (i), an exact solution of Einstein's equations (8.40) for all t and for $\Lambda = 0$?

Solution

(i) Rewriting the first of equations (8.40) for $\Lambda = 0$ as:

$$\dot{a}^2 = \frac{8\pi G_N}{3} \rho a^2 - k \quad (8.53)$$

is straightforward. Equation (8.52) is known as the *Friedmann equation*, and can be used in Cosmological models in order to give us information on the temporal evolution of the scale factor $a(t)$, once ρ and p are known. For the case of dust $p = 0$ substitute the solution ρ_{dust} for 'dust' (8.48), taking the limit $a \rightarrow 0$, and keeping the dominant terms. Obviously this is the first term on the right-hand side of the above equation, thereby yielding directly the required result

$$\dot{a}^2(t) \simeq \frac{(\text{const}) \times 8\pi G_N}{3a(t)}, \quad a(t) \rightarrow 0 \quad (8.54)$$

In (8.54) we can take the square root and keep only the positive sign, since we want an expanding universe for small t where this analysis is valid by assumption. Call $\mathcal{B} = 8\pi G_N \times \text{const}$ which is a positive constant. Then one has:

$$da(t)/dt = \sqrt{\mathcal{B}/3a}^{-1/2}$$

which can be straightforwardly integrated to give:

$$a(t) \simeq \left(\sqrt{\frac{3\mathcal{B}}{4}} t \right)^{2/3}, \quad \text{small } t \quad (8.55)$$

(ii) It is obvious from the form of Friedmann's equation (8.52) that (8.55) becomes (for $\Lambda = 0$) an *exact* solution for the case of 'dust' for all t if $k = 0$.

Exercise 8.5 Repeat the analysis in part (i) of exercise 8.4 for the case of pure radiation. Show in that case that for small t the scale factor scales with the Robertson–Walker time t as:

$$a(t)_{\text{radiation}} \propto t^{1/2}, \quad (8.56)$$

and determine the proportionality constant.

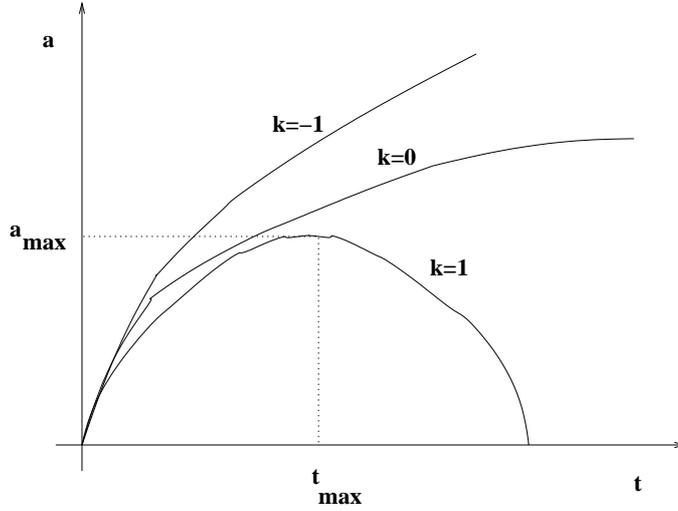


Figure 31: Qualitative behaviour of the scale factor in various types of universes in Friedmann–Robertson–Walker cosmologies.

From the previous analysis we can write the Friedmann equation (8.52) for both dust and pure radiation in the unified form:

$$\left(\frac{da(t)}{dt}\right)^2 = \frac{\kappa}{3a^b(t)} - k, \quad \kappa \equiv 8\pi G_N \quad (8.57)$$

where $b = 1$ for dust (matter dominated era) and $b = 2$ for pure radiation.

We can now qualitatively see the effects of the various types of universes (various k) that we have examined previously, in terms of the behaviour of $a(t)$.

- For small $a(t)$, we observe that $da(t)/dt \rightarrow \infty$ as $a(t) \rightarrow 0$, hence the perfect fluid universes (for all k) start from being pointlike at $t = 0$ and then there is a rapid expansion at the early times (*Big Bang*).
- For large $a(t)$ the behaviour depends only on k , since as we can see from (8.57) the a -dependent terms on the right-hand-side become negligible.

Let us now examine the fate of the universe for the various types of k and compare the results with our previous classification of universes based on geometry.

- For $k = 1 > 0$, as can be seen from (8.57) there is a point in time $t = t_{\max}$ at which $a(t_{\max})$ becomes maximum ($da(t_{\max})/dt = 0$) (cf figure 31). This conforms our earlier classification that this universe is closed and will eventually recontract.
- For $k = 0$, we observe from (8.57) that as $t \rightarrow \infty$, i.e. $a \rightarrow \infty$, $da(t)/dt \rightarrow 0$. This is the flat universe as we have seen previously (cf. figure (31)).
- Finally, for $k = -1$, we observe that as $t, a(t) \rightarrow \infty$ then $da(t)/dt \rightarrow 1$. This is the open universe according to our previous classification (c.f. figure 31).

We plot qualitatively these three cases in figure 31.

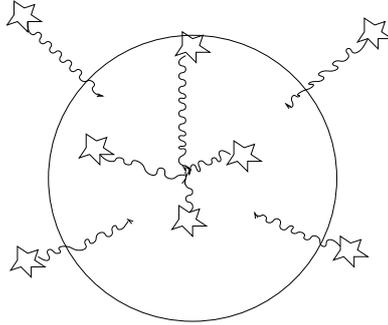


Figure 32: Schematic representation of the fact that the Sky is Dark at night in the Big Bang Universe.

8.6 The Big-Bang model in modern Cosmology: why is the sky dark at night?

As we see from figure 31, the scale factor of the Universe may cross zero at early times. This is, in fact, what happens in the modern theory of *Big Bang*, according to which our Universe started from a big explosion, at the ‘beginning of time’, where the Universe was point like in space (initial ‘singularity’). According to the theory of Big Bang, the Universe

- (a) has finite age, because it started with an initial big explosion, i.e. a cosmically catastrophic event, before which we have no idea how to describe space and time. In the context of a Robertson-Walker Universe, such an era would correspond to early times for which the scale factor $a(t) \rightarrow 0$.
- (b) there is a cosmological expansion which causes the cosmological redshift in radiation, i.e. the energies of the photons received from stars are smaller than the corresponding emission energies (energy = $h\nu$, h =Planck constant, ν frequency, and the redshift implies: $\nu a(t)=\text{constant}$, $a(t)$ a scale factor, increasing with cosmic time in the Big Bang theory of expanding Universe).

The most important feature of the Big Bang is therefore the fact that, as a consequence of (a) (finite age), as well as of the finite speed of light, there is a cosmic event horizon, due to which the light from objects beyond the horizon did not have the time to reach us at present (see figure 32).

Also, due to (b), the luminosity of the night sky is also diminished significantly. In particular it can be shown that photons emitted from stars on the horizon arrive to Earth with vanishing energy (infinite wavelength), hence unobserved.

In fact it can be shown mathematically that if the Universe age is 10^{10} years (which is the order of magnitude that recent observations have indicated), then the distance by which one should extend a random line of vision until it reaches the surface of a star, is much larger than the cosmic horizon radius (see figure 32). Hence the night sky appears mostly dark.

It must be noted at this stage that in the Steady State Theory, which preceded the Big bang Theory, according to which the Universe was eternal, the luminosity of the night sky would be that of the sky during day time when one looks towards the direction of our Sun, the latter assumed a middle range star (as far as luminosity is concerned).

Therefore we can safely say that *the fact that the night sky is dark is a pretty good evidence for the Big bang Theory of Modern Cosmology.*

8.7 The Critical Density of the Universe and Cosmological Observations

We discuss at this point an important quantity in Cosmology, which allows direct connection with observations. Let us differentiate equation (8.57) with respect to the time t ,

$$2 \frac{da(t)}{dt} \frac{d^2a(t)}{dt^2} = -b \frac{8\pi G_N}{3} \frac{1}{a^{b+1}} \frac{da(t)}{dt} \quad (8.58)$$

Re-expressing it in terms of the density $\rho = 1/a(t)^{b+2}$ and solving for ρ one obtains:

$$\rho = -\frac{3}{4\pi G_N b} \left(\frac{d^2a(t)/dt^2}{a(t)} \right) \quad (8.59)$$

We now define the *deceleration parameter* q as:

$$q \equiv -\frac{a(d^2a/dt^2)}{(da/dt)^2} = -\frac{\ddot{a}}{aH}, \quad (8.60)$$

in terms of which (8.59) becomes

$$\rho = \frac{3}{4\pi G_N b} q \left(\frac{da/dt}{a} \right)^2 \quad (8.61)$$

The Hubble parameter $\frac{da/dt}{a} \equiv H$ (8.25) is an observable, as can be measured by means of the Cosmological Redshift, as we have seen previously.

The *critical density* ρ_c is defined as the density the Universe should have in order to be spatially flat, i.e. $k = 0$, which, on account of (8.57), can be expressed in terms of $\frac{da/dt}{a}$:

$$\left(\frac{da/dt}{a} \right)^2 = \frac{8\pi G_N}{3} \rho_c \quad (8.62)$$

from which, by means of (8.61),

$$\rho = (2q/b)\rho_c.$$

If in the current era the Universe were matter dominated, i.e. $b = 1$, as was the belief up to 1998, then, in that case one may define the ratio Ω

$$\boxed{\Omega \equiv \frac{\rho}{\rho_c} = 2q} \quad (8.63)$$

Thus, by measuring q , then one could determine, in the context of the Friedmann model, whether a matter-dominated Universe was open or closed. Indeed, an open Universe corresponds to $\Omega < 1$, whilst a closed Universe has $\Omega > 1$.

At present there is good experimental evidence, from a plethora of cosmological observations that we shall discuss briefly in subsequent sections, that the simple Friedmann model works very well, but only upon the inclusion of a cosmological constant contribution to its energy budget, in addition to the matter contributions today. Moreover, the observations point out that the Universe is very nearly critical $\Omega \simeq 1$ (spatially flat $k = 0$), but the issue as to whether it is open or closed is still unsettled, due to experimental limitations, of course. In such a case, the deceleration parameter q (8.60) is not related simply to Ω for matter, as in (8.63), but it involves the contributions from the cosmological constant, which tend to make the deceleration parameter negative, and thus to *accelerate* the Universe. In the next subsections we discuss the properties of a Universe with a non zero (positive) cosmological constant, and we give a sketchy overview of the relevant up to date cosmological measurements of the Universe's energy budget.

8.8 The Age of a Big-Bang Universe: a model dependent quantity

Assuming a Big-Bang Universe, one may compute the resulting *finite Age* by using the Friedmann equation, i.e. the first of the equations (8.40). As we shall demonstrate below, the age of the Universe is a quantity that depends highly on the underlying cosmological model. For our purposes in this subsection we retain all the terms in the Friedmann equation, including the cosmological constant Λ .

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G_N}{3}\rho - \frac{k}{a^2} + \frac{\Lambda}{3} \quad (8.64)$$

Solving for $\dot{a} > 0$ (expanding universe), we obtain:

$$\dot{a} = +\sqrt{\frac{8\pi G_N}{3}\rho - \frac{k}{a^2} + \frac{\Lambda}{3}}.$$

From this, we obtain the age t_{Age} of a Big-Bang Universe, by integrating the above relation from the singularity $a(t=0) = 0$ till present value a_0 , i.e.:

$$\int_{a=0}^{a_0} \frac{da}{a\sqrt{\frac{8\pi G_N}{3}\rho - \frac{k}{a^2} + \frac{\Lambda}{3}}} = \int_0^{t_{\text{Age}}} dt = t_{\text{Age}} \quad (8.65)$$

This formula clearly demonstrates the model dependence, since the details of the Hubble parameter in terms of the various energy density components are model dependent (above we used the simple Friedmann-Robertson-Walker model. In more complicated models the dependence of the various energy density components on the scale factor is more involved).

As an illustration, consider the simple example of a spatially flat dust(matter)-dominated Universe, with $\Lambda = 0$. In such a case, from (8.48) and (8.65) we obtain:

$$t_{\text{Age-dust}} = \int_0^{a_0} \frac{da}{a\sqrt{\frac{8\pi G_N}{3}a^{-3}}} = \frac{2}{3} \left(\frac{3}{8\pi G_N}\right)^{1/2} a_0^{2/3} \quad (8.66)$$

From (8.64), evaluating at the present era, we obtain:

$$\left(\frac{\dot{a}}{a}\right)_0^2 \equiv H_0^2 = \frac{8\pi G_N}{3} \frac{1}{a_0^3} \quad (8.67)$$

where H_0 is the value of the Hubble parameter today. From (8.67) and (8.66), then, we obtain for the Age of a matter-dominated Universe:

$$t_{\text{Age-dust}} = \frac{2}{3} H_0^{-1} \quad (8.68)$$

This simple exercise indicates that the so-called “*Hubble time*” H_0^{-1} sets the scale for the age of a Friedmann-Robertson-Walker Universe. In a more realistic Universe, the above integration is more complicated, since one has to take into account the various epochs, radiation dominance at an early era,, succeeded by an era of matter-radiation equality, followed by a matter dominated era, and finally, according to present observation, an era where the cosmological constant (or more generally dark energy) contribution begins to dominate. Measurements of the present-day Hubble parameter, then, are crucial for calculating (within well defined theoretical models) the age of the Big-Bang Universe.

A cautionary remark is in place at this point: It must be noted that the *inflationary era* of the Universe, i.e. an early phase characterised by an exponential expansion of the scale factor, which is described by a de Sitter metric (c.f. section 9.3 below) is still *terra incognita*, and thus the actual Age of an inflationary Universe cannot be computed. It depends crucially on the details of the Early universe model, in particular the precise form of the inflationary potential. In such scenarios what one calls age is calculated after the end of the inflationary era. An estimate of the actual Age of an inflationary Universe cannot be made, unless knowledge of the inflationary potential is given. Flat directions in the latter result in the Universe being much older than one calculates usually.

8.9 Basic Thermodynamics of Robertson-Walker-Friedmann Universe

It is straightforward to see that the expansion of the Friedmann-Robertson-Walker (FRW) Universe is consistent with the basic concepts and laws of Thermodynamics (c.f. Appendix B for a more detailed review of the relevant formulae and concepts). Indeed, consider first the radiation era of the Universe. In that epoch, the energy density of radiation scales as we have seen with the fourth inverse power of the scale factor

$$\rho_{\text{rad}} \propto a^{-4} \quad (8.69)$$

From the cosmic redshift relation (8.37), we then have that

$$a \propto \lambda \quad (8.70)$$

where λ the wavelength of radiation.

From Stefan-Boltzman law (c.f. Appendix B), on the other hand, the density of radiation scales with the fourth power of the temperature T , assuming a heat bath:

$$\rho_{\text{rad}} = \sigma T^4, \quad \text{Stefan - Boltzmann law, } \sigma = \text{radiation constant} \quad (8.71)$$

From (8.71) and (8.69) we then have:

$$a \propto T^{-1} \quad (8.72)$$

which describes the cooling law of the expanding FRW Universe. This behaviour is then consistent with the Wien's law, according to which the maximum of the thermal radiation spectrum has a wavelength λ_{max} which changes with the temperature T_{rad} of radiation according to: $\lambda_{\text{max}} T_{\text{rad}} = \text{constant}$. This follows immediately from (8.72) and the red-shift relation (8.70).

From the time dependence of the scale factor on the cosmic time (8.56), $a \propto t^{1/2}$, the cooling law (8.72) yields for the radiation era of the Universe:

$$t \propto T^{-2} \quad (8.73)$$

It can be shown [1] that during the radiation era, the proportionality coefficient in (8.73) is: $0.301 g_{\star}^{-1/2} M_{\text{Pl}}$, where g_{\star} counts the total number of effectively massless degrees of freedom (those species with masses $m_i \ll T$. These will dominate the radiation-era energy density and pressure, given that the contributions from non-relativistic species will be suppressed by exponential terms of the form $e^{-m_i/T}$, which are negligible if $m_i \gg T$). Taking into account the difference between Bose and Fermi statistics, one has (c.f. Appendix B):

$$g_{\star} = \sum_{i=\text{Bosons}} g_i \left(\frac{T_i}{T}\right)^4 + \frac{7}{8} \sum_{j=\text{fermions}} g_j \left(\frac{T_j}{T}\right)^4 \quad (8.74)$$

and we have for those thermalised relativistic species in the radiation era (at thermal equilibrium):

$$\rho = \frac{\pi}{30} g_{\star} T^4 \quad (8.75)$$

During the expansion of the Universe, the total entropy inside the proper co-moving volume a^3 of the FRW Universe remains constant, as can be immediately deduced from Einstein's equations, in particular (8.50). The latter admits a thermodynamics interpretation, as we have already discussed in previous sections, namely it is consistent with the *second law* of thermodynamics, but with *constant entropy* enclosed in the proper volume a^3 :

$$d(\rho a^3) + p d(a^3) = 0 = dE_{\text{total}} + p dV = T dS \quad (8.76)$$

where $E_{\text{total}} = \rho a^3$ denotes the total energy of the cosmic fluid included in the co-moving (proper) volume a^3 , with energy density ρ and pressure p , and S is the total entropy of the volume a^3 , which is thus constant.

One may determine the entropy density per co-moving volume in the Einstein Universe, as follows. Ignoring for a moment Einstein's equations, which imply $dS = 0$, we first re-write (8.76) as:

$$dS = \frac{1}{T}[d((\rho + p)V) - Vdp] = \frac{V}{T} \frac{d\rho}{dT} dT + \frac{(\rho + p)}{T} dV \quad (8.77)$$

and make use of the integrability conditions

$$\frac{\partial^2 S}{\partial V \partial T} = \frac{\partial^2 S}{\partial T \partial V} \quad (8.78)$$

which on account of (8.77) and the fact that $\rho = \rho(T)$, $p = p(T)$, implies:

$$\frac{dp}{dT} = \frac{\rho + p}{T} \quad (8.79)$$

Substituting (8.79) into (8.77) we then have:

$$dS = \frac{V}{T} d(\rho + p) + \frac{\rho + p}{T} dV - \frac{V(\rho + p)}{T^2} dT = d\left(\frac{V[\rho + p]}{T}\right) \quad (8.80)$$

implying that the entropy inside the proper volume is constant,

$$S = V(\rho + p)/T = \text{constant} \quad (8.81)$$

with the entropy density

$$s \equiv \frac{S}{V} = \frac{\rho + p}{T} \quad (8.82)$$

We shall make use of these relations later on, when we discuss the calculation of the thermal relic abundances in an expanding Universe.

A final remark is in order, before closing this subsection. The entropy density is dominated by the contribution of relativistic species, so that (c.f. Appendix B, Eq. (12.22) and relevant discussion) [1]

$$s = \frac{2\pi^2}{45} g_{*S} T^3 \quad (8.83)$$

where $g_{*S} = \sum_{i=\text{Bosons}} g_i \left(\frac{T_i}{T}\right)^3 + \frac{7}{8} \sum_{j=\text{fermions}} g_j \left(\frac{T_j}{T}\right)^3$.

9 Including a Cosmological Constant in Robertson-Walker-Friedmann Cosmology

In this section we shall include a cosmological constant $\Lambda > 0$ into the Friedmann equation, which is the case suggested by a plethora of measurements over the last decade, and see how the underlying physics is modified, as compared to the case with $\Lambda = 0$.

9.1 Historical remark: Einstein's reasoning for introducing the Cosmological Constant Λ

Let us first, as a historical remark, discuss what was the problem that Einstein wanted to solve by introducing the cosmological constant term Λ into the theory. Suppose that $\Lambda = 0$. Then from (8.42) we have

$$\frac{\ddot{a}}{a} = -\frac{4\pi G_N}{3}(\rho + 3p) .$$

So, if we want the universe to have non-zero matter we must require $p > 0$, $\rho > 0$, which implies that $\ddot{a} < 0$. This is the problem that bothered Einstein, who wanted to have a *static* universe (i.e. $\dot{a} = 0$) with matter in it, because this was the common belief at the time. From (8.42), this obviously requires the introduction of a *positive* cosmological constant term $\Lambda = 4\pi G_N(\rho + 3p) > 0$.

Subsequent observations about an *expanding* universe essentially eliminated the above reasoning for the necessity of introducing the cosmological constant, and made Einstein characterize his whole reasoning about its introduction “the biggest blunder of his life”.

At present there is experimental (observational) evidence that the cosmological constant, if exists, it should be very small, in Planck units of order $\Lambda/M_P^4 < 10^{-120}$ (Planck energy scale $M_P \sim 10^{19}$ GeV). This is an extremely small number, and hence most theorists believe that there must be some *symmetry* that prevents the appearance of such a constant. This issue is still unresolved, and is one of the biggest and most challenging issues in theoretical cosmology to date.

9.2 Overview of Recent Experimental Evidence for a Cosmological Constant

Observations of light from high-redshift supernovæ ($z \sim 1$) appear to provide ‘evidence’ for a non-zero cosmological constant consistent with the above bound. The evidence initially came from observations by two teams, and the interested reader can find details in: A. G. Riess *et al.*, *Astroph. J.* **117**, 707 (1999), and S. Perlmutter *et al.*, *Astroph. J.* **517**, 565 (1999).

The evidence from supernovae data may be summarized as follows: looking at the apparent brightness of distant (redshift $z \sim 1$) supernovae Ia one finds that such objects appear *dimer* that they should be, if the Universe at such early times was expanding with the same rate as it does now. This evidence is, of course, currently not confirmed, mainly because the nuclear physics of the mechanisms of production of supernovae Ia is not quite understood, as yet, so as to allow for detailed quantitative models to be developed, and this is the main reason for uncertainty at present. If, however, one accepts that the observed anomalies (see figure 33) in the apparent magnitude of the supernovae with the redshift z are due to different rates of expansion of the Universe then and now, then it becomes obvious that such observations point towards the conclusion that currently our Universe is *accelerating*.

These data can be compiled with other data implying, as mentioned previously, that our Universe is flat $k = 0$ (within the current experimental accuracy). The fit of the experimental data from supernovae Ia, then, to a spatially flat $k = 0$ Friedmann-Robertson-Walker model, with a cosmological constant Λ leads to the following: $\Omega_{M,0} \simeq 0.3$, $\Omega_{\Lambda,0} \simeq 0.7$, where the suffix M stands for matter contributions, and Λ for cosmological constant contributions. It is remarkable that the data imply that 70% of the energy our Universe is due to a Cosmological Constant, if one believes the above fit. Fitting the data results in a present-era *acceleration* of the Universe, quantified by the deceleration parameter $q_0 = -0.55 < 0$ (the subscript 0 indicate present values).

The best confirmation for the above-mentioned energy budget of the Universe, in which the vacuum (“dark energy”) contribution, of unknown microscopic origin, dominates the current era, came from a set of independent measurements of the temperature fluctuations of the cosmic microwave background, by means of the Wilkinson Microwave Anisotropy Probe (WMAP-satellite). Up to date, after six years of running, the WMAP satellite provides very accurate measurements (c.f. figure 34 of the temperature fluctuations, and through those, one can fit the predictions of a specific cosmological model. The best fit model in *all* currently available, including data from large galactic surveys, is provided by the standard Cosmology of Friedmann-Robertson-Walker with a Cosmological constant, which today occupies 74% of its energy budget (c.f. figure 35), whilst 26% consists of matter (4% ordinary observable matter and 22% ‘dark’ matter, whose nature is also a mystery).

In fact the observations are compatible even with models in which the cosmological “constant” is *not constant*, but is relaxing to zero (quintessence models), e.g. as $1/t^2$ at late eras, where t is the Robertson–Walker time (8.11). If one sets t of the order of the age of the Universe today, i.e. $t \sim 10^{60}$ in Planckian units (1 Planck time $t_p \sim 10^{-43}$ s), then such relaxation models, if they turn

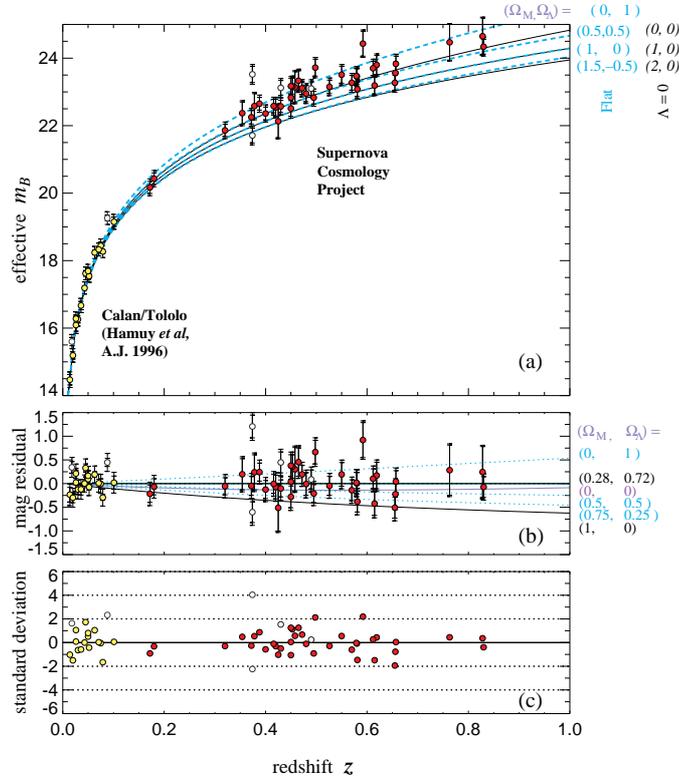


Figure 33: Initial Experimental evidence for the present-era acceleration of the Universe (and hence for a non-zero positive cosmological constant). Distant supernovae Ia seem dimmer than they should be if the Universe did not accelerate (picture from S. Perlmutter *et al.*, *Astroph. J.* **517**, 565 (1999)). Similar results exist for the second group A. G. Riess *et al.*, *Astroph. J.* **117**, 707 (1999)).

out to be correct, they can provide an explanation of the small value of the cosmological constant today by starting from Planckian values M_P^4 of Λ in the very Early Universe, the latter being the natural scale of quantum corrections. In general, such contributions could be contributions from energy of the ‘vacuum’, or in the latter case of a time dependent constant, from the potential of a *quintessence field* which has not yet relaxed to its equilibrium state. It is in this sense that people are talking about recent evidence of a ‘dark energy’ component of the Universe rather than simply a cosmological constant (which is more than 70% of the total energy).

As the Universe evolves, i.e. the scale factor a increases, the matter contributions to the energy density, which scale like a^{-3} (8.48), will become *subdominant* compared with the cosmological constant (or ‘vacuum energy’ or ‘dark energy’) contributions, which are supposed to remain constant, independent of a . Thus, eventually the Universe will be dominated by the Λ -term alone. It will be interesting to study the properties of such a Universe. We do so briefly in the next subsection.

9.3 Properties of a Universe with a Cosmological Constant term and no matter: de Sitter Universe

To understand better the properties of a Universe in which the cosmological constant is a dominant factor in Einstein’s equations (6.5), with $\kappa = 8\pi G_N$, and in which the matter contributions are

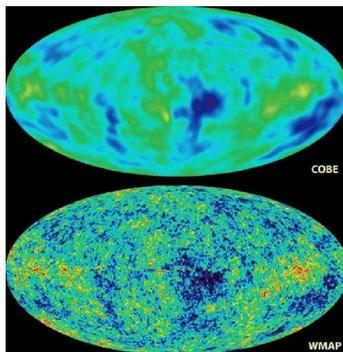


Figure 34: Upper Figure: temperature fluctuations (anisotropies) in the Cosmic Microwave Background Radiation (CMB) as measured by the COBE experiment (Nobel prize for Physics 2006); Lower figure: the significant improvement (already after one year of running, 2003) in measuring CMB temperature fluctuations by the WMAP satellite.

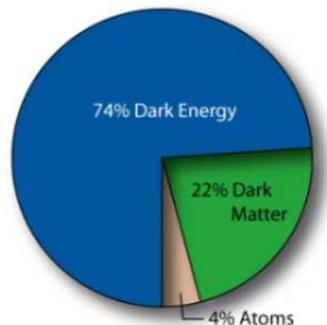


Figure 35: The energy content of our Universe as obtained by fitting data of WMAP satellite. The chart is in perfect agreement with earlier claims made by direct measurements of a current era acceleration of the Universe from distant supernovae type Ia (courtesy of <http://map.gsfc.nasa.gov/>).

negligible by comparison, we observe that in such a case the equations acquire the form

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = -\Lambda g_{\mu\nu}, \quad \Lambda > 0 \quad (9.1)$$

One may now interpret the right-hand-side as a ‘vacuum’ contribution to the stress energy tensor, which will then be of the form:

$$T_{\mu\nu}^{\Lambda\text{-vacuum}} \simeq -\frac{1}{8\pi G_N}g_{\mu\nu}\Lambda \quad (9.2)$$

In the above formula $g_{\mu\nu}$ is the *diagonal* Robertson-Walker metric (8.11). This Universe is called *de Sitter*, in honour of W. de Sitter, who constructed this Universe mathematically in 1917 (W. de Sitter, “On the curvature of space” *Proc. Kon. Ned. Akad. Wet.* 20, 29 (1917)).

Comparing then (9.2) with (6.2), we then observe that the Λ -dominated Universe, as the de Sitter Universe is called alternatively, is a perfect fluid, because its stress tensor is *diagonal*. Taking into account that the Robertson-Walker frame is a comoving frame, as mentioned above, we observe that in this mcrf one has an *equation of state*, i.e. a relation $p = f(\rho)$ between the pressure p_Λ and the energy density ρ_Λ of the Λ -dominated Universe of the form:

$$p_\Lambda = -\rho_\Lambda = -\frac{1}{8\pi G_N}\Lambda \quad (9.3)$$

where $\rho_\Lambda \equiv T_{00}^{\Lambda\text{-vacuum}}$ and $p_\Lambda = T_{ii}^{\Lambda\text{-vacuum}}/g_{ii}$ as usual for an ideal fluid in a mcrf (6.2), with g_{ii} the spatial components of the Robertson-Walker metric (8.11).

From the first of the Einstein's equations (8.40) in that case, i.e. setting $\rho \rightarrow 0$ we observe that, for large enough times t , where the Λ -dominated Universe is expected to occur, the k term is negligible (notice that the following results are exact for flat universes $k = 0$, which is the recent experimental evidence, as mentioned previously ⁵):

$$\frac{\dot{a}}{a} = +\sqrt{\frac{\Lambda}{3}}, \quad \Lambda > 0, \quad t \rightarrow \infty \quad (9.5)$$

where the positive root has been taken due to the expanding nature of the Universe. This implies an exponentially expanding Universe for large times, i.e. a scale factor

$$a(t) = a_0 e^{\sqrt{\frac{\Lambda}{3}}t}, \quad \Lambda > 0 \quad (9.6)$$

which, notably, is of the same nature as that in the so-called *inflationary period*. Indeed, the inflationary period of the Universe, is also a period in the very early Universe's history, where the Universe expands exponentially. We do not have the time, neither the student the expertise, to analyze this phase in these notes. This is the topic of a graduate course in Cosmology.

A de-Sitter metric is *maximally symmetric*, in the sense that, if one computes the four-dimensional Riemann tensor for such Universes (c.f. (8.18) using (9.6), and then computing the curvature tensor), then (s)he finds:

$$R_{\mu\nu\rho\sigma}^{\text{dS}} = \frac{\Lambda}{3} (g_{\mu\rho}g_{\nu\sigma} - g_{\nu\rho}g_{\mu\sigma}) \quad (9.7)$$

Thus, we are dealing with a space-time of constant curvature (positive if $\Lambda > 0$), in which no space or time directions are singled out. Moreover, if one computes the spatial three-curvature tensor for the de-Sitter metric (9.6), it becomes clear that in the three cases of a FRW Universe, $k = \pm 1, 0$, the corresponding metrics (9.4) correspond to three different sections of the *same* four-dimensional space with constant positive curvature. The reader should notice that the de-Sitter space is *not asymptotic* to Minkowski space (in the asymptotic time limits $t \rightarrow \pm\infty$ the metric diverges, as can be seen from (9.6)).

The important point to notice is that such de-Sitter Universes, with $\Lambda > 0$ a time-independent constant, are *eternally accelerating*, as can be directly seen by computing $\ddot{a} \propto a \rightarrow +\infty$ as $t \rightarrow \infty$. In such Universes there is a *cosmic horizon*, in other words there is a *maximum* distance beyond which the cosmological observers *cannot see*. In general, the question whether or not there is a cosmic horizon is determined by the finiteness of the quantity (in units of $c = 1$ we work here):

$$\delta = a(t) \int_{t_0}^{\infty} \frac{dt'}{a(t')} \quad (9.8)$$

where t_0 is the (present) time moment at which a light signal has been sent out in a FRW Universe. The question is whether the light signal reaches all points of the Universe before its end, which we assume here occurs at $t_{\text{end}} = \infty$. The finiteness of the integral (9.8) implies the existence of an *event (cosmic) horizon*, since in that case we shall never be able to learn anything about events situated at distances larger than δ , whilst in cases where the integral diverges the horizon is

⁵For finite times, the solution to the equations for the three cases of k are:

$$\begin{aligned} a(t) &= \frac{1}{b} \cosh(bt), & k &= +1 \\ a(t) &= \frac{1}{b} \sinh(bt), & k &= -1 \\ a(t) &= a_0 e^{bt}, & k &= 0, a_0 = \text{constant} \end{aligned} \quad (9.4)$$

where $b^2 = \Lambda/3$.

absent. As we have just seen de Sitter Universes are characterized asymptotically in cosmological frame time $t \rightarrow \infty$ by $a(t) = e^{bt}$, and hence for such scale factors the integral in (9.8) converges, and thus there is a cosmic horizon. This presents serious problems in formulating a consistent quantum theory in such Universes.

It must be noted, though, that in certain quintessence models, where the cosmological constant is time dependent, and in fact relaxes to zero (from Planckian values at early times), the possibility of exiting from a de Sitter phase occurs, and the horizon in such cases disappears. On the theoretical side, these issues are unsettled at present, especially because we do not have control of the Physics at such Planckian energy scales. On the experimental side, it must also be said that at present such observational evidence for a non-zero cosmological constant must be treated with caution, given the large experimental uncertainties of the observations.

9.4 Astrophysical Measurements of the Universe Energy Budget: some details

As we have already mentioned, on Large scales our Universe looks isotropic and homogeneous. A good formal description, which does not depend on the detailed underlying microscopic model, is provided by the Robertson-Walker (RW) metric, according to which the geometry of the Universe is described by means of the space-time invariant element (8.11), which we repeat here for the convenience of the reader:

$$ds^2 = -dt^2 + a(t)^2 R_0^2 \left[\frac{dr^2}{1 - k r^2} + r^2 (d\theta^2 + \sin^2\theta d\varphi^2) \right] \quad (9.9)$$

where $a(t) = \frac{R(t)}{R_0} = \frac{1}{1+z}$ is the scale factor, $H \equiv \frac{\dot{a}}{a}$ is the Hubble Parameter, t is the Cosmological Observer time, R_0 denotes the present-day scale factor, $z =$ is the redshift, related to the scale factor by the cosmic redshift relation (8.37), and k denotes the *Spatial Curvature*, which (by normalization) can take on the values: $k=0$ for a flat Universe (required by inflationary models), $k=1$ for a closed and $k=-1$ for an open Universe. In this section we shall outline the main Cosmological Measurements and the pertinent quantities, of interest to us in these Lectures. For more details we refer the reader to the literature [1, 2, 3].

9.4.1 Model Independent (Geometric) Considerations

An important quantity, which we shall make extensive use of in the following, when we use astrophysical data to constrain theoretical models, is the so-called *Luminosity Distance*, d_L , defined as:

$$d_L = \sqrt{\frac{\mathcal{L}}{4\pi\mathcal{F}}} , \quad (9.10)$$

where \mathcal{L} is the energy per unit time emitted by the source, at the source's rest frame, and \mathcal{F} is the flux measured by detector, i.e. the energy per unit time per unit area measured by the detector.

To determine the effects of the expansion of the Universe on the flux \mathcal{F} measured by the detector, we should take into account that a cosmic observer is co-moving with (i.e. is static with respect to) the expanding universe. Consider now a photon wave being emitted at time t_1 from a source at a coordinate $r = r_1$, and being observed at a detector located at $r = 0$. Due to the null-ness of the photon geodesics paths ($ds^2 = 0$), we have:

$$f(r_1) \equiv \int_0^{r_1} \frac{dr}{\sqrt{1 - kr}} = \int_{t_0}^{t_1} \frac{dt}{a(t)} . \quad (9.11)$$

Consider now a second wavecrest emitted at a time $t_1 + \delta t_1$, with δt_1 infinitesimally small. It will arrive at the detector at time $t_0 + \delta t_0$. The equation of motion for the photon will now be (9.11)

with $t_1 \rightarrow t_1 + \delta t_1$ and $t_0 \rightarrow t_0 + \delta t_0$, while $f(r_1)$ will be the same, because the *source is fixed* for a co-moving observer:

$$\int_{t_1}^{t_0} \frac{dt}{a(t)} = \int_{t_1 + \delta t_1}^{t_0 + \delta t_0} \frac{dt}{a(t)} \Rightarrow \int_{t_1}^{t_1 + \delta t_1} \frac{dt}{a(t)} = \int_{t_0}^{t_0 + \delta t_0} \frac{dt}{a(t)},$$

after rearranging appropriately the integration limits. For small δt 's one can assume the scale factors as approximately constant and take them out of the time integrations. That is to say, two events separated via time interval $\delta t_1 \ll 1$, when the universe had scale factor $a(t_1)$, will be separated by a time interval $\delta t_0 \ll 1$ at the observation point, when the Universe has scale factor $a(t_0) > a(t)$ for $t_0 > t_1$, given by:

$$\frac{\delta t_0}{a(t_0)} = \frac{\delta t_1}{a(t_1)}.$$

This implies that the *time dilation* induced by the *expansion* of the Universe is given by:

$$\delta t_{\text{detector}} = (\delta t)_{\text{source}}(1 + z). \quad (9.12)$$

The above relations also are responsible for the cosmic red-shift (8.37, which implies a reduced energy of photons at the detector as compared with that at emission from the source. *Both* effects, namely time dilation (9.12) and cosmological red-shift (8.37), then imply that for the flux \mathcal{F} measured by a detector one obtains:

$$\mathcal{F} = \frac{\mathcal{L}}{4\pi a(t_0)^2 r_1^2 (1 + z)^2}. \quad (9.13)$$

where we took into account energy conservation (implied by (6.38)), as well as the fact that at the detection time t_0 , due to the scale-factor effects, the fraction of the area of a two-sphere surrounding the source covered by the detector is $dA/4\pi a(t_0)^2 r_1^2$, with dA the area of the detector.

From this we obtain, on account of (9.10):

$$d_L^2 = a(t_0)^2 r_1^2 (1 + z)^2 \quad (9.14)$$

Another commonly used quantity in Astrophysics is the *Angular Diameter*, which is defined as follows: A celestial object (cluster of galaxies etc.) has proper diameter D at $r = r_1$ and emits light at $t = t_1$. The observed angular diameter by a detector at $t = t_0$ is:

$$\delta = \frac{D}{a(t_1) r_1}. \quad (9.15)$$

From this one defines the *Angular Diameter Distance*:

$$d_A = \frac{D}{\delta} = a(t_1) r_1 = d_L (1 + z)^2. \quad (9.16)$$

where in the last equality we used (9.14) and the cosmological redshift relation (8.37).

A final quantity which we would like to define is the *Horizon Distance*, d_H , beyond which light cannot reach us. This is calculated as follows: as already mentioned, for radial motion of light, pertinent to most observations, along null geodesics, $ds^2 = 0 = dt^2 - a^2(t) \left(\frac{dr^2}{1 - kr^2} \right)$, we have:

$\int_0^t \frac{dt'}{a(t')} = \int_0^{r_H} \frac{dr}{\sqrt{1 - kr^2}}$ from which

$$d_H = a(t) \int_0^{r_H} \sqrt{g_{rr}} = a(t) \int_0^t \frac{dt'}{a(t')}. \quad (9.17)$$

If d_H is finite, then our past light cone is limited by an Horizon, which acts as the boundary between the visible Universe and its part from which light has not reached us. The finiteness or not of d_H is determined mainly by the behaviour of the scale factor of the cosmological model

under consideration near the initial singularity. In Standard Big-Bang Cosmology $d_H \sim t_{\text{Age}} < \infty$ due to the *finite age* of the Big-Bang Universe, i.e. there is an Horizon.

The above quantities are related among themselves [1], as follows from the cosmic redshift phenomenon, the fact that photons follow null geodesics $ds^2 = 0$ etc. These leads to relations among H_0, d_L and the redshift z , some of which are model independent and follow from pure geometrical considerations, relying on the assumption of a RW homogeneous and isotropic cosmology.

For instance, by considering nearby measurements, i.e. small red-shifts, we can implement a Taylor expansion in the Hubble parameter with respect to the cosmic time (present day quantities are denoted, as usual, by the suffix 0):

$$\frac{a(t)}{a(t_0)} = 1 + H_0(t - t_0) - \frac{1}{2}q_0 H_0^2 (t - t_0)^2 + \dots, \quad H_0 \equiv \frac{\dot{a}(t_0)}{a(t_0)}, \quad (9.18)$$

where q_0 denotes the present-day value of the Universe deceleration parameter (8.60).

From the cosmic redshift relation (8.37) we obtain:

$$\begin{aligned} z &= -H_0(t - t_0) + \left(1 + \frac{q_0}{2}\right)H_0^2(t_0 - t)^2 + \dots \implies \\ t_0 - t &= H_0^{-1} \left(z - \left(1 + \frac{q_0}{2}\right)z^2 + \dots \right) \end{aligned} \quad (9.19)$$

To find a relation H_0, d_L and z we consider photon emitted at $t = t_1, r = r_1$ and received at $t = t_0, r = 0$ in a Robertson-Walker Universe with parameter k . Since the photon follows null geodesics, we have by integrating from emission till observation points:

$$\int_{t_1}^{t_0} dt/a(t) = \int_0^{r_1} dr/\sqrt{1 - kr^2} \equiv f(r_1),$$

where

$$f(r_1) = \begin{cases} \sin^{-1}r_1 \simeq r_1 + \frac{r_1^3}{6} + \dots, & k = +1 \text{ (closed)} \\ r_1, & k = 0 \text{ (flat)} \\ \sinh^{-1}r_1 \simeq r_1 - \frac{r_1^3}{6} + \dots, & k = -1 \text{ (open)} \end{cases}.$$

From this we obtain:

$$r_1 \simeq a^{-1}(t_0) \left((t_0 - t_1) + \frac{1}{2}H_0(t_0 - t_1)^2 + \dots \right) \simeq a^{-1}(t_0)H_0^{-1} \left(z - \frac{1}{2}(1 + q_0)z^2 + \dots \right),$$

from which the required relation between the Hubble parameter (today) and the luminosity distance is derived, upon using (9.14):

$$H_0 d_L = z + \frac{1}{2}(1 - q_0)z^2 + \dots, \quad (9.20)$$

which is essentially Hubble's law (8.1).

It should be stressed once again that the above relations (which are valid for small red-shifts) are model independent, and they follow from pure geometrical considerations (assuming of course the Cosmological principle, that is a Friedmann-Robertson-Walker Universe. But no information on the precise energy budget and the form of the various energy-density contributions to the vacuum of the Universe is required. These details, however, are required when one considers much higher red-shifts, z , pertaining to the very early epochs of the Universe, where the exact relations are necessary.

In the next subsection we discuss how a specific dynamical model of the Universe affects the cosmological measurements. In particular, as we shall show, model dependence is hidden inside the details of the dependence of the Hubble parameter H on the various components of the Universe's energy budget. This property is a consequence of the pertinent dynamical equations of motion of the gravitational field.

9.4.2 Cosmological Measurements: Model Dependence

Within the standard General-Relativistic framework, according to which the dynamics of the gravitational field is described by the Einstein-Hilbert action, the gravitational (Einstein) equations in a Universe with cosmological constant Λ read: $R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + g_{\mu\nu}\Lambda = 8\pi G_N T_{\mu\nu}$, where G_N is (the four-dimensional) Newton's constant, $T_{00} = \rho$ is the energy density of matter, and $T_{ii} = a^2(t)p$ with $p =$ the pressure, and we assumed that the Universe and matter systems behave like ideal fluids in a co-moving cosmological frame, where all cosmological measurements are assumed to take place. From the RW metric (9.9), we arrive at the Friedman equation:

$$3 \left(\frac{\dot{a}}{a} \right)^2 + 3 \frac{k}{a^2} - \Lambda = 8\pi G_N \rho \quad (9.21)$$

From this equation one obtains the expression for the Critical density (i.e. the total density required for flat $k = \Lambda = 0$ Universe): $\rho_c = \frac{3}{8\pi G_N} \left(\frac{\dot{a}}{a} \right)^2$.

From the dynamical equation (9.21) one can obtain various relations between the Hubble parameter $H(z)$, the luminosity distance d_L , the deceleration parameter $q(z)$ and the energy densities ρ at various epochs of the Universe. For instance, for matter dominating flat ($k = 0$) Universes with $\Lambda > 0$ and various (simple, z -independent) equations of state $p = w_i \rho$, ($w_r = 1/3$ (radiation), $w_m = 0$ (matter-dust), $w_\Lambda = -1$ (cosmological constant (de Sitter))) we have for the Hubble parameter:

$$H(z) = H_0 \left(\sum_i \Omega_i (1+z)^{3(1+w_i)} \right)^{1/2} \quad (9.22)$$

with the notation: $\Omega_i \equiv \frac{\rho_i^0}{\rho_c^0}$, $i = r(\text{adiation}), m(\text{atter}), \Lambda, \dots$

From equations (8.42), by dividing by the Hubble parameter H , we can readily obtain an expression for the deceleration parameter at late eras, where radiation is negligible, in terms of Ω_m and Ω_Λ :

$$q(z) \equiv -\frac{\ddot{a}a}{(\dot{a})^2} = \left(\frac{H_0}{H(z)} \right)^2 \left(\frac{1}{2} \Omega_m (1+z)^3 - \Omega_\Lambda \right), \quad (9.23)$$

with $q_0 = \frac{1}{2}\Omega_m - \Omega_\Lambda$. Thus, it becomes evident that Λ acts as ‘‘repulsive’’ gravity, tending to accelerate the Universe currently, and eventually dominates, leading to an eternally accelerating de Sitter type Universe, with a future cosmic horizon. At present in the data there is also evidence for past deceleration ($q(z) > 0$, for some $z > z^* > 0$), which is to be expected if the dark energy is (almost) constant, due to matter dominance in earlier eras: $q(z) > 0 \Rightarrow (1+z)^3 > 2\Omega_\Lambda/\Omega_m \Rightarrow z > z^* = \left(\frac{2\Omega_\Lambda}{\Omega_m} \right)^{1/3} - 1$.

Finally, let us consider the case of a spatially flat Universe, $k = 0$, which is favoured by the current data, and give an expression for the luminosity distance in terms of the Hubble parameter $H(z)$ that we shall make use in the following. Using the null-geodesics of photons in a FRW Universe (9.11), we can solve for r_1 in (9.14) for the flat case $k = 0$, and obtain for the luminosity distance the important relation:

$$d_L = (1+z) \int_{t_1}^{t_0} \frac{a(t_0)dt}{a(t)} = -(1+z) \int_{a(t_1)}^{a(t_0)} \frac{d\left(\frac{a(t_0)}{a}\right)}{(\dot{a}/a)}, \quad (9.24)$$

from which

$$d_L = (1+z) \int_0^z \frac{dz}{H(z)} \quad (9.25)$$

We shall use this relation in the following, in order to constrain various theoretical cosmological models by means of astrophysical observations.

To give an idea how the various exact relationships differ from the approximate ones stated in the previous subsection, consider the case of a matter-dominated Universe. The Hubble-law relation between d_L and H measured today, which for measurements of nearby sources was given by (9.20), becomes in the exact case of a matter-dominated Universe [1]:

$$H_0 d_L = q_0^{-2} \left(z q_0 + (q_0 - 1) \left(\sqrt{2q_0 z + 1} - 1 \right) \right) \quad (9.26)$$

This completes our short discussion on the general concepts and methods used in astrophysical measurements. In the following subsections we discuss first the supernovae measurements of the cosmic acceleration, followed by a brief discussion on Cosmic Microwave Background measurements.

9.4.3 Supernovae Ia Measurements of Cosmic Acceleration

Type Ia Supernovae (SNe) behave as Excellent Standard Candles, and thus can be used to measure directly the expansion rate of the Universe at high redshifts ($z \geq 1$) and compare it with the present rate, thereby providing direct information on the Universe's acceleration. SNe type Ia are very bright objects, with absolute magnitude $M \sim 19.5$, typically comparable to the brightness of the entire host galaxy! This is why they can be detected at high redshifts $z \sim 1$, i.e. 3000 Mpc , $1 \text{ pc} \sim 3 \times 10^{16} m$. Detailed studies of the luminosity profile [4, 5] of each SNe suggests a strong relation between the width of the light curve and the absolute luminosity of SNe. This implies an accurate determination of its absolute luminosity. For each supernova one measures an effective (rest frame) magnitude in blue wavelength band, m_B^{eff} , which is then compared with the theoretical expectation (depending on the underlying model for the Universe) to yield information on the various Ω^i . The larger the magnitude the dimmer the observed SNe.

To understand the pertinent measurements recall the relation between the observed (on Earth) and emitted wavelengths $\lambda_{\text{obs}} = (1+z)\lambda_{\text{emit}}$, as a result of the cosmic redshift phenomenon (8.37). In a magnitude-redshift graph, if nothing slowed down matter blasted out of the Big Bang, we would expect a straight line. The data from High-redshift ($z \sim 1$) SNe Ia, showed that distant SNe lie slightly above the straight line. Thus they are moving away slower than expected. So at those early days ($z \sim 1$) the Universe was expanding at a slower rate than now. *The Universe accelerates today!* In such measurements, one needs the Hubble-Constant-Free Luminosity Distance:

$$\mathcal{D}_L(z; \Omega_M, \Omega_\Lambda) = \frac{H_0}{c} d_L, \quad d_L \equiv \sqrt{\frac{\mathcal{L}}{4\pi\mathcal{F}}},$$

with \mathcal{L} the intrinsic luminosity of the source, \mathcal{F} the measured flux and d_L the luminosity distance (9.10),(9.14). In Friedman models \mathcal{D}_L is parametrically known in terms of Ω_M, Ω_Λ , as a result of the corresponding dependence of d_L on $H(z)$, which in turn depends on Ω_i (c.f. (9.22)). An important quantity used in measurements is the Distance Modulus $m - M$, where

$$m = M + 25 + 5 \log \left(\frac{d_L}{1 \text{ Mpc}} \right) = \mathcal{M} + 5 \log \mathcal{D}_L,$$

with m =Apparent Magnitude of the Source, M the Absolute Magnitude, and $\mathcal{M} \equiv M - 5 \log H_0 + 25$ the fit parameter. Comparison of theoretical expectations with data restricts Ω_M, Ω_Λ . An important point to notice is that for fixed redshifts z the eqs. $\mathcal{D}_L(z; \Omega_M, \Omega_\Lambda) = \text{constant}$ yields *degeneracy curves* \mathcal{C} in the Ω -plane, of small curvature to which one associates a small slope, with the result that even very accurate data can at best select a narrow strip in Ω -plane parallel to \mathcal{C} . The results (2004) are summarized in figure 36 In the early works (1999) it was claimed that the best fit model, that of a FRW Universe with matter and cosmological constant for $z \leq 3$ (where the SNe data are valid) yields the following values: $0.8\Omega_M - 0.6\Omega_\Lambda \simeq -0.2 \pm 0.1$, for $\Omega_M \leq 1.5$. Assuming a flat model ($k=0$) the data imply: $\Omega_M^{Flat} = 0.28_{-0.08}^{+0.09}$ ($1\sigma \text{ stat}$) $_{-0.04}^{+0.05}$ (identified syst.), that is the *Universe accelerates today*

$$q_0 = \frac{1}{2}\Omega_M - \Omega_\Lambda \simeq -0.6 < 0$$

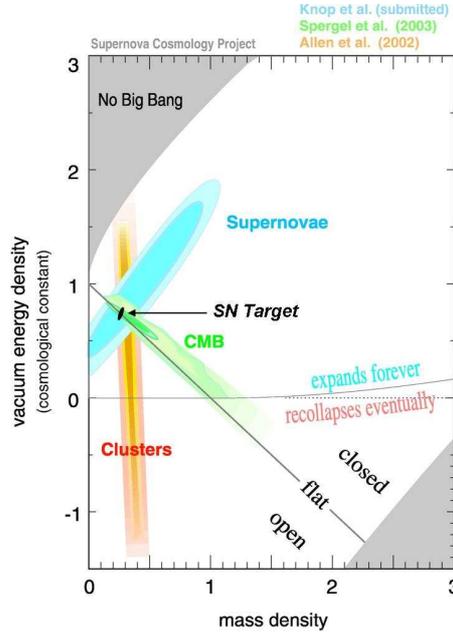


Figure 36: Supernovae (and other) measurements on the Universe’s energy budget.

Further support on these results comes, within the SNe measurement framework, from the recent (> 2004) discovery [5], by Hubble Space Telescope, ESSENCE and SNLS Collaborations, of more than 100 high- z ($2 > z \geq 1$) supernovae, pointing towards the fact that for the past 9 billion years the energy budget of the Universe is dominated by an approximately constant dark energy component.

9.4.4 CMB Anisotropy Measurements by WMAP1,3: brief comments

After three years of running, WMAP provided a much more detailed picture of the temperature fluctuations than its COBE predecessor, which can be analyzed to provide best fit models for cosmology, leading to severe constraints on the energy content of various model Universes, useful for particle physics, and in particular supersymmetric searches. Theoretically [1], the temperature fluctuations in the CMB radiation are attributed to: (i) our velocity w.r.t cosmic rest frame, (ii) gravitational potential fluctuations on the last scattering surface (Sachs-Wolf effect), (iii) Radiation field fluctuations on the last scattering surface, (iv) velocity of the last scattering surface, and (v) damping of anisotropies if Universe re-ionizes after decoupling. A Gaussian model of fluctuations [1], favored by inflation, is in very good agreement with the recent WMAP data (see figure 37). The perfect fit of the first few peaks to the data allows a precise determination of the total density of the Universe, which implies its spatial flatness. The various peaks in the spectrum of fig. 37 contain interesting physical signatures:

- (i) The angular scale of the first peak determines the curvature (but not the topology) of the Universe.
- (ii) The second peak –truly the ratio of the odd peaks to the even peaks– determines the reduced baryon density.
- (iii) The third peak can be used to extract information about the dark matter (DM) density (this is a model-dependent result, though –standard local Lorentz invariance assumed, see discussion in later sections on Lorentz-violating alternative to dark matter models).

The measurements of the WMAP [9] on the cosmological parameters of interest to us here are

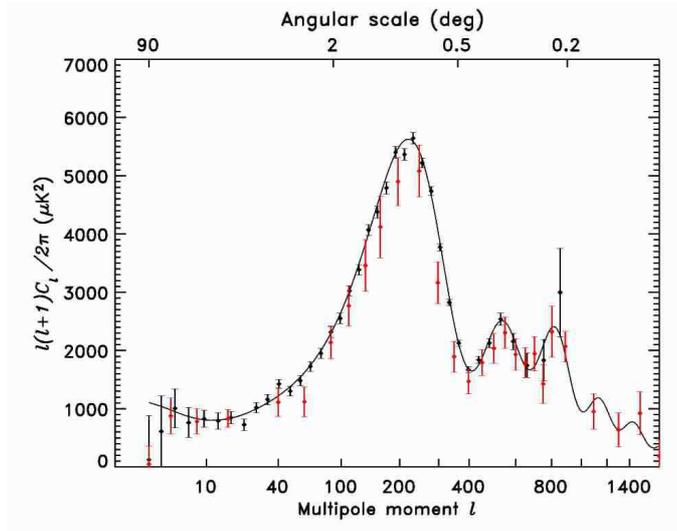


Figure 37: Red points (larger errors) are previous measurements. Black points (smaller errors) are WMAP measurements (G. Hinshaw, *et al.* arXiv:astro-ph/0302217).

given in [9], and reviewed in [3]. The WMAP results constrain severely the equation of state $p = w\rho$ (p = pressure), pointing towards $w < -0.78$, if one fits the data with the assumption $-1 \leq w$ (we note for comparison that in the scenarios advocating the existence of a cosmological constant one has $w = -1$). Many quintessence models can easily satisfy the criterion $-1 < w < -0.78$, especially the supersymmetric ones, which we shall comment upon later in the article. Thus, at present, the available data are not sufficient to distinguish the cosmological constant model from quintessence (or more generally from relaxation models of the vacuum energy). The results lead to the chart for the energy and matter content of our Universe depicted in figure 35, and are in perfect agreement with the Supernovae Ia Data [4]. The data of the WMAP satellite lead to a new determination of $\Omega_{\text{total}} = 1.02 \pm 0.02$, where $\Omega_{\text{total}} = \rho_{\text{total}}/\rho_c$, due to high precision measurements of secondary (two more) acoustic peaks as compared with previous CMB measurements (c.f. figure 37). Essentially the value of Ω is determined by the position of the first acoustic peak in a Gaussian model, whose reliability increases significantly by the discovery of secondary peaks and their excellent fit with the Gaussian model [9].

Finally we mention that the determination of the cosmological parameters by the WMAP team [9], after three years of running, favors, by means of best fit procedure, *spatially flat* inflationary models of the Universe [14]. In general, WMAP gave values for important inflationary parameters, such as the running spectral index, $n_s(k)$, of the primordial power spectrum of scalar density fluctuations $\delta_{\vec{k}}$ [15] $P(k) \equiv |\delta_{\vec{k}}|^2$. The running scalar spectral index $n_s(k)$ is $n_s(k) = \frac{d \ln P(k)}{d \ln k}$, where k is the co-moving scale. Basically inflation implies $n_s = 1$. WMAP measurements yield $n_s = 0.96$, thus favoring Gaussian primordial fluctuations, as predicted by inflation. For more details we refer the reader to the literature [9, 3].

To summarize, WMAP-CMB measurements, combined with high-redshift supernovae ones and others (c.f. below), gave a pretty detailed information on the history of our Big-Bang Universe. A schematic time-line of the Universe, according to such measurements, is given in fig. 38.

9.4.5 Baryon Acoustic Oscillations (BAO)

Further evidence for the energy budget of the Universe is obtained by Detection of the baryon acoustic peak in the large-scale correlation function of SDSS luminous red galaxies [7]. The underlying Physics of BAO can be understood as follows: Because the universe has a significant fraction of baryons, cosmological theory predicts that the acoustic oscillations (CMB) in the plasma will

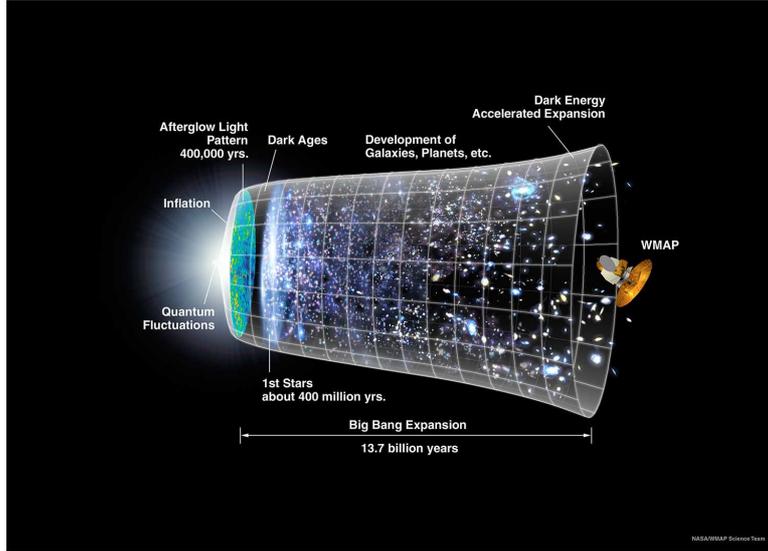


Figure 38: The time-line of the Universe according to the WMAP satellite measurements (picture from <http://map.gsfc.nasa.gov/resources/imagetopics.html>).

also be imprinted onto the late-time power spectrum of the non-relativistic matter: from an initial point perturbation common to the dark matter and the baryons, the dark matter perturbation grows in place while the baryonic perturbation is carried outward in an expanding spherical wave. At recombination, this shell is roughly 150 Mpc in radius. Afterwards, the combined dark matter and baryon perturbation seeds the formation of large-scale structure. Because the central perturbation in the dark matter is dominant compared to the baryonic shell, the acoustic feature is manifested as a small single spike in the correlation function at 150 Mpc separation [7].

The acoustic signatures in the large-scale clustering of galaxies yield three more opportunities to test the cosmological paradigm with the early-universe acoustic phenomenon:

- (i) They would provide smoking-gun evidence for the theory of gravitational clustering, notably the idea that large-scale fluctuations grow by linear perturbation theory from $z \sim 1000$ to the present;
- (ii) they would give another confirmation of the existence of dark matter at $z \sim 1000$, since a fully baryonic model produces an effect much larger than observed;
- (iii) they would provide a characteristic and reasonably sharp length scale that can be measured at a wide range of redshifts, thereby determining purely by geometry the angular-diameter-distance-redshift relation and the evolution of the Hubble parameter.

In the current status of affairs of the BAO measurements it seems that there is an underlying-theoretical-model dependence of the interpretation of the results, as far as the predicted energy budget for the Universe is concerned. This stems from the fact that for small deviations from $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, the change in the Hubble parameter at $z = 0.35$ is about half of that of the angular diameter distance. Eisenstein *et al.* in [7] modelled this by treating the dilation scale as the cubic root of the product of the radial dilation times the square of the transverse dilation. In other words, they defined

$$D_V(z) = \left[D_M(z)^2 \frac{cz}{H(z)} \right]^{1/3}, \quad H = H_0 \left(\sum_i \Omega_i (1+z)^{3(1+w_i)} \right)^{1/2} \quad (9.27)$$

where $H(z)$ is the Hubble parameter and $D_M(z)$ is the co-moving angular diameter distance. As the typical redshift of the sample is $z = 0.35$, we quote the result [7] for the dilation scale

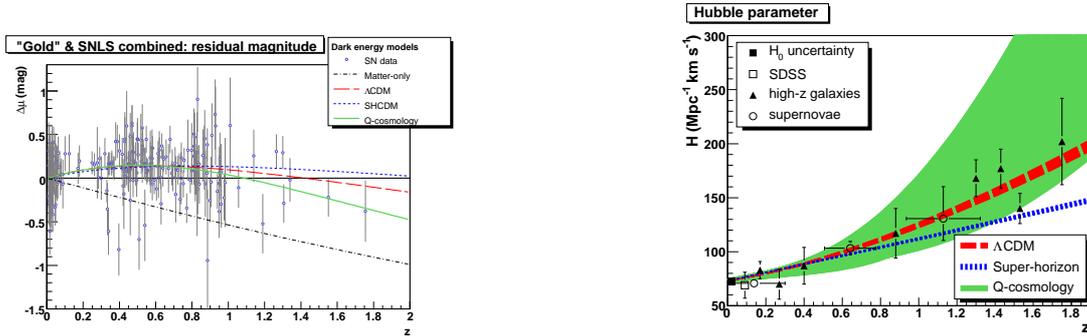


Figure 39: Left: Residual magnitude versus redshift for supernovae from the ‘gold’ and the SNLS datasets for various cosmological models. Right: The Hubble-parameter vs. redshift relation for these models and observational data. The bands represent 68% confidence intervals derived by the SN analysis for the standard Λ CDM, the super-horizon (no DE) and the Q-cosmology models. The black rectangle shows the WMAP3 estimate for H_0 , the squares show the measurements from SDSS galaxies, the triangles result from high- z red galaxies, and the circles correspond to a combined analysis of supernovae data (from [16]).

as $D_V(0.35) = 1370 \pm 64 \text{Mpc}$. The BAO measurements from Large Galactic Surveys and their results for the dark sector of the Universe are consistent with the WMAP data, as far as the energy budget of the Universe is concerned, but the reader should bear in mind that they based their parametrization on standard FRW cosmologies, so the consistency should be interpreted within that theory framework.

9.4.6 Measuring $H(z)$: an important constraint on models

The previous results, based on SNe, CMB and BAO measurements, relied on the standard FRW Cosmological model for the Universe as the underlying theory. However, in modern approaches to (quantum) gravity, such as brane and string theories, the underlying dynamics may no longer be described by the simple Einstein-Hilbert action. One may have extra fields, such as the dilaton or moduli fields in theories with extra dimensions, plus higher order curvature terms which could become important in the early Universe. Moreover, there have been suggestions in the literature [16] that the claimed Dark Energy may not be there, but simply be the result of temperature fluctuations in a (flat) Universe filled with matter $\Omega_M = 1$ (“super-horizon model”). All such alternative theories should be tested against each one of the above-mentioned categories of measurements together with an independent measurement of the behavior of the Hubble parameter vs. the redshift $H(z)$, the latter coming from large galactic surveys. This latter measurement provides an *important constraint* which could differentiate among the potential Dark Energy (DE)/Dark Matter (DM) models and their alternatives. This extra measurement has the potential of ruling out alternative models (to DM and DE) that otherwise fit the supernova data alone (in a m_{eff} vs z plot). This happens, for instance, with the super-horizon model of [16]. I mention in passing that other non-equilibrium stringy cosmologies [12], with relaxing to zero dark energy (quintessence-like due to the dilaton field) survive at present this constraint, as illustrated in figure 39. For more details I refer the reader to [17] and references therein.

9.4.7 Cosmic Coincidence and Cosmological Constant Issues

There may be several possible explanations regarding the Dark Energy part of the Universe’s energy budget:

(i) The dark energy is an “Honest” Cosmological Constant $\Lambda \sim 10^{-122} M_{\text{Pl}}^4$, strictly unchanging through space and time. This has been the working hypothesis of many of the best fits so far, but

I stress it is *not* the only explanation consistent with the data.

(ii) Quintessence: The Cosmological constant is mimicked by a slowly-varying field, ϕ , whose time until it reaches its potential minimum is (much) longer than the Age of Universe. Simplest Quintessence models assume exponential potentials of the scalar field representing quintessence:

$$V(\phi) \sim e^\phi .$$

In such a case the pertinent equation of state reads:

$$w = \frac{\frac{(\dot{\phi})^2}{2} - V(\phi)}{\frac{(\dot{\phi})^2}{2} + V(\phi)} .$$

For $\phi = -2\ln t$ one has a relaxing-to-zero vacuum energy $\Lambda(t) \sim \text{const}/t^2$ (in Planck units), of the right order of magnitude today. Such a situation could be met [12] in some models of string theory, where the rôle of the quintessence field could be played by the dilaton [18], i.e. the scalar field of the string gravitational multiplet.

(iii) Einstein-Friedman model is incorrect, and one could have modifications in the gravitational law at galactic or supergalactic scales. Models of this kind have been proposed as alternatives to dark matter, for instance Modified Newtonian Dynamics (MOND) by Milgrom [19], and its field theory version by Bekenstein [20], known as Tensor-Vector-Scalar (TeVeS) theory, which however, is Lorentz Violating, as it involves a preferred frame. Other modifications from Einstein theory, which however maintain Lorentz invariance of the four-dimensional world, could be brane models for the Universe, which are characterized by non-trivial, and in most cases time dependent, vacuum energy. It should be noted that such alternative models may lead to completely different energy budget [21, 22]. We shall discuss one such case of a non-critical string inspired (non-equilibrium, relaxation) cosmology (Q-cosmology) in a subsequent section, where we shall see that one may still fit the astrophysical data with exotic forms of “dark matter”, not scaling like dust with the redshift at late epochs, and different percentages of dark (dilaton quintessence) energy (c.f. also fig. 39).

Given that from most of the standard best fits for the Universe it follows that the energy budget of our Cosmos today is characterized by 73 – 74% vacuum energy, i.e. an energy density of order

$$\rho_{vac} \simeq (10^{-3} \text{ eV})^4 = 10^{-8} \text{ erg/cm}^3 ,$$

and about 27 – 26% matter (mostly dark, only about 4% visible, ordinary matter), this implies the *Coincidence Problem*: “The vacuum energy density today is approximately equal (in order of magnitude) to the current matter density.” As the Universe expands, this relative balance is lost in models with a cosmological constant, such as the standard Λ CDM model, since the matter density scales with the scale factor as

$$\frac{\Omega_\Lambda}{\Omega_M} = \frac{\rho_\Lambda}{\rho_M} \propto a^3 .$$

In this framework, at early times we have that the Vacuum Energy is much more suppressed as compared with that of Matter and Radiation, while at late times it dominates. There is only one brief epoch for which the transition from domination of one component to the other can be witnessed, and this epoch, according to the Λ CDM model, happened to be the present one! This calls for a microscopic Explanation, which is still lacking.

The smallness of the value of the Dark Energy today is another big mystery of particle physics. For several years the particle physics community thought that the vacuum energy was exactly zero, and in fact they were trying to devise microscopic explanations for such a vanishing by means of some symmetry. One of the most appealing, but eventually failed in this respect, symmetry justifications for the vanishing of the vacuum energy was that of *supersymmetry* (SUSY): if unbroken, supersymmetry implies strictly a vanishing vacuum energy, as a result of the cancelation among boson and fermion vacuum-energy contributions, due to opposite signs in the respective quantum

loops. However, this cannot be the correct explanation, given that SUSY, if it is to describe Nature, must be broken below some energy scale M_{susy} , which should be higher than a few TeV, as partners have not been observed as yet. In broken SUSY theories, in four dimensional space times, there are contributions to vacuum energy

$$\rho_{\text{vac-SUSY}} \propto \hbar M_{\text{susy}}^4 \sim (\text{few TeV})^4 ,$$

which is by far greater than the observed value today of the dark energy

$$\Lambda \sim 10^{-122} M_{\text{Pl}}^4 , \quad M_{\text{Pl}} \sim 10^{19} \text{ GeV} .$$

Thus, SUSY does not solve the *Cosmological Constant Problem*, which at present remains one of the greatest mysteries in Physics.

In my opinion, the smallness today of the value of the “vacuum” energy density might point towards a relaxation problem. Our world may have not yet reached equilibrium, from which it departed during an early-epoch cosmically catastrophic event, such as a Big Bang, or—in the modern version of string/brane theory—a collision between two brane worlds. This non equilibrium situation might be expressed today by a quintessence(ϕ)-like exponential potential

$$V(\phi) \sim \exp(\phi) ,$$

where ϕ could be the dilaton field, which in some models [12] behave at late cosmic times t as

$$\phi \sim -2\ln t .$$

This would predict a vacuum energy today of order $1/t^2$, which has the right order of magnitude, if t is of order of the Age of the Universe, i.e. $t \sim 10^{60}$ Planck times. Supersymmetry in such a picture may indeed be a symmetry of the vacuum, reached asymptotically, hence the asymptotic vanishing of the dark energy. SUSY breaking may not be a spontaneous breaking but an *obstruction*, in the sense that only the excitation particle spectrum has mass differences between fermions and bosons. To achieve phenomenologically realistic situations, one may exploit [23] the string/brane framework, by compactifying the extra dimensions into manifolds with non-trivial “fluxes” (these are not gauge fields associated with electromagnetic interactions, but pertain to extra-dimensional unbroken gauge symmetries characterizing the string models). In such cases, fermions and bosons couple differently, due to their spin, to these flux gauge fields (a sort of generalized “Zeeman” effects). Thus, they exhibit mass splittings proportional to the square of the “magnetic field”, which could then be tuned to yield phenomenologically acceptable SUSY-splittings, while the relaxation dark energy has the cosmologically observed small value today. In such a picture, SUSY is needed for stability of the vacuum, although today, in view of the landscape scenarios for string theory, one might not even have supersymmetric vacua at all. However, there may be another reason why SUSY could play an important physical rôle, that of dark matter. I now come to discuss this important issue, mainly from a particle physics perspective.

10 Dark Matter (DM)

In this section I will discuss issues pertaining to dark matter and supersymmetry. I will first make the case for Dark Matter, starting historically from discrepancies concerning rotational curves of galaxies. Then I will move to describe possible candidates, and based on standard models for cosmology to exclude many of them, by means of WMAP data, arguing that supersymmetric dark matter remains compatible with such data. I will again emphasize, however, the model dependence of such conclusions. Then I will proceed to discuss supersymmetric particle physics constraints in various frameworks by describing the underlying general framework for calculating thermal dark matter relics and compare them with WMAP data. For a more complete discussion on direct searches for dark matter the reader is referred to [24], and references therein.

10.1 The Case for DM

Dark Matter (DM) is defined as a Non luminous massive matter, of unknown composition, that does not emit or reflect enough electromagnetic radiation to be observed directly, but whose presence can be inferred from gravitational effects on visible matter. Observed phenomena consistent with the existence of dark matter are:

- (i) rotational speeds of galaxies and orbital velocities of galactic clusters,
- (ii) gravitational lensing of background objects by galaxy clusters such as the Bullet cluster of galaxies, and
- (iii) the temperature distribution of hot gas in galaxies and clusters of galaxies.
- (iv) As we have seen, DM also plays a central role in structure formation and galaxy evolution, and has measurable effects on the anisotropy of the cosmic microwave background, especially the third peak in the anisotropy spectrum (c.f. fig. 37).

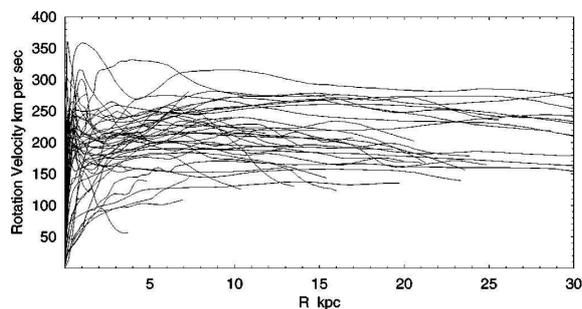


Figure 40: Collage of Rotational Curves of nearby spiral galaxies obtained by combining Doppler data from CO molecular lines for the central regions, optical lines for the disks, and HI 21 cm line for the outer (gas) disks. Graph from Y. Sofue and V. Rubin (Annual Review of Astronomy and Astrophysics, Volume 31 (c)2001, 127).

Historically, the first evidence for DM came [25] from discrepancies concerning the Rotational Curves (RC) of Galaxies. If all matter were luminous then the rotational speed of the galactic disc would fall with the (radial) distance r from the center as $v(r) \sim r^{-1/2}$ but observations show that $v(r) \sim \text{const}$, as seen clearly in figure 40, where the rotation velocity in units of km s^{-1} is plotted vs galactocentric radius R in kiloparsecs (kpc); $1 \text{ kpc} \approx 3000 \text{ light years}$. It is seen that the RCs are flat to well beyond the edges of the optical disks ($\sim 10 \text{ kpc}$). Further Evidence for DM is provided by the Matter oscillation spectrum in galaxies, depicted in figure 41. The observed spectrum does not have the pronounced wiggles predicted by a baryon-only model, but it also has significantly higher power than does the model. In fact, $\Delta^2 = k^3 P(k)/(2\pi^2)$, which is a dimensionless measure of the clumping, never rises above one in a baryon-only model, so we could not see any large structures (clusters, galaxies, people, etc.) in the universe in such a model [26].

However, at this stage we should mention the alternatives to Dark Matter models, the MOND [19], and its Lorentz-violating TeVeS field theory version [20], which could also reproduce the rotational curves of galaxies, by assuming modified Newtonian dynamics at galactic scales for small gravitational accelerations, smaller than a universal value $\gamma < \gamma_0 \sim (200 \text{ km sec}^{-1})^2 / (10 \text{ kpc})$. MOND theories have been claimed to fit most of the rotational curves of galaxies (fig. 40), with few notable exceptions, though, e.g. the bullet cluster. It should be mentioned that TeVeS models, due to their preferred-cosmic-frame features, are characterized by ‘‘Aether’’-Lorentz violating isotropic vector fields $A_\mu = (f(t), 0, 0, 0)$, $A_\mu A^\mu = -1$, whose cosmic instabilities are also claimed [22] to reproduce the enhanced growth of perturbations observed in galaxies (c.f. fig. 41). In these lectures I will not discuss such models. It should be noted at this point that such issues, namely whether there are dark matter particles or not, could be resolved in principle by particle physics searches at colliders or direct dark matter searches, which I will now come to.

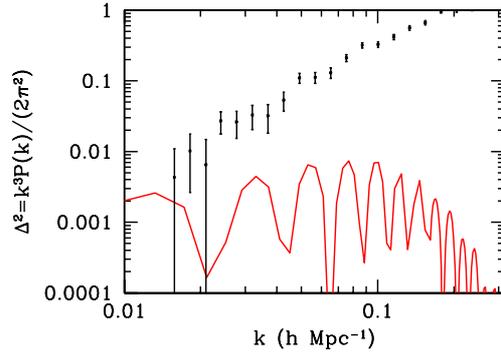


Figure 41: Power spectrum of matter fluctuations (red curve, with wiggles) in a theory without dark matter as compared to observations of the galaxy power spectrum.

10.2 Types of DM and Candidates

From nucleosynthesis constraints we can estimate today the baryonic energy density contribution to be of order: $\Omega_{baryons} = 0.045 \pm 0.01$, and this in fact is the dominant form of ordinary matter in the Universe. Thus, barring alternatives, 90% of the alleged matter content of the Universe seems to be dominated by DM of unknown composition at present. There are several dark matter candidates, which can be classified into two large categories depending on their origin and properties:

(I) Astrophysical: (i) MASSive Compact Halo ObectS (MACHOS): Dwarf stars and Planets (Baryonic Dark Matter) and Black Holes, (ii) Non-luminous Gas Clouds.

(II) Particles (Non-Baryonic Dark Matter): Weakly Interacting Massive Particles (WIMP), which might be the best candidates for DM: should not have electromagnetic or strong interactions. May have weak and gravitational interactions. WIMPS might include axions, neutrinos stable supersymmetric partners *etc.* If these WIMPS are thermal relics from the Big Bang then we can calculate their relic abundance today and compare with CMB and other astrophysical data. Non-thermal relics may also exist in some cosmological models but will not be the subject of our discussion in these lectures.

There is an alternative classification of DM, depending on the energetics of the constituting particles:

(i) Hot Dark Matter (HDM): form of dark matter which consists of particles that travel with ultra-relativistic velocities: e.g. neutrinos.

(ii) Cold Dark Matter (CDM): form of dark matter consisting of slowly moving particles, hence cold, e.g. WIMPS (stable supersymmetric particles (e.g. neutralinos *etc.*) or MACHOS.

(iii) Warm Dark Matter (WDM): form of dark matter with properties between those of HDM and CDM. Examples include sterile neutrinos, light gravitinos in supergravity theories *etc.*

Particle physics and/or astrophysics should provide candidates for DM and also explain the relic densities of the right order as that predicted by the data. Currently, the most favorite SUSY candidate for non baryonic CDM are neutralinos [27] $\tilde{\chi}$. These particles could be a WIMP if they are stable, which is the case in models where they are the Lightest SUSY Particles (LSP) (with typical masses $m_{\tilde{\chi}} > 35 \text{ GeV}$). Most of supersymmetric model constraints come from the requirement that a neutralino is the dominant astrophysical DM, whose relic abundance can explain the missing Universe mass problem. I mention at this stage that direct searches for $\tilde{\chi}$ involve, among others, the recoil of nucleons during their interaction with $\tilde{\chi}$ in cryogenic materials. In these lectures we shall concentrate mainly on colliders DM searches. I refer the reader to ref. [24] for direct DM searches and other pertinent terrestrial and extraterrestrial experiments.

10.3 WIMP DM: thermal properties and relic densities: The Boltzmann equation for species abundances

In all the searches we shall deal with in the present set of notes, which are also the most commonly studied in the literature, one makes the standard assumption that the dark matter particle, χ , is a *thermal relic* of the Big Bang of mass m_χ : when the early Universe was dense and hot, with temperature $T \gg m_\chi$, χ was in *thermal equilibrium*; annihilation of χ and its antiparticle $\bar{\chi}$ into lighter particles, $\chi\bar{\chi} \rightarrow \bar{l}l$, and the inverse process $\bar{l}l \rightarrow \chi\bar{\chi}$ proceeded with equal rates [1]. As the Universe expanded and cooled down to a temperature $T < m_\chi$, the number density of χ dropped exponentially, $n_\chi \sim e^{-m_\chi/T}$. Eventually, the temperature became too low for the annihilation to keep up with the expansion rate and the species χ ‘froze out’ with the cosmological *abundance* (“relic”) observed today.

As we shall prove below, the time evolution of the number density $n_\chi(t)$ is determined by the Boltzmann equation [1],

$$\boxed{dn_\chi/dt + 3Hn_\chi = -\langle\sigma_A v\rangle [(n_\chi)^2 - (n_\chi^{\text{eq}})^2]}, \quad (10.1)$$

where H is the Hubble expansion rate, n_χ^{eq} the equilibrium number density and $\langle\sigma_A v\rangle$ is the thermally averaged annihilation cross section summed over all contributing channels. It turns out that the relic abundance today is inversely proportional to the thermally averaged annihilation cross section, $\Omega_\chi h^2 \sim 1/\langle\sigma_A v\rangle$. The situation is depicted in fig. 42. When the properties and interactions of the WIMP are known, its thermal relic abundance can hence be computed from particle physics’ principles and compared with cosmological data.

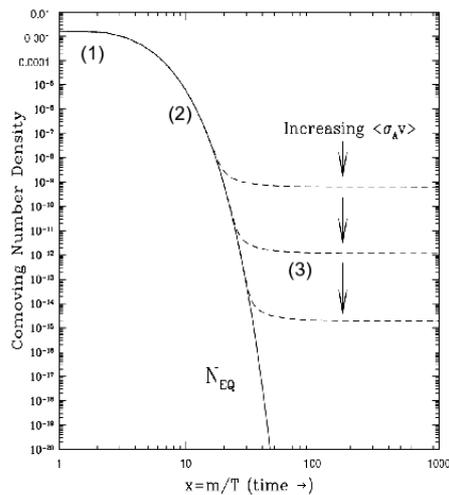


Figure 42: The full line is the equilibrium abundance; the dashed lines are the actual abundance after freeze-out. As the annihilation cross section $\langle\sigma_A v\rangle$ is increased, the WIMP stays in equilibrium longer, leading to a smaller relic density (from ref. [1]).

Derivation of the Boltzmann Equation in RW space-time

We now proceed to discussing briefly a derivation of Eq. (10.1), based on our knowledge of general relativity and cosmology so far. We shall be very sketchy in our discussion, providing only essential information, appropriate for a graduate student to understand the basic concepts and techniques involved. For more details we refer the interested reader in the literature [1].

The Boltzmann equation essentially expresses the action of the so-called Liouville operator $\hat{L}[f]$ on the phase-space density of the species χ , $f(\vec{x}, |\vec{p}|, t)$, in terms of the so-called collision operator, $C[f]$, monitoring the deviation from equilibrium in the reactions that the species χ participates.

In the non-relativistic (Newtonian) case, the Liouville operator is a total time derivative, time is universal, and $\vec{x}(t)$, $\vec{p}(t)$ depend on time (phase-space trajectories of the particle): so its action on $f(\vec{x}, |\vec{p}|, t)$ is given by:

$$\widehat{L}[f] = \frac{d}{dt}f = \frac{\partial}{\partial t}f + \vec{v} \cdot \vec{\nabla}f + \frac{\vec{F}}{m_\chi} \cdot \nabla_{\vec{v}}f \quad (10.2)$$

where $\vec{v} = d\vec{x}/dt$ is the velocity, and $\vec{F} = d\vec{p}/dt$ is the (Newtonian) force acting on the particle.

The extension of (10.2) to the general-relativistic case, that will allow treatment in the Robertson-Walker Universe, is straightforward. Essentially, the Newtonian total time derivative of the non-relativistic case is replaced by a total *derivative* with respect the *proper time*, which plays the rôle of the universal Newtonian time, as we have repeatedly stressed in these Lectures. The resulting Liouville operator is essentially,

$$\widehat{L}[f] \rightarrow m_\chi \frac{d}{d\tau}f = m_\chi u^\alpha \partial_\alpha f + m_\chi \frac{dp^\alpha}{d\tau} \frac{\partial}{\partial p^\alpha}f, \quad (10.3)$$

where u^μ is the four-velocity (3.17) and $p^\mu = m_\chi u^\mu$ the four-momentum (3.19). In (10.3) we took into account that any dependence of the phase-space density f on the proper time τ is through the dependence of $x^\mu(\tau)$, $p^\alpha(\tau)$ on τ .

Based on our discussion so far, then, the combination $\frac{\partial}{\partial t}f + \vec{v} \cdot \vec{\nabla}$ of the Newtonian case is replaced in General Relativity by $p^\alpha \partial_\alpha$, whilst the ‘force’ term is expressed in terms of the Christoffel symbols by means of the geodesic equation (5.19), i.e. $m_\chi \frac{dp^\alpha}{d\tau} = -\Gamma_{\alpha\beta}^\mu p^\alpha p^\beta$.

The result for the general-relativistic Liouville operator is, therefore:

$$\widehat{L}[f] = [p^\alpha \partial_\alpha - \Gamma_{\mu\nu}^\alpha p^\mu p^\nu \frac{\partial}{\partial p^\alpha}]f. \quad (10.4)$$

For a homogeneous and isotropic Robertson-Walker Universe, we have that $f = f(t, |\vec{p}|)$ or, equivalently, upon using the RW-space-time on-shell condition for the massive species χ , $f = f(E, t)$, where E denotes the energy of the dark matter particle and t is the co-moving frame RW cosmic time. In such a case, upon using the Christoffel symbols for the Robertson-Walker metric (8.18), we obtain from (10.4):

$$\widehat{L}[f] = E \frac{\partial f}{\partial t} - \frac{\dot{a}}{a} |\vec{p}|^2 \frac{\partial f}{\partial E}. \quad (10.5)$$

The number density of species n_χ is defined as:

$$n_\chi = \frac{g}{8\pi^3} \int d^3p f(E, t) \quad (10.6)$$

where g is the number of degrees of freedom of the species χ . Acting on n_χ with the general-relativistic Liouville operator for the RW Universe (10.5), we obtain:

$$\begin{aligned} \widehat{L}[n_\chi] &= E \frac{dn_\chi}{dt} - \frac{\dot{a}}{a} \int d|\vec{p}| d\Omega |\vec{p}|^4 \frac{\partial |\vec{p}|}{\partial E} \frac{\partial f}{\partial |\vec{p}|} = \\ &= E \frac{dn_\chi}{dt} - E \frac{\dot{a}}{a} \int d|\vec{p}| d\Omega |\vec{p}|^3 \frac{\partial f}{\partial |\vec{p}|} \end{aligned} \quad (10.7)$$

where in the last step we have split the momentum integration into momentum-amplitude ($|\vec{p}|$) and angular (Ω) parts, and transformed the E -differentiation to a $|\vec{p}|$ -differentiation, using $\frac{\partial |\vec{p}|}{\partial E} = \frac{E}{|\vec{p}|}$, as a result of the (on-shell) dispersion relation $|\vec{p}|^2 + m_\chi^2 = E^2$ for the dark matter species χ , with the notation $|\vec{p}|^2 \equiv p_i p_j h^{ij}$, where h_{ij} is the spatial part of the RW metric in the notation of (8.17). By partially integrating the last term on the right-hand-side of (10.7), we then obtain the Boltzmann equation for the number density in the form:

$$dn_\chi/dt + 3Hn_\chi = \frac{g}{8\pi^3} \int \frac{d^3p}{E} C[f], \quad H = \frac{\dot{a}}{a}. \quad (10.8)$$

The Collision Term

We next discuss the collision operator $C[f]$, following [1] and references therein. Consider the process: $\chi + I_1 + I_2 + \dots \rightleftharpoons F_1 + F_2 + \dots$. The relevant collision term is then given by:

$$\begin{aligned} & \frac{g}{(2\pi)^3} \int C[f] \frac{d^3 p_\chi}{E_\chi} = \\ & - \int d\Pi_\chi d\Pi_{I_1} d\Pi_{I_2} \dots d\Pi_{F_1} d\Pi_{F_2} \dots \times (2\pi)^4 \delta^{(4)}(p_\chi + p_{I_1} + p_{I_2} + \dots - p_{F_1} - p_{F_2} - \dots) \times \\ & (|\mathcal{M}|_{\chi+I_1+I_2+\dots \rightarrow F_1+F_2+\dots}^2 f_{I_1} f_{I_2} \dots f_\chi (1 \pm f_{F_1})(1 \pm f_{F_2}) \dots - \\ & |\mathcal{M}|_{F_1+F_2+\dots \rightarrow \chi+I_1+I_2+\dots}^2 f_{F_1} f_{F_2} \dots (1 \pm f_{I_1})(1 \pm f_{I_2}) \dots (1 \pm f_\chi)) \end{aligned} \quad (10.9)$$

where f_k denotes the appropriate phase-space densities of the particle-species k , the sign (+) refers to bosons, whilst (-) refers to fermions, and $d\Pi_k \equiv \frac{g_k}{(2\pi)^3} \frac{d^3 p_k}{E_k}$, with g_k the internal degrees of freedom of the species k . The four-dimensional δ -function is the result of energy-momentum conservation in the interactions, and $|\mathcal{M}|_{\text{reaction}}^2$ denotes the matrix element squared of the pertinent interaction, including average over initial and final spin states, appropriate symmetry factors for identical particles in the initial and final states (if any). The rules for calculating such matrix elements can be found in any graduate textbook on quantum field theory, and has been covered in your field theory courses in the doctorate programme, where you are referred for further details.

The expression (10.9) is greatly simplified if we make some assumptions about discrete symmetries that may characterise the interactions. If one assumes Time Reversal Invariance (T -invariance), then, on account of CPT symmetry, which is assumed to characterise the quantum field theory under consideration this will also imply CP invariance⁶. Upon the T -reversal invariance assumption, we have equality of the scattering amplitudes describing the two-different directions of the reaction involving the dark-matter species χ :

$$|\mathcal{M}|_{\chi+I_1+I_2+\dots \rightarrow F_1+F_2+\dots}^2 = |\mathcal{M}|_{F_1+F_2+\dots \rightarrow \chi+I_1+I_2+\dots}^2 \equiv |\mathcal{M}|^2 \quad (10.10)$$

Another simplifying assumption, common in dark matter studies, is the replacement of the Fermi-Dirac or Bose statistics (c.f. Appendix B) of the individual species by the common Maxwell-Boltzmann statistics for *all species*. This assumption is a good approximation in the absence of Bose condensation or Fermi degeneracy, and implies that one may approximate the factors $1 \pm f_k \simeq 1$, and

$$f_i(E_i) \simeq e^{-(E_i - \mu_i)/k_B T} \quad (10.11)$$

for all species in *kinetic equilibrium*, where T is the temperature, E_i is the particle energy, μ_i is the species chemical potential and k_B is the Boltzmann constant, assumed one from now on (choice of units).

Under the above simplifications, the Boltzmann equation (10.8),(10.9) becomes:

$$\begin{aligned} \dot{n}_\chi + 3Hn_\chi = & - \int d\Pi_\chi d\Pi_{I_1} d\Pi_{I_2} \dots d\Pi_{F_1} d\Pi_{F_2} \dots \times |\mathcal{M}|^2 \times \\ & (2\pi)^4 \delta^{(4)}(p_\chi + p_{I_1} + p_{I_2} + \dots - p_{F_1} - p_{F_2} - \dots) (f_{I_1} f_{I_2} \dots f_\chi - f_{F_1} f_{F_2} \dots) . \end{aligned} \quad (10.12)$$

⁶We note at this stage that CP invariance is relaxed in modern Cosmology, when Baryogenesis is considered [1]. In view of CPT symmetry, this implies also T -reversal violation. We also note that in some non-equilibrium theories of Cosmology, involving quantum decoherence of matter due to environmental entanglement with quantum-gravitational degrees of freedom, CPT might also be violated, and this might be the case of very early epochs of the Universe, where strong quantum gravitational fluctuations (that we ignored throughout our notes here) might induce CPT violation, as a result of their potential singular-curvature characteristics. If CPT is violated, the rules on computing the Boltzmann equation change, and the equation itself gets modified by CPT -violating source terms [28]. Such issues are at present mere speculations, as currently there is no experimental evidence for CPT violation. However, it should be noted that such theories lead to relaxation of some of the stringent constraints imposed on particle physics models from astrophysical measurements of dark matter, due to the induced modifications of the amount of dark matter relics in such models, which is found smaller as compared to that of standard cosmology scenarios.

We next consider case of a *stable* dark matter species χ , which is of interest to us here ⁷ and which will lead to the form (10.1) of the collision term. Since the species is stable, the only interactions it can participate are *annihilation* with its antiparticle and its *inverse*, namely

$$\chi\bar{\chi} \rightleftharpoons X\bar{X} \quad (10.13)$$

We also assume that the species X have zero chemical potential, for simplicity. In most cases, the species X are characterised by much stronger interactions than χ , so the assumption of equilibrium for them is a good one. Indeed, as we shall discuss below in sec. 10.5, in case the dark matter particle represents a neutralino of a supersymmetric theory, its annihilation reactions involve in the final state electromagnetically charged Standard Model particles, which have electromagnetic interactions, stronger than the weak interactions of the neutralino *etc.* In this case:

$$f_X = e^{-E_X/T} \quad (10.14)$$

and similarly for \bar{X} . Energy conservation, enforced by the δ -function in (10.12), implies: $E_\chi + E_{\bar{\chi}} = E_X + E_{\bar{X}}$, so that:

$$f_X f_{\bar{X}} = e^{-(E_X + E_{\bar{X}})/T} = e^{-(E_\chi + E_{\bar{\chi}})/T} = f_\chi^{\text{eq}} f_{\bar{\chi}}^{\text{eq}} \quad (10.15)$$

where $f_\chi^{\text{eq}} = e^{-E_\chi/T}$ for a species in thermal equilibrium (c.f. Appendix B). Therefore,

$$f_\chi f_{\bar{\chi}} - f_X f_{\bar{X}} = f_\chi f_{\bar{\chi}} - f_\chi^{\text{eq}} f_{\bar{\chi}}^{\text{eq}} \quad (10.16)$$

With these in mind, we can now write the interaction term in the Boltzmann equation in the form appearing in (10.1) above:

$$\begin{aligned} \frac{dn_\chi}{dt} + 3Hn_\chi &= -\langle \sigma_{\chi\bar{\chi} \rightarrow X\bar{X}} |v| \rangle (n_\chi^2 - (n_\chi^{\text{eq}})^2) , \\ \sigma_{\chi\bar{\chi} \rightarrow X\bar{X}} |v| &\equiv (n_\chi^{\text{eq}})^{-2} \int d\Pi_\chi d\Pi_{\bar{\chi}} d\Pi_X d\Pi_{\bar{X}} (2\pi)^4 \times \\ &\delta^{(4)}(p_\chi + p_{\bar{\chi}} - p_X - p_{\bar{X}}) |\mathcal{M}|^2 e^{-\frac{E_\chi}{T}} e^{-\frac{E_{\bar{\chi}}}{T}} . \end{aligned} \quad (10.17)$$

The situation can be straightforwardly generalised to the case (which is appropriate for neutralino annihilations, c.f. section 10.5) where the final state F of the annihilation $\chi\bar{\chi} \rightarrow F$ involves more products. In such a case, the only change in (10.17) is the replacement of the thermally averaged cross section $\langle \sigma_{\chi\bar{\chi} \rightarrow X\bar{X}} |v| \rangle$ by the more general one $\langle \sigma_{\chi\bar{\chi} \rightarrow F} |v| \rangle$. Summing over all annihilation channels, then, yields the Boltzmann equation (10.1), with the final annihilation cross section entering the collision term being denoted by $\langle \sigma_A |v \rangle$.

Solving the Boltzmann equation and calculating the thermal relic abundance

In this part of this section we shall outline a method of solving the Boltzmann equation (10.1), thus calculating the thermal dark matter relic abundance of the single dominant DM species χ . More complicated situations, where one may have more than one species in the Boltzmann equation, will not be discussed here.

To this end, we first notice that, upon exploiting the fact that the entropy of an Einstein Universe is constant (8.81), one may construct the entropy density s , for which $sa^3 = \text{constant}$, with $a(t)$ the scale factor of the Universe. Then, one changes variables in the Boltzmann equation, from the number density of species n_χ to:

$$Y \equiv \frac{n_\chi}{s} \quad (10.18)$$

in terms of which the left-hand-side of the Boltzmann equation reads:

$$\dot{n}_\chi + 3Hn_\chi = s\dot{Y} \quad (10.19)$$

⁷We shall not discuss, for the sake of brevity and lack of time, unstable species cases here. The interested reader can find the relevant details in [1].

Exercise 10.1 Prove eq. (10.19).

It is also convenient to pass from an equation with respect to cosmic time variations to that of temperature variations, given that the collision term depends explicitly on temperature, as we have seen above. To this end, we need to know what is the relation between the cosmic time and the temperature of our Universe. This is a model dependent question, and it depends crucially on the epoch of the Universe we are considering.

For instance, in a standard Robertson-Walker-Einstein Universe, we are restricting ourselves here, during the radiation era, one has the following relation between temperature and the scale factor, as we have discussed in previous sections (8.73) [1]:

$$t = 0.30g_\star^{-1/2} \frac{M_{\text{Pl}}}{T^2} \quad (10.20)$$

where g_\star the effective number of degrees of freedom of the species χ and M_{Pl} is the four-dimensional Planck mass ($M_{\text{Pl}} \sim 10^{19}$ GeV). The reader is referred to Appendix B for a better explanation of this formula.

It is also convenient to pass into dimensionless variables,

$$x \equiv m_\chi/T \quad (10.21)$$

where m_χ the mass of the species (or in general some other convenient mass scale).

Thus, the Boltzmann equation can then be re-written as:

$$\begin{aligned} \frac{dY}{dx} = & -\frac{x}{H(m_\chi)s} \int d\Pi_\chi d\Pi_{I_1} d\Pi_{I_2} \dots d\Pi_{F_1} d\Pi_{F_2} \dots \times |\mathcal{M}|^2 \times \\ & (2\pi)^4 \delta^{(4)}(p_\chi + p_{I_1} + p_{I_2} + \dots - p_{F_1} - p_{F_2} - \dots) (f_{I_1} f_{I_2} \dots f_\chi - f_{F_1} f_{F_2} \dots) . \end{aligned} \quad (10.22)$$

where $H(m_\chi) = 1.67g_\star^{1/2} m_\chi^2/M_{\text{Pl}}$, with

$$H(x) = H(m_\chi)x^{-2} . \quad (10.23)$$

In this form, the Boltzmann equation is easier to solve and yield the required relic abundance.

In the case of stable species, in terms of these variables the Boltzmann equation (10.17) becomes (for the more general case where one sums up over annihilation channels):

$$\frac{dY}{dx} = -\frac{x\langle\sigma_A|v\rangle s}{H(m_\chi)}(Y^2 - Y_{\text{eq}}^2) , \quad (10.24)$$

in an obvious notation. Recalling (10.23), we can then re-write (10.24) in the following suggestive form:

$$\frac{x}{Y_{\text{eq}}} \frac{dY}{dx} = -\frac{\Gamma_A}{H(T)} \left(\frac{Y^2}{Y_{\text{eq}}^2} - 1 \right) , \quad (10.25)$$

with $\Gamma_A \equiv n_{\text{eq}}\langle\sigma_A|v\rangle$ denoting the total annihilation interaction rate. . This form expresses the well-known fact that the change of the number density of (stable) dark matter species per co-moving volume in an expanding Universe is controlled by the ratio Γ/H , which indicates the *effectiveness of annihilations* of the stable species, times a measure from deviation from equilibrium. When the ratio $\Gamma/H \lesssim 1$ the relative change of the number of species χ in a co-moving volume becomes relatively small and the species are said to decouple, or *freeze in* (equivalently the annihilations *freeze out*). As the Universe expands, originally $\Gamma \gtrsim H$, but eventually the expansion wins over the scattering rate, since the relative distance among the species overcomes significantly the scattering length, and thus $\Gamma < H$, and the species freeze out. Their numbers then remain constant from that moment until the present era. The *relic abundance* of the species χ is given by the value of the corresponding energy density today

$$(\rho_\chi)_0 = s_0 Y(x \rightarrow \infty) m_\chi , \quad \text{or equivalently } \Omega_\chi h^2 , \quad (10.26)$$

with h the reduced Hubble constant, defined in (8.26).

It is this *relic abundance* that the Boltzmann equation helps evaluating, by means of solving it, after integrating it from the freezeout point x_f till today, where to a very good approximation we may assume the temperature to approach zero, $x \rightarrow \infty$ (actually one can use the CMB temperature of $\mathcal{O}(1)$ K as the end point of integration, in numerical solutions). It must be noted at this point that the Boltzmann equation is a particular form of a Ricatti equation, and there are *no general solutions known* in closed form. It is mostly solved by means of *approximate* or *numerical* methods, although in some limiting cases one can obtain analytic forms of the solution. For details we refer the interested reader to the literature, see refs. [1] and references therein.

Below we shall sketch an *approximate* solution [1] of the Boltzmann, equation, which however allows for analytic solution, as well as an understanding of the basic steps in such a calculation. To this end, let us first write the equation in terms of the deviation from the (thermal) equilibrium, by defining:

$$\Delta \equiv Y - Y_{\text{eq}} . \quad (10.27)$$

We also take into account that the total annihilation cross section can be expanded as (c.f. (10.68) in sub-section 10.6):

$$\sigma_A |v| \sim v^p$$

where $p = 0$ for s -wave annihilators, and $p = 2$ for p -wave annihilators. These two types are usually the dominant ones in such an expansion, and in this course we restrict our attention on them. Taking into account that temperature is proportional to the average kinetic energy of a particle, that is $\langle v^2 \rangle \sim T$, we may write : ⁸

$$\langle \sigma_A |v| \rangle = \sigma_0 \left(\frac{T}{m_\chi} \right)^n \equiv \sigma_0 x^{-n} , \quad n = 0 \text{ (1) for } s \text{ (p) - annihilators} . \quad (10.28)$$

We concentrate on the case of non-relativistic, massive, species, for which (c.f. Eq. (12.23), and relevant discussion, in Appendix B):

$$Y_{\text{eq}} = 0.145 \left(\frac{g}{g_{\star S}} \right) x^{3/2} e^{-x} \quad (10.29)$$

where g is the number of internal d.o.f. of the species χ whose relic we want to calculate.

From (10.27), (10.28) and (10.29), upon taking into account that $Y^2 - Y_{\text{eq}}^2 = (Y - Y_{\text{eq}})(Y + Y_{\text{eq}}) = \Delta(\Delta + 2Y_{\text{eq}})$, the Boltzmann equation (10.24) can be written in terms of the difference Δ as :

$$\Delta' = -Y'_{\text{eq}} - \lambda x^{-n-2} \Delta (\Delta + 2Y_{\text{eq}}) , \quad (10.30)$$

where the prime denotes differentiation w.r.t. x , and $\lambda \equiv \frac{x \langle \sigma_A |v| \rangle s}{H(m_\chi)} |_{x=1} = 0.264 \left(\frac{g_{\star S}}{g_{\star}^{1/2}} \right) M_{\text{Pl}} m_\chi \sigma_0$, where we took into account the formulae for the entropy density at equilibrium (8.83) (c.f. also (12.22) and relevant discussion in Appendix B).

To solve (10.30) analytically, we make the approximation that above the freeze-out temperature $T \gg T_f$, such that $1 < x \ll x_f \equiv m_\chi/T_f$, *i.e.* at *early times* of the Universe, $Y \sim Y_{\text{eq}}$. Thus, in such a regime of temperatures, $\Delta' \simeq 0$, Δ is small, $\Delta \ll Y_{\text{eq}}$, and thus (10.30) can be solved immediately:

$$\text{For } T \gg T_f : \quad \Delta \simeq -\lambda^{-1} x^{n+2} \frac{Y'_{\text{eq}}}{2Y_{\text{eq}} + \Delta} \simeq \frac{x^{n+2}}{2\lambda} , \quad (10.31)$$

⁸In some cases, both s - and p -wave annihilators are simultaneously present and of comparable strength. Then, the total annihilation cross section may be parametrised as [1]: $\langle \sigma_A |v| \rangle = \sigma_0 x^{-n} (1 + b x^{-m})$. The modifications to the relic density induced by such a case are straightforward to compute, but we shall not give them here.

where in the last equality we took into account (10.29), as well as that $\Delta \ll Y_{\text{eq}}$, and $1 < x \ll x_f \equiv m_\chi/T_f$, thus keeping only dominant terms in x , *i.e.*

$$\frac{Y'_{\text{eq}}}{2Y_{\text{eq}} + \Delta} \simeq -1/2 + \mathcal{O}\left(x^{-1/2}e^{-x}\right), \quad 1 < x \ll x_f. \quad (10.32)$$

At late times, $x \gg x_f$, the quantity $Y \gg Y_{\text{eq}}$, hence to a good approximation we may assume:

$$\Delta \simeq Y, \quad \text{For : } T \ll T_f \quad (10.33)$$

In this regime, the terms involving Y'_{eq} and Y_{eq} can safely neglected in (10.30), which now reads:

$$\Delta' \simeq -\lambda x^{-n-2} \Delta^2, \quad (10.34)$$

This can be integrated straightforwardly from the freeze-out point x_f until today $x_0 \equiv m_\chi/T_0$. Today, the temperature of the Universe is that of the CMB, $T_0 \sim 2.7\text{K}$. To a very good approximation, then, we may consider this temperature to be sufficiently small, as compared to the mass of m_χ (which, if it is a supersymmetric partner should have masses at least of order of a few hundreds of GeV usually, c.f. sub-sections 11.1, 11.1.6 below), such that $x_0 \rightarrow \infty$ to a very good approximation. Hence, the integration over x should be extended from x_f till $x \rightarrow \infty$ and thus one can derive an expression for the relic density today in terms of the freeze-out temperature x_f :

$$Y_\infty = \Delta_\infty = \frac{n+1}{\lambda} x_f^{n+1}. \quad (10.35)$$

To complete our task and evaluate (10.35) we need to provide an estimate of the freeze-out temperature x_f of the species χ . This can be done approximately, if we observe that at freeze-out, Δ becomes of order Y_{eq} . A good approximation, therefore, is to set [1]:

$$\Delta(x_f) = cY_{\text{eq}}(x_f) \quad (10.36)$$

where c a numerical constant of order unity, which can be determined by a *best-fit procedure*, in order to get satisfactory agreement between the above-described *analytic (approximate)* result and the *numerical solution* of the Boltzmann equation (10.1).

Matching the solutions for early and late times at the freeze-out point $x = x_f$, we then obtain from (10.36) and (10.31) at $x = x_f$, on account of the approximation (10.32):

$$cY_{\text{eq}}(x_f) = \Delta(x_f) \simeq \frac{x_f^{n+2}}{\lambda(2+c)}. \quad (10.37)$$

From (10.29) then, we can solve for x_f to obtain:

$$x_f \simeq \ln\left(0.145\left(\frac{g}{g_{\star S}}\right)[2+c]c\lambda\right) - (n+1/2)\ln\left(\ln\left(0.145\left(\frac{g}{g_{\star S}}\right)[2+c]c\lambda\right)\right). \quad (10.38)$$

The best fit value of c , which yields better than 5% agreement with the numerical solution is [1]:

$$c(c+2) = n+1,$$

which yields for the freeze-out temperature:

$$x_f = \ln\left(0.038(n+1)(g/g_{\star S}^{1/2})M_{\text{Pl}}m_\chi\sigma_0\right) - (n+\frac{1}{2})\ln\left(\ln\left(0.038(n+1)(g/g_{\star S}^{1/2})M_{\text{Pl}}m_\chi\sigma_0\right)\right). \quad (10.39)$$

From this, the relic density of the species χ is evaluated by means of (10.35):

$$Y_\infty = \frac{3.79(n+1)x_f^{n+1}}{(g_{\star S}/g_{\star}^{1/2})M_{\text{Pl}}m_\chi\sigma_0} = \frac{3.79(n+1)x_f(g_{\star S}/g_{\star}^{1/2})}{M_{\text{Pl}}m_\chi\langle\sigma_A|v\rangle}. \quad (10.40)$$

from which the present-epoch number density n_0 and mass densities ρ_0 of the species χ can be readily evaluated:

$$n_{\chi,0} = s_0 Y_\infty = 1.13 \times 10^4 \frac{(n+1)x_f^{n+1}}{(g_{*S}/g_{*}^{1/2}) M_{\text{Pl}} m_\chi \sigma_0} \text{ cm}^{-3},$$

$$\boxed{\Omega_\chi h^2 = 1.07 \times 10^9 \frac{(n+1)x_f^{n+1} \text{ GeV}^{-1}}{(g_{*S}/g_{*}^{1/2}) M_{\text{Pl}} m_\chi \sigma_0}}. \quad (10.41)$$

Some important remarks are now in order. The relic abundance (10.40),(10.41) depends on the mass m_χ of the dark matter species and is *inversely proportional* to the total annihilation *cross section* σ_0 . bounded by astrophysical and other measurements, by means of relations of the form

$$A \leq \Omega_\chi h^2 \leq B \quad (10.42)$$

which, in turn, imply corresponding *allowed ranges* of m_χ (through the dependence of the freeze-out temperature on m_χ (c.f. (10.39)). The reader should also notice that the quantity $\Omega_\chi h^2$ is independent of the present-epoch value of Hubble parameter, H_0 , by construction.

If the dark matter, then, is attributed to supersymmetric particles, as in the above example, the corresponding allowed ranges of m_χ , are translated into *cosmologically allowed regions* in the parameter space of the model, which however are highly model dependent, as they depend crucially on the details of the microscopic model being tested. As we have seen above, the mere presence of the source in the Boltzmann equation (10.43), which is model dependent, does affect the relic abundance, as becomes clear by comparing (10.62) with the standard cosmology (no-source) result (10.63). This is the basic philosophy of such collider dark-matter searches, and in what follows we shall discuss briefly some very simple but indicative models, namely (i) the minimal supersymmetric standard model within the standard cosmology (for a review see [3]), and (ii) a string model, in which dark matter is coupled to the dilaton field ϕ , leading to non-trivial source terms in the Boltzmann equation (10.43), of the form $\Gamma = \dot{\phi}$ [28]. In this respect, we now discuss the respective modifications in the Boltzmann equation and the associated solutions induced by such source term, with the point of demonstrating the underlying model dependence of the thermal relic abundance.

Advanced topic: Modifications to Boltzmann equation due to external sources

As we have mentioned previously, there are theoretical models in which there are modifications in the Boltzmann equation, due to the appearance of extra sources, which may come for instance from the coupling of dark matter to scalar fields in the gravitational sector of string theories, or may represent some non equilibrium off-shell effects, remnants from an early-Universe catastrophic event [28]. We shall explicitly discuss one such model in section 11.2 below. In what follows, therefore, we shall consider a more general equation than (10.1), by including a source Γ . The conventional cosmology Boltzmann equation, then, corresponds to the case $\Gamma = 0$. The more general equation for the number density of species n , which we shall solve along the lines of our previous discussion for the conventional case, reads:

$$\frac{dn}{dt} = -3 \frac{\dot{a}}{a} n - \langle v\sigma \rangle (n^2 - n_{eq}^2) + \Gamma n \quad (10.43)$$

As discussed previously, before the decoupling time t_f , $t < t_f$, equilibrium is maintained and thus $n = n_{eq}$ for such an era. However, it is crucial to observe that, as a result of the presence of the source Γ terms, n_{eq} no longer scales with the inverse of the cubic power of the expansion radius a , which was the case in conventional (on-shell) cosmological models.

To understand this, let us assume that $n = n_{eq}^{(0)}$ at a very early epoch t_0 . Then the solution of the modified Boltzmann equations at all times $t < t_f$ is given by

$$n_{eq} a^3 = n_{eq}^{(0)} a^3(t_0) \exp \left(\int_{t_0}^t \Gamma dt \right). \quad (10.44)$$

The time t_0 characterizes a very early time, which is not unreasonable to assume that it signals the exit from the inflationary period. Soon after the exit from inflation, all particles are in thermal equilibrium, for all times $t < t_f$, with the source term modifying the usual Boltzmann distributions in the way indicated in Eq. (10.44) above. It has been tacitly assumed that the entropy is conserved despite the presence of the source. This is a good *approximation*, given that the entropy increase is most significant during the inflationary era of the Universe, and hence it is not inconsistent to assume that, for all practical purposes, sufficient for our phenomenological analysis in this work, there is no significant entropy production after the exit from inflation. This is a necessary ingredient for our approach, since without such an assumption no predictions can be made, even in the conventional cosmological scenarios. Thus, the picture we envisage is that at t_0 the Universe entered an equilibrium phase, the entropy is conserved to a good approximation, and hence all particle species find themselves in thermal equilibrium, despite the presence of the Γ source, which *slowly* pumps in or sucks out energy, without, however, disturbing the particles' thermal equilibrium.

From the above discussion it becomes evident that it is of paramount importance to know the behaviour of the source term at all times, in order to extract information for the relic abundances, especially those concerning Dark Matter, and how these are modified from those of the standard Cosmology. Before embarking on such a task and study the phenomenological consequences of particular models predicting the existence of Dark Matter, especially Supersymmetry-based ones, we must first proceed in a general way to set up the stage and discuss how the Boltzmann equation is solved in general, as well as how the relic density is affected by the presence of the non-conventional source terms present in (10.43).

For the sake of brevity, we shall not deploy all the details of the derivation of the relic density, as these parallel the conventional-cosmology case, studied above. Instead, we shall demonstrate the most important features and results of our approach, paying particular attention to exhibiting the differences from the conventional case. Generalizing the standard techniques [1], mentioned above, we assume that above the freeze-out point the density is the equilibrium density as provided by Eq. (10.44), while below this the interaction terms starts becoming unimportant.

Following [28] we define

$$x \equiv T/m_{\tilde{\chi}} , \quad (10.45)$$

and restrict the discussion on a particular species $\tilde{\chi}$ of mass $m_{\tilde{\chi}}$, which eventually may play the role of the dominant Dark Matter candidate. Notice that this definition of x is inverse to the definition of x (10.21), used in the conventional cosmology treatments of the Boltzmann equation (10.1). As we shall see, this is done for mere convenience in calculating the source term. Clearly one can pass from one definition to the other trivially. In what follows in this section, however, we shall use (10.45), which should be remembered, when one compares the results with the conventional cosmology. It goes without saying that the final results for the relic densities in both approaches are expressed unambiguously in terms of T .

As in the conventional-Cosmology treatments, it also proves convenient to trade the number density n for the quantity $Y \equiv n/s$, that is the number per entropy density [1]. The equation for Y , derived from (10.43), is given by:

$$\frac{dY}{dx} = m_{\tilde{\chi}} \langle v\sigma \rangle \left(\frac{45}{\pi} G_N \tilde{g}_{eff} \right)^{-1/2} \left(h + \frac{x}{3} \frac{dh}{dx} \right) (Y^2 - Y_{eq}^2) - \frac{\Gamma}{Hx} \left(1 + \frac{x}{3h} \frac{dh}{dx} \right) Y \quad . \quad (10.46)$$

where $G_N = 1/M_{\text{Pl}}^2$ is the four-dimensional gravitational constant, the quantity H is the Hubble expansion rate, h denote the entropy degrees of freedom, and $\langle v\sigma \rangle$ is the thermal average of the relative velocity times the annihilation cross section and \tilde{g}_{eff} is simply defined by the relation [1]

$$\varrho + \Delta\varrho \equiv \frac{\pi^2}{30} T^4 \tilde{g}_{eff} \quad . \quad (10.47)$$

The reader should notice at this point that $\Delta\varrho$ incorporates the effects of the additional contributions due to the dissipative source, which are not accounted for in the g_{eff} of conventional

Cosmology [1], hence the notation \tilde{g}_{eff} . We next remark that ρ , as well as $\Delta\rho$, as functions of time are known, once one solves the cosmological equations. However, only the degrees of freedom involved in ρ are thermal, the rest, like the cosmological-constant term if present in a model, are included in $\Delta\rho$. Therefore, the relation between temperature and time is provided by

$$\rho = \frac{\pi^2}{30} T^4 g_{eff}(T) \quad (10.48)$$

while $\rho + \Delta\rho$ are involved in the evolution through the modified Friedmann equation, we assume for our case here

$$H^2 = \frac{8\pi G_N}{3} (\rho + \Delta\rho) \quad (10.49)$$

where the terms $\Delta\rho$ are responsible for the presence of the source Γ in (10.43). In the simplest case of scalar fields in the gravitational multiplet of string theory (the so-called dilatons, spin-zero particle excitations of the string spectrum) the terms $\Delta\rho$ contain the contributions of such dilaton fields to the relevant generalizations of Einstein's equations (i.e. equations stemming from the variation with respect to the gravitational field in the model).

It is important for the reader to bear in mind that $\Delta\rho$ contributes to the dynamical expansion, through Eq. (10.49), but not to the thermal evolution of the Universe. The quantity \tilde{g}_{eff} , defined in (10.47), is therefore given by [28]

$$\tilde{g}_{eff} = g_{eff} + \frac{30}{\pi^2} T^{-4} \Delta\rho \quad (10.50)$$

The meaning of the above expression is that time has been replaced by temperature, through Eq. (10.48), after solving the dynamical equations. In terms of \tilde{g}_{eff} the expansion rate H is written as

$$H^2 = \frac{4\pi^3 G_N}{45} T^4 \tilde{g}_{eff} \quad (10.51)$$

This is used in the Boltzmann equation for Y and the conversion from the time variable t to temperature or, equivalently, the variable x .

For x above the freezing point x_f , $Y \approx Y_{eq}$ and, upon omitting the contributions of the derivative terms dh/dx , an approximation which is also adopted in the standard cosmological treatments [1], we obtain for the solution of (10.46)

$$Y_{eq} = Y_{eq}^{(0)} \exp\left(-\int_x^\infty \frac{\Gamma H^{-1}}{x} dx\right) \quad (10.52)$$

Here, $Y_{eq}^{(0)}$ corresponds to $n_{eq}^{(0)}$ and in the non-relativistic limit is given by

$$Y_{eq}^{(0)} = \frac{45}{2\pi^2} \frac{g_s}{h} (2\pi x)^{-3/2} \exp(-1/x) \quad (10.53)$$

where g_s counts the particle's spin degrees of freedom.

In the regime $x < x_f$, $Y \gg Y_{eq}^{(0)}$ the equation (10.46) can be written as

$$\frac{d}{dx} \frac{1}{Y} = -m_{\tilde{\chi}} \langle v\sigma \rangle \left(\frac{45}{\pi} G_N \tilde{g}_{eff}\right)^{-\frac{1}{2}} h + \frac{\Gamma H^{-1}}{xY} \quad (10.54)$$

Applying (10.54) at the freezing point x_f and using (10.52) and (10.53), leads, after a straightforward calculation, to the determination of $x_f = T_f/m_{\tilde{\chi}}$ through

$$x_f^{-1} = \ln \left[0.03824 g_s \frac{M_{\text{Pl}} m_{\tilde{\chi}}}{\sqrt{g_*}} x_f^{1/2} \langle v\sigma \rangle_f \right] + \frac{1}{2} \ln \left(\frac{g_*}{\tilde{g}_*} \right) + \int_{x_f}^{x_{in}} \frac{\Gamma H^{-1}}{x} dx \quad (10.55)$$

As usual, all quantities are expressed in terms of the dimensionless $x \equiv T/m_{\tilde{\chi}}$ and x_{in} corresponds to the time t_0 discussed previously, taken to represent the exit from the inflationary period of the Universe.

The first term on the right-hand-side of (10.55) is that of a conventional Cosmology for, say, an LSP carrying g_s spin degrees of freedom, playing the rôle of the dominant Cold Dark Matter species in our concrete and physically promising example, which we use here. The quantity $\langle v\sigma \rangle_f$ is the thermal average of $v\sigma$ at x_f and g_* is g_{eff} of conventional Cosmology at the freeze-out point. The same notation holds for \tilde{g}_* . In our treatment above, we chose in (10.55) to present x_f in such a way so as to separate the conventional contributions, which reside in the first term, from the contributions of the source, which are contained within the last two terms. The latter induce a shift in the freeze-out temperature. The penultimate term on the right hand side of (10.55), due to its logarithmic nature, does not affect much the freeze-out temperature. The last term, on the other hand, is more important and, depending on its sign, may shift the freeze-out point to earlier or later times.

In order to calculate the relic abundance, we must solve (10.54) from x_f to today's value x_0 , corresponding to a temperature $T_0 \approx 2.7^0 K$, which is the CMB temperature. Following the usual approximations we arrive at the result:

$$Y^{-1}(x_0) = Y^{-1}(x_f) + \left(\frac{\pi}{45}\right)^{\frac{1}{2}} m_{\tilde{\chi}} M_{\text{Pl}} \tilde{g}_*^{-\frac{1}{2}} h(x_0) J - \int_{x_0}^{x_f} \frac{\Gamma H^{-1}}{x Y} dx \quad . \quad (10.56)$$

In conventional Cosmology [1] \tilde{g}_* is replaced by g_* and the last term in (10.56) is absent. The quantity J is $J \equiv \int_{x_0}^{x_f} \langle v\sigma \rangle dx$. By replacing $Y(x_f)$ by its equilibrium value (10.52) the ratio of the first term on the r.h.s. of (10.56) to the second is found to be exactly the same as in the no-source case. Therefore, by the same token as in conventional Cosmology, the first term can be safely omitted, as long as x_f is of order of 1/10 or less. Furthermore, the integral on the r.h.s. of (10.56) can be simplified if one uses the fact that $\langle v\sigma \rangle n$ is small as compared with the expansion rate \dot{a}/a after decoupling. For the purposes of the evaluation of this integral, therefore, this term can be omitted in (10.54), as long as we stay within the decoupling regime, and one obtains:

$$\frac{d}{dx} \frac{1}{Y} = \frac{\Gamma H^{-1}}{x Y} \quad . \quad (10.57)$$

By integration this yields $Y(x) = Y(x_0) \exp(-\int_{x_0}^x \Gamma H^{-1} dx/x)$. Using this inside the integral in (10.56) we get

$$(h(x_0)Y(x_0))^{-1} = \left(1 + \int_{x_0}^{x_f} \frac{\Gamma H^{-1}}{\psi(x)} dx\right)^{-1} \left(\frac{\pi}{45}\right)^{\frac{1}{2}} m_{\tilde{\chi}} M_{\text{Pl}} \tilde{g}_*^{-\frac{1}{2}} J \quad (10.58)$$

where the function $\psi(x)$ is given by $\psi(x) \equiv x \exp(-\int_{x_0}^x \Gamma H^{-1} dx/x)$. With the exception of the prefactor on the r.h.s. of (10.58), this is identical in form to the result derived in standard treatments, if \tilde{g}_* is replaced by g_* and the value of x_f , implicitly involved in the integral J , is replaced by its value found in ordinary treatments in which the dilaton-dynamics and non-critical-string effects are absent.

The matter density of species $\tilde{\chi}$ is then given by

$$\rho_{\tilde{\chi}} = f \left(\frac{4\pi^3}{45}\right)^{1/2} \left(\frac{T_{\tilde{\chi}}}{T_{\gamma}}\right)^3 \frac{T_{\gamma}^3}{M_{\text{Pl}}} \frac{\sqrt{\tilde{g}_*}}{J} \quad (10.59)$$

where the prefactor f is:

$$f = 1 + \int_{x_0}^{x_f} \frac{\Gamma H^{-1}}{\psi(x)}$$

It is important to recall that the thermal degrees of freedom are counted by g_{eff} (c.f. (10.48)), and not \tilde{g}_{eff} , the latter being merely a convenient device connecting the total energy, thermal and

non-thermal, to the temperature T (c.f. (10.47)). Hence,

$$\left(\frac{T_{\tilde{\chi}}}{T_{\gamma}}\right)^3 = \frac{g_{eff}(1MeV)}{g_{eff}(T_{\tilde{\chi}})} \frac{4}{11} = \frac{43}{11} \frac{1}{g_*} . \quad (10.60)$$

In deriving (10.60) only the thermal content of the Universe is used, while the dilaton and the non-critical terms do not participate. Therefore the $\tilde{\chi}$'s matter density is given by

$$\rho_{\tilde{\chi}} = f \left(\frac{4\pi^3}{45}\right)^{1/2} \frac{43}{11} \frac{T_{\gamma}^3}{M_{\text{Pl}} g_*} \frac{\sqrt{\tilde{g}_*}}{J} . \quad (10.61)$$

This formula tacitly assumes that the $\tilde{\chi}$ s decoupled before neutrinos. For the relic abundance, then, we derive the following approximate result

$$\Omega_{\tilde{\chi}} h_0^2 = (\Omega_{\tilde{\chi}} h_0^2)_{no-source} \times \left(\frac{\tilde{g}_*}{g_*}\right)^{1/2} \left(1 + \int_{x_0}^{x_f} \frac{\Gamma H^{-1}}{\psi(x)} dx\right) . \quad (10.62)$$

The quantity referred to as *no-source* is the well known no-source expression (10.41),

$$(\Omega_{\tilde{\chi}} h_0^2)_{no-source} = \frac{1.066 \times 10^9 \text{ GeV}^{-1}}{M_{\text{Pl}} \sqrt{g_*} J} \quad (10.63)$$

where $J \equiv \int_{x_0}^{x_f} \langle v\sigma \rangle dx$. However, as already remarked, the end point x_f in the integration is the shifted freeze-out point as determined by Eq. (10.55). The merit of casting the relic density in such a form is that it clearly exhibits the effect of the presence of the source. Certainly, if an accurate result is required, one can proceed without approximations and handle the problem numerically as in the standard treatments.

After this discussion on modified Boltzmann equations, which may characterise some non-standard Cosmologies, we are now ready to consider explicit examples of the above-described calculations, in terms of phenomenologically semi-realistic models of Cold Dark Matter (CDM). Before doing so, though, we should first discuss briefly for completeness, how the recent measurements of CMB temperature fluctuations by the WMAP satellite exclude warm and hot forms of Dark Matter.

10.4 Hot and Warm DM Excluded by WMAP

The WMAP/CMB results on the cosmological parameters discussed previously disfavor strongly Hot Dark Matter (neutrinos), as a result of the new determination of the upper bound on neutrino masses. The contribution of neutrinos to the energy density of the Universe depends upon the sum of the mass of the light neutrino species [1, 9]:

$$\Omega_{\nu} h^2 = \frac{\sum_i m_i}{94.0 \text{ eV}} \quad (10.64)$$

where the sum includes neutrino species that are light enough to decouple while still relativistic.

The combined results from WMAP and other experiments [9] on the cumulative likelihood of data as a function of the energy density in neutrinos lead to $\Omega_{\nu} h^2 < 0.0067$ (at 95% confidence limit). Adding the Lyman α data, the limit weakens slightly [9]: $\Omega_{\nu} h^2 < 0.0076$ or equivalently (from (10.64)): $\sum_i m_{\nu_i} < 0.69 \text{ eV}$, where, we repeat again, the sum includes light species of neutrinos. This may then imply an average upper limit on electron neutrino mass $< m_{\nu} >_e < 0.23 \text{ eV}$. These upper bounds strongly disfavor Hot Dark Matter scenarios.

Caution should be exercised, however, when interpreting the above WMAP result. There is the underlying-theoretical-model dependence of these results, which stems from the assumption of an Einstein-FRW Cosmology, characterized by local Lorentz invariance. If Lorentz symmetry is violated, as, for instance, is the case of the TeVeS models alternative to DM, then neutrinos

with (rest) masses of up to 2 eV could have an abundance of $\Omega_\nu \sim 0.15$ in order to reproduce the peaks in the observed CMB spectrum (fig. 37) [21] and thus being phenomenologically acceptable, at least from the CMB measurements viewpoint.

At this juncture we note that another important result of WMAP is the evidence for early re-ionization of the Universe at redshifts $z \simeq 20$. If one assumes that structure formation is responsible for re-ionization, then such early re-ionization periods are compatible only for high values of the masses m_X of *Warm Dark Matter*. Specifically, one can exclude models with $m_X \leq 10$ KeV based on numerical simulations of structure formation for such models [29]. Such simulations imply that dominant structure formation responsible for re-ionization, for Warm Dark Matter candidates with $m_X \leq 10$ KeV, occurs at much smaller z than those observed by WMAP. In view of this, one can therefore exclude popular particle physics models employing light gravitinos ($m_X \simeq 0.5$ KeV) as the Warm Dark Matter candidate. It should be noted at this stage that such structure formation arguments can only place a lower bound on the mass of the Warm Dark Matter candidate. The reader should bear in mind that Warm Dark Matter with masses $m_X \geq 100$ KeV becomes indistinguishable from Cold Dark Matter, as far as structure formation is concerned.

10.5 Cold DM in Supersymmetric Models: Neutralino

After the exclusion of Hot and Warm Dark Matter, the only type of Dark matter that remains consistent with the recent WMAP results [9] is the *Cold Dark Matter*, which in general may consist of axions, superheavy particles (with masses $\sim 10^{14\pm 5}$ GeV) [30, 31] and stable supersymmetric partners. Indeed, one of the major and rather unexpected predictions of Supersymmetry (SUSY), broken at low energies $M_{SUSY} \approx \mathcal{O}(1\text{TeV})$, while R -parity is conserved, is the existence of a stable, neutral particle, the lightest neutralino ($\tilde{\chi}$), referred to as the lightest supersymmetric particle (LSP) [27]. Such particle is an ideal candidate for the Cold Dark Matter in the Universe [27]. Such a prediction fits well with the fact that SUSY is not only indispensable in constructing consistent string theories, but it also seems unavoidable at low energies (~ 1 TeV) if the gauge hierarchy problem is to be resolved. Such a resolution provides a measure of the SUSY breaking scale $M_{SUSY} \approx \mathcal{O}(1\text{TeV})$.

This type of Cold Dark Matter will be our focus from now on, in association with the recent results from WMAP3 on relic densities [6, 9]. The WMAP3 results, combined with other existing data, yield for the baryon and matter densities (including dark matter) at 2σ -level: $\Omega_m h^2 = 0.1268^{+0.0072}_{-0.0095}$ (matter), $100\Omega_b h^2 = 2.233^{+0.072}_{-0.091}$ (baryons). One assumes that CDM is given by the difference of these two. As mentioned already, in supersymmetric (SUSY) theories the favorite candidate for CDM is the lightest of the Neutralinos $\tilde{\chi}$ (SUSY CDM), which is stable, as being the Lightest SUSY particle (LSP) (There are cases where the stau or the sneutrino can be the lightest supersymmetric particles. These cases are not favored [32] and hence are not considered). From the WMAP3 results [6], then, assuming $\Omega_{CDM} \simeq \Omega_\chi$, we can infer stringent limits for the neutralino χ relic density:

$$0.0950 < \Omega_\chi h^2 < 0.1117, \quad (10.65)$$

It is important to notice that in this inequality only the upper limit is *rigorous*. The lower Limit is *optional*, given that there might (and probably do) exist other contributions to the overall (dark) matter density. It is imperative to notice that all the constraints we shall discuss in this review are *highly model dependent*. The results on the minimal SUSY extensions of the standard model [3], for instance, cannot apply to other models such as superstring-inspired ones, including non equilibrium cosmologies, which we shall also discuss here. However, formally at least, most of the analysis can be extrapolated to such models, with possibly different results, provided the SUSY dark matter in such models is thermal. Before moving into such a discussion we consider it as instructive to describe briefly various important properties of the Neutralino DM.

The Neutralino is a superposition of SUSY partner states. Its mass matrix in bino–wino–

higgsinos basis $\psi_j^0 = (-i\lambda', -i\lambda^3, \psi_{H_1}^0, \psi_{H_2}^0)$ is given by

$$\mathcal{M}_N = \begin{pmatrix} M_1 & 0 & -m_Z s_W c_\beta & m_Z s_W s_\beta \\ 0 & M_2 & m_Z c_W c_\beta & -m_Z c_W s_\beta \\ -m_Z s_W c_\beta & m_Z c_W c_\beta & 0 & -\mu \\ m_Z s_W s_\beta & -m_Z c_W s_\beta & -\mu & 0 \end{pmatrix}$$

where M_1, M_2 : the U(1) and SU(2) gaugino masses, μ : higgsino mass parameter, $s_W = \sin \theta_W$, $c_W = \cos \theta_W$, $s_\beta = \sin \beta$, $c_\beta = \cos \beta$ and $\tan \beta = v_2/v_1$ ($v_{1,2}$ v.e.v. of Higgs fields $H_{1,2}$). The mass matrix is diagonalized by a unitary mixing matrix N , $N^* \mathcal{M}_N N^\dagger = \text{diag}(m_{\tilde{\chi}_1^0}, m_{\tilde{\chi}_2^0}, m_{\tilde{\chi}_3^0}, m_{\tilde{\chi}_4^0})$, where $m_{\tilde{\chi}_i^0}$, $i = 1, \dots, 4$, are the (non-negative) masses of the physical neutralino states with $m_{\tilde{\chi}_1^0} < \dots < m_{\tilde{\chi}_4^0}$. The lightest neutralino is then:

$$\tilde{\chi}_1^0 = N_{11} \tilde{B} + N_{12} \tilde{W} + N_{13} \tilde{H}_1 + N_{14} \tilde{H}_2.$$

To calculate relic densities it is assumed that the initial number density of neutralinos χ particle in the Early Universe has been in thermal equilibrium: interactions creating χ usually happen as frequently as the reverse interactions which destroy them. Eventually the temperature of the expanding Universe drops to the order of the neutralino (rest) mass $T \simeq m_\chi$. In such a situation, most particles no longer have sufficient energy to create neutralinos. Now neutralinos can only *annihilate*, until their rate becomes smaller than the Hubble expansion rate, $H \geq \Gamma_{\text{ann}}$. Then, Neutralinos are being separated apart from each other too quickly to maintain equilibrium, and thus they reach their freeze-out temperature, $T_F \simeq m_\chi/20$, which characterizes this type of cold dark matter.

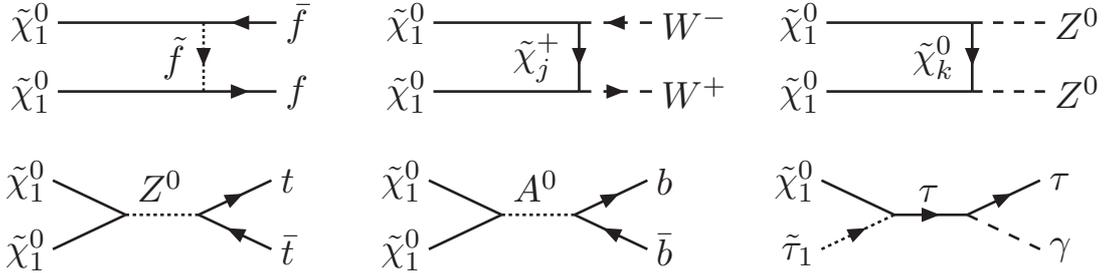


Figure 43: Basic Neutralino annihilations including stau co-annihilations in MSSM (from S. Kraml, Pramana 67, 597 (2006) [hep-ph/0607270]).

In most neutralino relic density calculations, the only interaction cross sections that need to be calculated are annihilations of the type $\chi\chi \rightarrow X$ where χ is the lightest neutralino and X is any final state involving only Standard Model particles. However, there are scenarios in which other particles in the thermal bath have important effects on the evolution of the neutralino relic density. Such a particle annihilates with the neutralino into Standard Model particles and is called a *co-annihilator* (c.f. figure 43). In order for a particle to be an *Effective co-annihilator*, it must have direct interactions with the neutralino and must be *nearly degenerate* in mass: Such degeneracy happens in the Minimal Supersymmetric Standard Model (MSSM), for instance, with possible co-annihilators being the lightest stau, the lightest stop, the second-to-lightest neutralino or the lightest chargino. When this degeneracy occurs, the neutralino and all relevant co-annihilators form a coupled system.

Without co-annihilations the evolution of a relic particle number density, n , is governed, as mentioned previously, by a single-species Boltzmann equation (10.1). It should be noted that the relic-particle number density is modified by the Hubble expansion and by direct and inverse annihilations of the relic particle. The Relic particle is assumed stable, so relic decay is neglected. Also commonly assumed is time-reversal (T) invariance, which relates annihilation and inverse

annihilation processes. In the presence of co-annihilators the Boltzmann equation gets more complicated but it can be simplified using stability properties of relic particle and co-annihilators (using $n = \sum_{i=1}^N n_i$):

$$\frac{dn}{dt} = - 3Hn - \sum_{i,j=1}^N \langle \sigma_{ij} v_{ij} \rangle (n_i n_j - n_i^{\text{eq}} n_j^{\text{eq}}) . \quad (10.66)$$

To a very good approximation, one can use an effective single species Boltzmann equation for this case if $\langle \sigma v \rangle = \sum_{i,j} \langle \sigma_{ij} v_{ij} \rangle \frac{n_i^{\text{eq}} n_j^{\text{eq}}}{n^{\text{eq}} n^{\text{eq}}}$.

The Boltzmann equation (10.66) can be solved numerically, but in most cases even analytically. Details on how to solve the Boltzmann equation are given abundantly in the cosmology literature [1] and will not be repeated here. We shall only outline the most important results that will be essential for our discussion in these lectures. One should determine the freeze-out temperature $x_F = m_\chi/T_F$: $x_F = \ln \left(\frac{0.038 g m_{\text{pl}} m_\chi \langle \sigma v \rangle}{\sqrt{g_*} x_F} \right)$, with m_{pl} the Planck mass, g the total number of degrees of freedom of the χ particle (spin, color, etc.), g_* the total number of effective relativistic degrees of freedom at freeze-out, and the thermally averaged cross section is evaluated at the freeze-out temperature. For most CDM candidates, $x_F \simeq 20$. The total (co)annihilation depletion of neutralino number density is calculated by integrating the thermally averaged cross section from freeze-out to the present temperature:

$$\Omega_\chi h^2 = 40 \sqrt{\frac{\pi}{5}} \frac{h^2}{H_0^2} \frac{s_0}{m_{\text{pl}}^3} \frac{1}{(g_{*S}/g_*^{1/2}) J(x_F)} = \frac{1.07 \times 10^9 \text{ GeV}^{-1}}{g_*^{1/2} m_{\text{pl}} J(x_F)} ,$$

$$J(x_F) = \int_{x_F}^{\infty} \langle \sigma v \rangle x^{-2} dx . \quad (10.67)$$

where s_0 is the entropy density, g_{*S} denotes the number of effective relativistic d.o.f. contributing to the (constant) entropy of the universe and h is the reduced Hubble parameter: $H_0 = 100 h \text{ km sec}^{-1} \text{ Mpc}^{-1}$. This is the expression one compares with the experimental determination of the DM abundance via, e.g., WMAP data. It should be noted at this stage that the theoretical assumptions leading to the above results may not hold in general for all DM models and candidates: the missing non-baryonic matter in the universe may only partially, or not at all, consist of relic neutralinos. Also, as we shall discuss later on the article, in some off-shell, non-equilibrium relaxation stringy models of dark energy, the Boltzmann equation gets modified by off-shell, non-equilibrium terms as well as time-dependent dilaton-source terms. This leads to important modifications on the associated particle-physics models constraints.

10.6 Model-Independent DM Searches in Colliders

As we have discussed above, if dark matter comes from a thermal relic, its density is determined, to a large extent, by the dark matter annihilation cross section: $\sigma(\chi\chi \rightarrow SM SM)$. Indeed, as already mentioned, the present-day dark matter abundance is roughly inversely proportional to the thermally averaged annihilation cross section times velocity, $\Omega_\chi h^2 \propto 1/\langle \sigma v \rangle$. This latter quantity can be conveniently expanded in powers of the relative dark matter particle velocity:

$$\sigma v = \sum_J \sigma_{an}^{(J)} v^{(2J)} . \quad (10.68)$$

Usually, only the lowest order non-negligible power of v dominates. For $J = 0$, such dark matter particles are called s -annihilators, and for $J = 1$, they are called p -annihilators; powers of J larger than 1 are rarely needed.:

Figure 44 shows the constraint on the annihilation cross section as a function of dark matter mass that results from Eq. (10.65) [33]. The lower (upper) band of fig. 44 is for models where s -wave (p -wave) annihilation dominates. It is important to notice [33] that the total annihila-

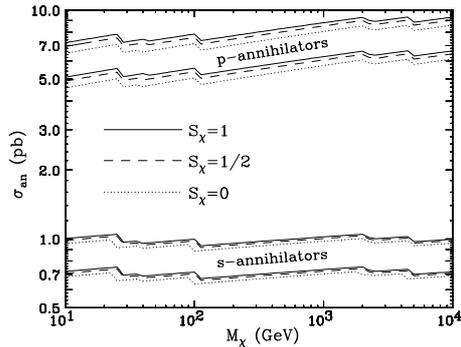


Figure 44: Values of the quantity σ_{an} allowed at 2σ level as a function of the DM mass.

tion cross section σ_{an} is virtually insensitive to dark-matter mass. This latter effect is due to the changing number of degrees of freedom at the time of freeze-out as the dark matter mass changes. It also points to cross sections expected from weak-scale interactions (around $0.8 pb$ for s -annihilators and $6 pb$ for p -annihilators), hence implying the possibility that DM is connected to an explanation for the weak scale and thus WIMPs [33]. Such WIMPs exist not only in supersymmetric theories, of course, but in a plethora of other models such as theories involving extra dimensions and 'little Higgs' models. The LHC and the ILC are specifically designed to probe the origin of the weak scale, so dark matter searches and future collider physics appear to be closely related. The next question one could ask is whether the above cross section could be

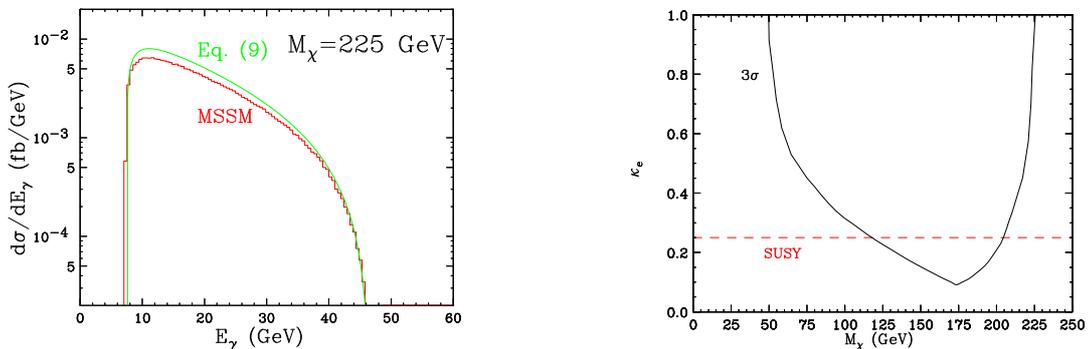


Figure 45: *Left panel:* Comparison between the photon spectra from the process $e^+e^- \rightarrow 2\chi_1^0 + \gamma$ in the explicit supersymmetric models defined in A. Birkedal, K. Matchev and M. Perelstein, Phys. Rev. D **70**, 077701 (2004) (red/dark-gray) and the predicted spectra for a p -annihilator of the corresponding mass and κ_e (green/light-gray). *Right panel:* The reach of a 500 GeV unpolarized electron-positron collider with an integrated luminosity of 500 fb^{-1} for the discovery of p -annihilator WIMPs, as a function of the WIMP mass M_χ and the e^+e^- annihilation fraction κ_e . The 3σ (black) contour is shown, along with an indication of values one might expect from supersymmetric models (red dashed line, labelled 'SUSY'). Only statistical uncertainty is included.

turned, within a WIMP working hypothesis framework, into a model-independent signature at colliders. This question was answered in the affirmative in [33]. One introduces the parameter $\kappa_e \equiv \sigma(\chi\chi \rightarrow e^+e^-)/\sigma(\chi\chi \rightarrow SM|SM)$ which relates dark matter annihilation processes to cross sections involving e^+e^- in the final state. Using crossing symmetries to relate $\sigma(\chi\chi \rightarrow e^+e^-)$ to $\sigma(e^+e^- \rightarrow \chi\chi)$ and co-linear factorization one can relate $\sigma(e^+e^- \rightarrow \chi\chi)$ to $\sigma(e^+e^- \rightarrow \chi\chi\gamma)$, thus connecting astrophysical data on σ_{an} to the process $e^+e^- \rightarrow \chi\chi\gamma$. The resulting differential cross

section reads [33]

$$\frac{d\sigma}{dx d\cos\theta}(e^+e^- \rightarrow 2\chi + \gamma) \simeq \frac{\alpha}{16\pi} \frac{\kappa_e \sigma_{\text{an}}}{x} \frac{1 + (1-x)^2}{\sin^2\theta} 2^{J_0} (2S_\chi + 1)^2 \left(1 - \frac{4M_\chi^2}{(1-x)s}\right)^{\frac{1}{2} + J_0} \quad (10.69)$$

with α the appropriate fine structure constant, $x = 2E_\gamma/\sqrt{s}$, θ angle between photon and incoming electron, S_χ spin of WIMP, J_0 is the dominant value of J in the velocity expansion of (10.68) (as discussed above, commonly $J = 0$ dominates, s -annihilator DM). The accuracy of the method and its predictions are illustrated in fig. 45, where the left panel illustrates the results obtained using the formula (10.69), which are then compared with those of an exact calculation, based on a supersymmetric MSSM model, with WIMP masses 225 GeV, whilst the right panel shows the expected reach in κ_e for a 500 GeV linear e^+e^- collider as a function of the WIMP mass. As we observe from such comparisons the results of the method and of the exact calculation are in pretty good agreement.

We note at this stage, however, that, although model independent, the above process is rarely the dominant collider signature of new physics within a given model. It therefore makes sense to look for model dependent processes at colliders, which we now turn to. In this last respect, it is important to realize [33] that a calculation of slepton masses is essential for computing accurately relic abundances in theoretical models; without a collider measurement of the slepton mass, there may be a significant uncertainty in the relic abundance calculation. This uncertainty results because the slepton mass should then be allowed to vary within the whole experimentally allowed range. We mention here that measuring slepton masses at LHC is challenging due to W^+W^- and $t\bar{t}$ production. However, as shown in [33], it is possible through the study of di-lepton mass distribution $m_{\ell\ell}$ in the decay channel $\tilde{\chi}_2^0 \rightarrow \ell^\pm \ell^\mp \tilde{\chi}_1^0$ and also at the International Linear Collider (ILC). The reader is referred to the literature [33] for further details on these important issues. We are now ready to start our discussion on model-dependent DM signatures at LHC and future colliders.

11 Model-Dependent WMAP SUSY Constraints

We shall concentrate on DM signatures at colliders, using WMAP1,3 data. To illustrate the underlying-theoretical-model dependence of the results we chose three representative theoretical models: (i) the mSUGRA (or constrained MSSM model) [10, 3] and (iii) a non-critical (non-equilibrium) cosmology, based on a particular model of strings, with running dilatons (implying a dilaton quintessence relaxation model for dark energy at late eras) [12]. It must be pointed out at this juncture that other models from critical string theory [11] have been extensively analysed in the literature, involving even non-thermal dark matter, which will not be the topic of our discussion here.

11.1 Constrained MSSM/mSUGRA Model

MSSM has too many parameters to be constrained effectively by data. To minimize the number of parameters one can “embed” this model by taking into account the gravity sector, which from a cosmological point of view is a physical necessity. Such an embedding in principle affects the dark energy sector of the cosmology, and in fact the minimal Supergravity model (mSUGRA) [10], used to yield the Constrained MSSM (CMSSM) predicts too large values of the cosmological constant at a quantum level, and hence it should not be viewed as the physical model. Nevertheless, as far as DM searches are concerned, such models give a pretty good idea of how astrophysical data can be used to constrain particle physics models, and this is the point of view we take in this work. mSUGRA is the best studied model so far as far as constraints on supersymmetric models using astrophysical CMB data are concerned. A relatively recent review on such approaches is given in

[3], where we refer the reader for details and further material and references. In our presentation here we shall be very brief and concentrate only on the basic conclusions of such analyses.

11.1.1 Basic Features: geometry of the parameter space

Before embarking into a detailed analysis of the constraints of the minimal supersymmetric standard model embedded in a minimal supergravity model (CMSSM) [10], we consider it useful to outline the basic features of these models, which will be used in this review. The embedding of SUSY models into the minimal supergravity (mSUGRA) model implies that there are five independent parameters: Three of them, the scalar and gaugino masses $m_0, m_{1/2}$ as well as the trilinear soft coupling A_0 , at the unification scale, set the size of the Supersymmetry breaking scale. In addition one can consider as input parameter $\tan\beta = \frac{\langle H_2 \rangle}{\langle H_1 \rangle}$, the ratio of the v.e.v's of the Higgses H_2 and H_1 giving masses to up and down quarks respectively. The sign (signature) of the Higgsino mixing parameter μ is also an input but not its size which is determined from the Higgs potential minimization condition [3]. The parameter space of mSUGRA can be effectively described in terms of two branches:

(i) An Ellipsoidal Branch (EB) of Radiative Symmetry Breaking, which exists for small to moderate values of $\tan\beta \lesssim 7$, where the loop corrections are typically small. One finds that the radiative symmetry breaking constraint demands that the allowed set of soft parameters m_0 and a combination [3] $m'_{12} = f(m_{1/2}, A_0, \tan\beta)$ lie, for a given value of μ , on the surface of an *Ellipsoid*. This places upper bounds on the sparticle masses for a given value of $\Phi \equiv \mu^2/M_Z^2 + 1/4$.

(ii) Hyperbolic Branch (HB) of Radiative Symmetry Breaking. This branch is realized [34] for large values of $\tan\beta \gtrsim 7$, where the loop corrections to μ are significant. In this branch, $(m_0, m'_{1/2})$

lie now on the surface of a *hyperboloid*: $\frac{m'_{1/2}{}^2}{\alpha^2(Q_0)} - \frac{m_0^2}{\beta^2(Q_0)} \simeq \pm 1$, $Q_0 = 0$ a fixed value of the running scale, α, β constant functions of Φ, M_Z, A_0 . For fixed A_0 , the $m_0, m_{1/2}$ lie on a *hyperbola*, hence they can get large for fixed μ or Φ . What is interesting in the HB case is the fact that m_0 and/or $m_{1/2}$ can become very large, while much smaller values for μ can occur.

(iia) A subset of HB is the so-called *high zone* [34]. In this case electroweak symmetry breaking (EWSB) can occur in regions where m_0 and $m_{1/2}$ can be in the several TeV range, with much smaller values for the parameter μ which however is much larger than M_Z . This has important consequences for phenomenology, as we shall see. In this zone the lightest of the neutralinos, χ_1 , is almost a Higgsino having a mass of order μ . This is called *inversion phenomenon* since the LSP is a Higgsino rather a Bino. The inversion phenomenon has dramatic effects on the nature of the particle spectrum and SUSY phenomenology in this HB. Indeed, as we discussed above, in mSUGRA one naturally has co-annihilation with the sleptons when the neutralino mass extends to masses beyond 150-200 GeV with processes of the type (c.f. fig. 43): $\chi_1^{\tilde{a}} \tilde{\ell}_R^a \rightarrow \ell^a \gamma, \ell^a Z, \ell^a h, \tilde{\ell}_R^a \tilde{\ell}_R^b \rightarrow \ell^a \ell^b$, and $\tilde{\ell}_R^a \tilde{\ell}_R^{b*} \rightarrow \ell^a \ell^b, \gamma\gamma, \gamma Z, ZZ, W^+W^-, hh$, where $\tilde{\ell}$ is essentially a $\tilde{\tau}$. Remarkably the relic density constraints can be satisfied on the hyperbolic branch also by co-annihilation. However, on the HB the co-annihilation is of an entirely different nature as compared with the stau co-annihilations discussed previously: instead of a neutralino-stau co-annihilation, and stau - stau in the HB one has co-annihilation processes involving the second lightest neutralino and chargino states [35], $\chi_1^0 - \chi_1^\pm$, followed by $\chi_1^0 - \chi_2^0, \chi_1^+ - \chi_1^-, \chi_1^\pm - \chi_2^0$. Some of the dominant processes that contribute to the above co-annihilation processes are [35]: $\chi_1^0 \chi_1^+, \chi_2^0 \chi_1^+ \rightarrow u_i \bar{d}_i, \bar{e}_i \nu_i, AW^+, ZW^+, W^+h$ and $\chi_1^+ \chi_1^-, \chi_1^0 \chi_2^0 \rightarrow u_i \bar{u}_i, d_i \bar{d}_i, W^+W^-$. Since the mass difference between the states χ_1^+ and χ_1^0 is the smallest the $\chi_1^0 \chi_1^+$ co-annihilation dominates. In such cases, the masses $m_0, m_{1/2}$ may be pushed beyond 10 TeV, so that squarks and sleptons can get masses up to several TeV, i.e. beyond detectability limits of immediate future accelerators such as LHC.

(iib) Except the high zone where the inversion phenomenon takes place the HB includes the so called *Focus Point (FP)* region [36], which is defined as a region in which some renormalization group (RG) trajectories intersect (FP region would be only a point, were it not for threshold effects which smear it out). We stress that the FP is *not* a fixed point of the RG. The FP region is a subset of the HB limited to relatively low values of $m_{1/2}$ and values of μ close to the electroweak scale, M_Z , while m_0 can be a few TeV but not as large as in the high zone due to the constraints

imposed by the EWSB condition. The LSP neutralino in this region is a mixture of Bino and Higgsino and the Higgsino impurity allows for rapid s-channel LSP annihilations, resulting to low neutralino relic densities at experimentally acceptable levels. This region is characterized by m_0 in the few TeV range, low values of $m_{1/2} \ll m_0$ and rather small values of μ close to M_Z . The LSP neutralino in this case is a mixture of Bino and Higgsino and its Higgsino impurity is adequate to give rise to rapid s-channel LSP annihilations so that the neutralino relic density is kept low at experimentally acceptable values. Since μ is small the lightest chargino may be lighter than 500 GeV and the FP region may be accessible to future TeV scale colliders. Also due to the relative smallness of $m_{1/2}$ in this region gluino pair production may occur at a high rate making the FP region accessible at LHC energies.

It should be pointed out that, although the HB may be viewed as fine tuned, nevertheless recent studies [37], based on a χ^2 analysis, have indicated that the WMAP data, when combined with data on $b \rightarrow s\gamma$ and $g_\mu - 2$, seem to favor the Focus Point HB region and the large $\tan \beta$ neutralino resonance annihilation of mSUGRA.

11.1.2 Muon's anomaly and SUSY detection prospects

Undoubtedly one of the most significant experimental results of the last years is the measurement of the anomalous magnetic moment of the muon [38]. Deviation of its measured value from the Standard Model (SM) predictions is evidence for new physics with Supersymmetry being the prominent candidate to play that role. Adopting Supersymmetry as the most natural extension of the SM, such deviations may be explained and impose at the same time severe constraints on the predictions of the available SUSY models by putting upper bounds on sparticle masses. Therefore knowledge of the value of $g_\mu - 2$ is of paramount importance for Supersymmetry and in particular for the fate of models including heavy sparticles in their mass spectrum, as for instance those belonging to the Hyperbolic Branch.

Unfortunately the situation concerning the anomalous magnetic moment is not clear as some theoretical uncertainties remain unsettled as yet. Until last year, as far as I am aware of, there were two theoretical estimates for the difference of the experimentally measured [38] value of $a_\mu = (g_\mu - 2)/2$ from the theoretically calculated one within the SM [39],

- Estimate (I) $a_\mu^{\text{exp}} - a_\mu^{\text{SM}} = 1.7(14.2) \times 10^{-10}$ $[0.4(15.5) \times 10^{-10}]$
- Estimate (II) $a_\mu^{\text{exp}} - a_\mu^{\text{SM}} = 24.1(14.0) \times 10^{-10}$ $[22.8(15.3) \times 10^{-10}]$

In (I) the τ -decay data are used in conjunction with Current Algebra while in (II) the $e^-e^+ \rightarrow$ Hadrons data are used in order to extract the photon vacuum polarization which enters into the calculation of $g_\mu - 2$. Within square brackets are updated values of Ref. [39]⁹. Estimate (I) is considered less reliable since it carries additional systematic uncertainties and for this reason in many studies only the Estimate (II) is adopted. Estimate (II) includes the contributions of additional scalar mesons not taken into account in previous calculations.

In order to get an idea of how important the data on the muon anomaly might be we quote Ref.[34] where both estimates have been used. If Estimate (II) is used at a 1.5σ range much of the HB and all of the inversion region can be eliminated. In that case the usually explored region of SUSY in the EB is the only one that survives, which, as we shall discuss below, can be severely constrained by means of the recent WMAP data. On the other hand, Estimate (I), essentially implies no difference from the SM value, and hence, if adopted, leaves the HB, and hence its high zone (inversion region), *intact*. In such a case, SUSY may not be detectable at colliders, at least in the context of the mSUGRA model, but may be detectable in some direct dark matter searches, to which we shall turn to later in the article.

For the above reasons, it is therefore *imperative* to determine unambiguously the muon anomalous magnetic moment $g_\mu - 2$ by reducing the errors in the leading order hadronic contribution, experimentally, and improving the theoretical computations within the standard model. In view

⁹Due to the rapid updates concerning $g_\mu - 2$ the values of $a_\mu^{\text{exp}} - a_\mu^{\text{SM}}$ used in previous works quoted in this article may differ from those appearing above.

of its importance for SUSY searches, it should also be necessary to have further experiments in the future, that could provide independent checks of the measured muon magnetic moment by the E821 experiment [38]. Quite recently (2006) a new measurement [13] of the $g_\mu - 2$ became available, which shows a clear discrepancy from the theoretically calculated Standard Model prediction by 3.4σ

$$1.91 \times 10^{-9} < \Delta a_\mu < 3.59 \times 10^{-9} \quad (11.1)$$

thereby pointing towards the elimination of the inversion region of the HB of mSUGRA, according to the above discussion.

11.1.3 WMAP mSUGRA Constraints in the EB

After the first year of running of WMAP, there have been two independent groups working on this update of the CMSSM in light of the WMAP data, with similar results [40, 41] and below we summarize the results of [40] in fig. 46 for some typical values of the parameters $\tan\beta$ and signature of μ . In such analyses one plots m_0 vs. $m_{1/2}$, taking into account the calculated relic abundance of neutralinos in the model and constraining it by means of the WMAP results (10.65). Details are given in [3] and references therein.

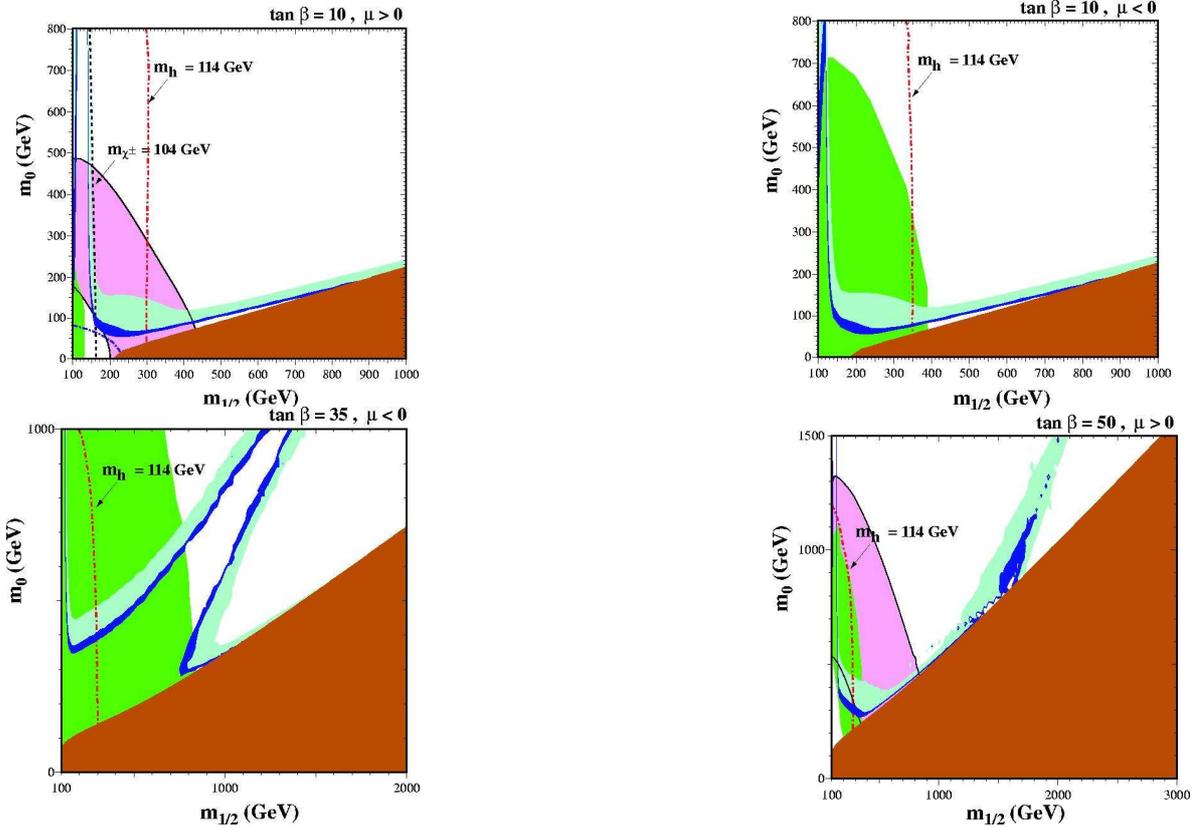


Figure 46: *mSUGRA/CMSSM constraints after WMAP from Ref. [38]. The Dark Blue shaded region is favored by WMAP1 ($0.094 \leq \Omega_\chi h^2 \leq 0.129$). Turquoise shaded regions have $0.1 \leq \Omega_\chi h^2 \leq 0.3$. Brick red shaded regions are excluded because LSP is charged. Dark green regions are excluded by $b \rightarrow s\gamma$. The Pink shaded region includes $2 - \sigma$ effects of $g_\mu - 2$. Finally, the dash-dotted line represents the LEP constraint on \tilde{e} mass.*

For the LSP, the lightest of the charginos, stops, staus and Higgses the upper bounds on their masses of order of a few hundreds of GeV [3], for various values of the parameter $\tan\beta$, if the new WMAP determination [6, 9] of the Cold Dark Matter (10.65) and the 2σ bound $149 <$

$10^{-11} \alpha_\mu^{SUSY} < 573$ of E821 is respected. The lightest of the charginos has a mass whose upper bound is ≈ 550 GeV, and this is smaller than the upper bounds put on the masses of the lightest of the other charged sparticles, namely the stau and stop. Hence the prospects of discovering CMSSM at a e^+e^- collider with center of mass energy $\sqrt{s} = 800$ GeV, are *not* guaranteed. Thus, a center of mass energy of at least $\sqrt{s} \approx 1.1$ TeV is required to discover SUSY through chargino pair production. Note that in the allowed regions the next to the lightest neutralino, $\tilde{\chi}'$, has a mass very close to the lightest of the charginos and hence the process $e^+e^- \rightarrow \tilde{\chi}\tilde{\chi}'$, with $\tilde{\chi}'$ subsequently decaying to $\tilde{\chi} + l^+l^-$ or $\tilde{\chi} + 2\text{jets}$, is kinematically allowed for such large $\tan\beta$, provided the energy is increased to at least $\sqrt{s} = 860$ GeV. It should be noted however that this channel proceeds via the t -channel exchange of a selectron and it is suppressed due to the heaviness of the exchanged sfermion. Therefore only if the center of mass energy is increased to $\sqrt{s} = 1.1$ TeV supersymmetry can be discovered in a e^+e^- collider provided it is based on the Constrained scenario [41].

An important conclusion, therefore, which can be inferred by inspecting the figures 46 is that the constraints implied by a possible discrepancy of $g_\mu - 2$ from the SM value, as seems to be supported by the 2006 data [13] (11.1), ($\alpha_\mu^{SUSY} \gtrsim 15.0 \times 10^{-10}$), when combined with the WMAP restrictions on CDM (neutralino) relic densities (10.65), imply severe restrictions on the available parameter space of the EB and lower significantly the upper bounds on the allowed neutralino masses $m_{\tilde{\chi}}$.

11.1.4 WMAP mSUGRA Constraints in the HB

Despite the above-mentioned good prospects of discovering minimal SUSY models at future colliders, if the EB is realized, however, things may not be that simple in Nature. χ^2 studies [37] of mSUGRA in light of the recent WMAP data has indicated (c.f. figure 47) that the HB/focus point region of the model's parameter space seems to be favored along with the neutralino resonance annihilation region for $\mu > 0$ and large $\tan\beta$ values. The favored focus point region corresponds to moderate to large values of the Higgs parameter μ^2 , and large scalar masses m_0 in the several TeV range. The situation in case the HB is included in the analysis is depicted in figure 48 [34],

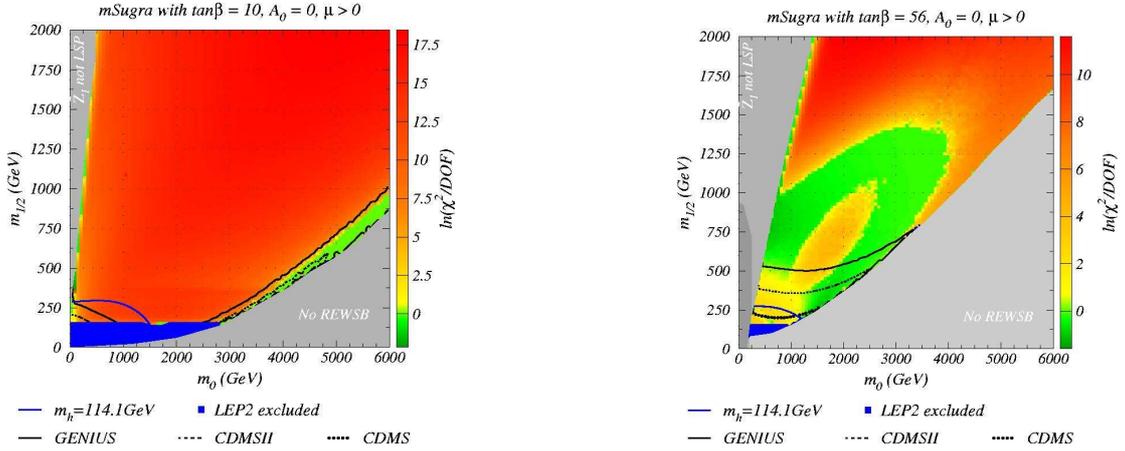


Figure 47: WMAP data seem to favor ($\frac{\chi^2}{\text{DOF}} < 4/3$) (green) the HB/focus point region (moderate to large values of μ , large m_0 scalar masses) for almost all $\tan\beta$ (Left), as well as s - channel Higgs resonance annihilation (Right) for $\mu > 0$ and large $\tan\beta$.

where we plot the $m_0 - m_{1/2}$ graphs, as well as graphs of $m_0, m_{1/2}$ vs the neutralino LSP mass. The neutralino density is that of the WMAP data.

We stress again that, in case the high zone (inversion) region of the HB is realized, then the detection prospects of SUSY at LHC are diminished significantly, in view of the fact that in such

regions slepton masses may lie in the several TeV range (see figure 48). Fortunately, as already mentioned, last years's $g_\mu - 2$ data [13] (11.1) seem to exclude this possibility.

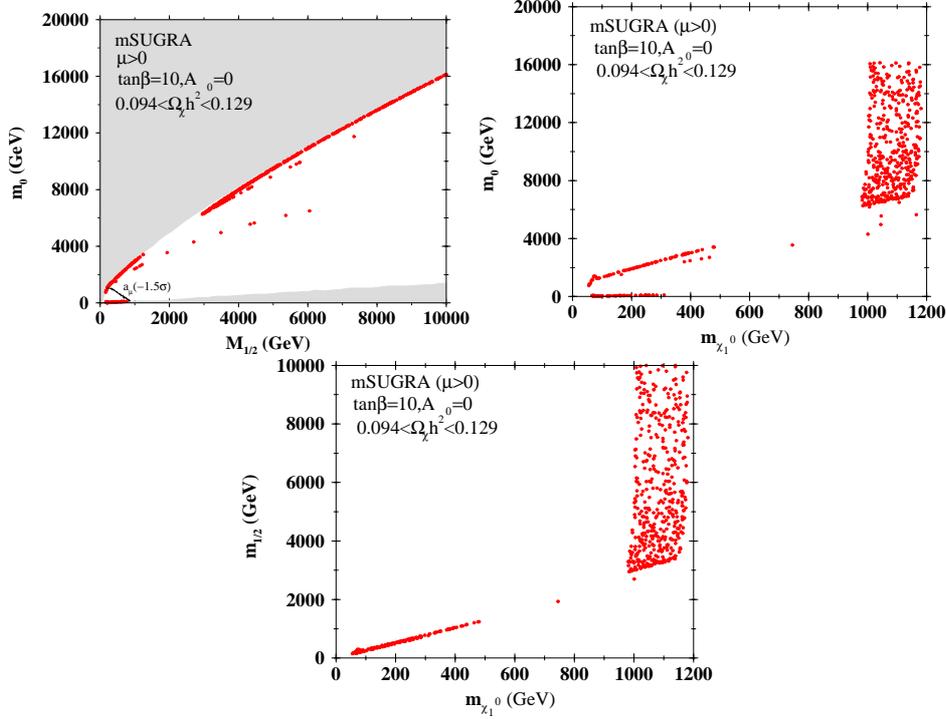


Figure 48: $m_0 - m_{1/2}$ graph, and m_0 and $m_{1/2}$ vs. m_{χ} graphs, including the HB of mSUGRA. Such regions are favored by the WMAP data.

11.1.5 Expected Reach of LHC and Tevatron

In view of the above results, an updated reach of LHC in view of the recent WMAP and other constraints discussed above (see figure 49) has been performed in [42], showing that a major part of the HB, but certainly not its high zone (which though seems to be excluded by means of the recent $g_\mu - 2$ data (11.1)), can be accessible at LHC. The conclusion from this study [42] is that for an integrated luminosity of $100 fb^{-1}$ values of $m_{1/2} \sim 1400$ GeV can be probed for small scalar masses m_0 , corresponding to gluino masses $m_{\tilde{g}} \sim 3$ TeV. For large m_0 , in the hyperbolic branch/focus point region, $m_{1/2} \sim 700$ GeV can be probed, corresponding to $m_{\tilde{g}} \sim 1800$ GeV. It is also concluded that the LHC (CERN) can probe the entire stau co-annihilation region and most of the heavy Higgs annihilation funnel allowed by WMAP data, except for some range of $m_0, m_{1/2}$ in the case $\tan\beta \gtrsim 50$. A similar updated reach study in light of the new WMAP data has also been done for the Tevatron [43], extending previous analyses to large m_0 masses up to 3.5 TeV, in order to probe the HB/focus region favored by the WMAP data [37]. Such studies (c.f. figure 50) indicate that for a 5σ (3σ) signal with 10 (25) fb^{-1} of integrated luminosity, the Tevatron reach in the triplepton channel extends up to $m_{1/2} \sim 190$ (270) GeV independent of $\tan\beta$, which corresponds to a reach in terms of gluino mass of $m_{\tilde{g}} \sim 575$ (750) GeV.

11.1.6 Astrophysical and Collider Dark Matter

Above we have analyzed constraints placed on supersymmetric particle physics models, in particular MSSM, by WMAP/CMB astrophysical data. The analysis made the assumption that neutralinos constitute exclusively the astrophysical DM. It would be desirable to inverse the logic

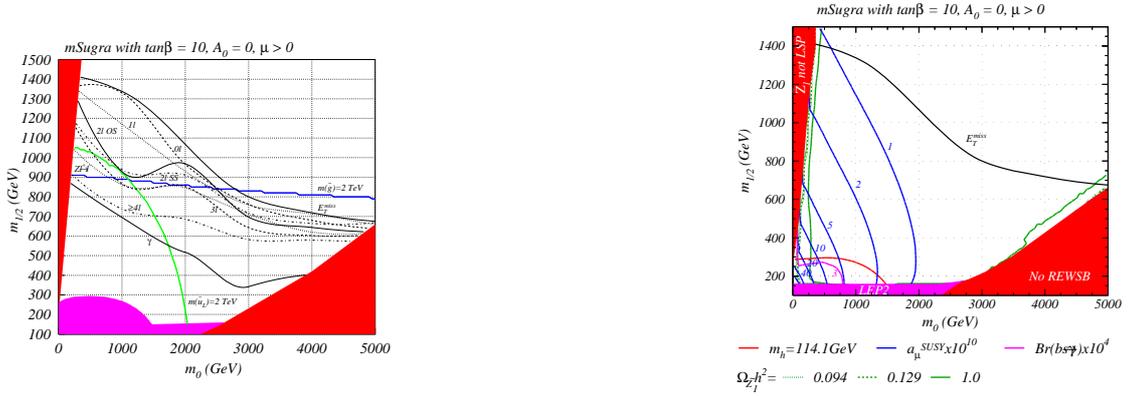


Figure 49: *Left:* The updated Reach in $(m_0, m_{1/2})$ parameter plane of mSUGRA assuming 100 fb^{-1} integrated luminosity. Red (magenta) regions are excluded by theoretical (experimental) constraints. *Right:* Contours (in view of the uncertainties) of several low energy observables : CDM relic density (green), contour of $m_h = 114.1 \text{ GeV}$ (red), contours of $a_\mu 10^{10}$ (blue) and contours of $b \rightarrow s\gamma$ BF ($\times 10^4$) (magenta).

and ask the question [44]: “are neutralinos produced at the LHC the particles making up the astronomically observed dark matter?”

To answer this question, let us first recall the relevant neutralino interactions (within the mSUGRA framework) that could take place in the Early universe (fig. 51). As we have discussed previously, the WMAP3 constraint (10.65) limits the parameter space to three main regions arising from the above diagrams (there is also a small “bulk” region): (1) The stau-neutralino ($\tilde{\tau}_1 - \tilde{\chi}_1^0$) co-annihilation region. Here m_0 is small and $m_{1/2} \leq 1.5 \text{ TeV}$. (2) The focus region where the neutralino has a large Higgsino component. Here $m_{1/2}$ is small and $m_0 \geq 1 \text{ TeV}$. (3) The funnel region where annihilation proceeds through heavy Higgs bosons which have become relatively light. Here both m_0 and $m_{1/2}$ are large. A key element in the co-annihilation region is the Boltzmann factor from the annihilation in the early universe at $kT \sim 20 \text{ GeV}$: $\exp[-\Delta M/20]$, $\Delta M = M_{\tilde{\tau}_1} - M_{\tilde{\chi}_1^0}$ implying that significant co-annihilation occurs provided $\Delta M \leq 20 \text{ GeV}$.

The accelerator constraints further restrict the parameter space and if the muon $g_{\mu-2}$ anomaly maintains [13], (c.f. (11.1)), then $\mu > 0$ is preferred and there remains mainly the co-annihilation region (c.f. figure 52). Note the cosmologically allowed narrow co-annihilation band, due to the Boltzmann factor for $\Delta M = 5 - 15 \text{ GeV}$, corresponding to the allowed WMAP range for $\Omega_{\tilde{\chi}_1^0} h^2$.

One may ask, then, whether: (i) such a small stau-neutralino mass difference (5-15 GeV) arise in mSUGRA, since one would naturally expect these SUSY particles to be hundreds of GeV apart and (ii) such a small mass difference be measured at the LHC. If the answer to both these questions is in the affirmative, then the observation of such a small mass difference would be a strong indication [44] that the neutralino is the astronomical DM particle, since it is the cosmological constraint on the amount of DM that forces the near mass degeneracy with the stau, and it is the accelerator constraints that suggest that the co-annihilation region is the allowed region.

As far as question (i) is concerned, one observes the following: In the mSUGRA models, at GUT scale we expect no degeneracies, the ΔM is large, since $m_{1/2}$ governs the gaugino masses, while m_0 the slepton masses. However, at the electroweak scale (EWS), the Renormalization Group Equation can modify this: e.g. the lightest selectron \tilde{e}^c at EWS has mass $m_{\tilde{e}^c}^2 = m_0^2 + 0.15m_{1/2}^2 + (37\text{GeV})^2$ while the $\tilde{\chi}_1^0$ has mass $m_{\tilde{\chi}_1^0}^2 = 0.16m_{1/2}^2$. The numerical accident that coefficients of $m_{1/2}^2$ is nearly the same for both cases allows a near degeneracy: for $m_0 = 0$, \tilde{e}^c and $\tilde{\chi}_1^0$ become degenerate at $m_{1/2} = (370-400) \text{ GeV}$. For larger $m_{1/2}$, near degeneracy is maintained by increasing m_0 to get the narrow corridor in m_0 - $m_{1/2}$ plane. Actually the case of the stau $\tilde{\tau}_1$ is more complicated [44]: large t-quark mass causes left-right mixing in the stau mass matrix and



Figure 50: Left: The reach of Fermilab Tevatron in the m_0 vs. $m_{1/2}$ parameter plane of the mSUGRA model, with $\tan\beta = 10$, $A_0 = 0$ and $\mu > 0$, assuming a 5σ signal at 10 fb^{-1} (solid) and a 3σ signal with 25 fb^{-1} of integrated luminosity (dashed). The red (magenta) region is excluded by theoretical (experimental) constraints. The region below the magenta contour has $m_h < 114.1$ GeV, in violation of Higgs mass limits from LEP2. Right: The reach of Fermilab Tevatron in the m_0 vs. $m_{1/2}$ parameter plane of the mSUGRA model, with $\tan\beta = 52$, $A_0 = 0$ and $\mu > 0$. The red (magenta) region is excluded by theoretical (experimental) constraints. The region below the magenta contour has $m_h < 114.1$ GeV, in violation of Higgs mass limits from LEP2.

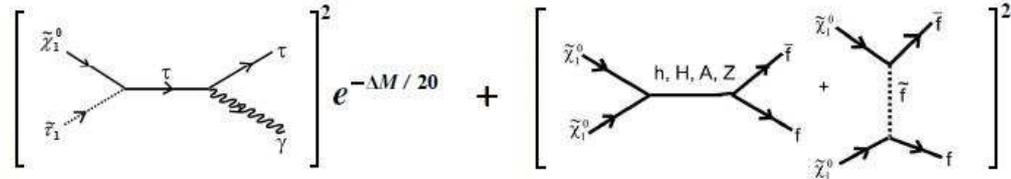


Figure 51: The Feynman diagrams for annihilation of neutralino dark matter in the early universe. The Boltzmann factor $e^{-\Delta M/20}$ in the stau-co-annihilation graph is explicitly indicated.

this results in the $\tilde{\tau}_1$ being the lightest slepton and not the selectron. However, a result similar to the above occurs, with a $\tilde{\tau}_1 - \tilde{\chi}_1^0$ co-annihilation corridor appearing.

We note that the above results depend only on the U(1) gauge group and so co-annihilation can occur even if there were non-universal scalar mass soft-breaking or non-universal gaugino mass soft breaking at M_G . Thus, co-annihilation can occur in a wide class of SUGRA models, not just in mSUGRA. Hence, in such models one has naturally near degenerate neutralino-staus, and hence the answer to question (i) above is affirmative.

Now we come to the second important question (ii), namely, whether LHC measurements have the capability of asserting that the neutralino (if discovered) is the astrophysical DM. To this end we note that, in LHC, the major SUSY production processes of neutralinos are interactions of gluinos (\tilde{g}) and squarks (\tilde{q}) (c.f. figure 53), e.g., $p + p \rightarrow \tilde{g} + \tilde{q}$. These then decay into lighter SUSY particles. The final states involve two neutralinos $\tilde{\chi}_1^0$ giving rise to missing transverse energy E_{miss}^T and four τ 's, two from the \tilde{g} and two from the \tilde{q} decay chain for the example of fig. 53.

In the co-annihilation region, two of the taus have a high energy (“hard” taus) coming from the $\tilde{\chi}_2^0 \rightarrow \tau\tilde{\tau}_1$ decay (since $M_{\tilde{\chi}_2^0} \simeq 2M_{\tilde{\tau}_1}$), while the other two are low energy particles (“soft” taus) coming from the $\tilde{\tau}_1 \rightarrow \tau + \tilde{\chi}_1^0$ decay, since ΔM is small.

The signal is thus $E_T^{\text{miss}} + \text{jets} + \tau$'s, which should be observable at the LHC detectors [44]. As seen above, we expect two pairs of taus, each pair containing one soft and one hard tau from each $\tilde{\chi}_2^0$ decay. Since $\tilde{\chi}_2^0$ is neutral, each pair should be of opposite sign. This distinguishes them from SM- and SUSY-backgrounds jets-faking taus, which will have equal number of like-sign as opposite-sign events [44]. Thus, one can suppress backgrounds statistically by considering the number of opposite-sign events N_{OS} minus the like-sign events N_{LS} (figure 54).

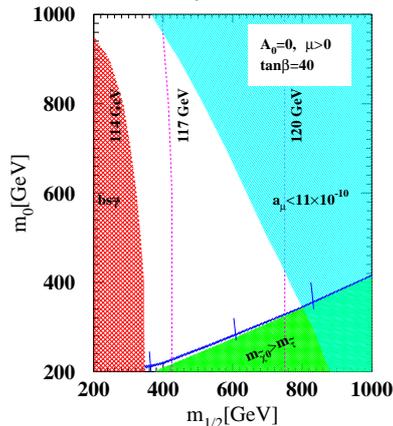


Figure 52: Allowed parameter space in mSUGRA. Dashed vertical lines are possible Higgs masses (from [42]).

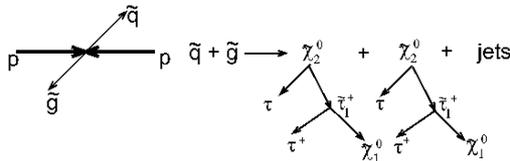


Figure 53: SUSY production of neutralinos and decay channels

The four τ final state has the smallest background but the acceptance and efficiency for reconstructing all four taus is low. Thus to implement the above ideas we consider here the three τ final state of which two are hard and one is soft. There are two important features: First, N_{OS-LS} increases with ΔM (since the τ acceptance increases) and N_{OS-LS} decreases with $M_{\tilde{g}}$ (since the production cross section of gluinos and squarks decrease with $M_{\tilde{g}}$). Second, one sees that N_{OS-LS} forms a peaked distribution. The di-tau peak position $M_{\tau\tau}^{\text{peak}}$ increases with both ΔM and $M_{\tilde{g}}$. This allows us to use the two observables N_{OS-LS} and $M_{\tau\tau}^{\text{peak}}$ to determine both ΔM and $M_{\tilde{g}}$ (c.f. figure 55). As becomes evident from the analysis [44] (c.f. fig. 56) it is possible to simultaneously determine ΔM and the gluino mass $M_{\tilde{g}}$. Moreover, one sees that at LHC even with 10 fb^{-1} (which should be available at the LHC after about two years running) one could determine ΔM to within 22%, which should be sufficient to know whether one is in the SUGRA co-annihilation region. The above analysis was within the mSUGRA model, however similar analyses for other SUGRA models can be made, provided the production of neutralinos is not suppressed. In fact, the determination of $M_{\tilde{g}}$ depends on mSUGRA universality of gaugino masses at GUT scale, M_G , to relate $M_{\tilde{\chi}_2^0}$ to $M_{\tilde{g}}$ thus a model independent method of determining $M_{\tilde{g}}$ would allow one to test the question of gaugino universality. However, it may not be easy to directly measure $M_{\tilde{g}}$ at the LHC for high $\tan\beta$ in the co-annihilation region due to the large number of low energy taus, and the ILC would require a very high energy option to see the gluino.

One can also measure [44] ΔM using the signal $E_T^{\text{miss}} + 2 \text{ jets} + 2\tau$. This signal has higher acceptance but larger backgrounds. With 10 fb^{-1} one can measure ΔM with 18% error at the benchmark point assuming a separate measurement of $M_{\tilde{g}}$ with 5% error has been made. While the benchmark point has been fixed in [44] at $M_{\tilde{g}} = 850 \text{ GeV}$ (i.e. $m_{1/2} = 360 \text{ GeV}$), higher gluino mass would require more luminosity to see the signal. One finds that with 100 fb^{-1} one can probe $m_{1/2}$ at the LHC up to $\sim 700 \text{ GeV}$ (i.e., $M_{\tilde{g}}$ up to $\simeq 1.6 \text{ TeV}$). Finally it should be mentioned that measurements of ΔM at the ILC could be made if a very forward calorimeter is implemented to reduce the two γ background. In such a case, ΔM can be determined with 10% error at the benchmark point, thereby implying that [44] in the co-annihilation region, the determination of

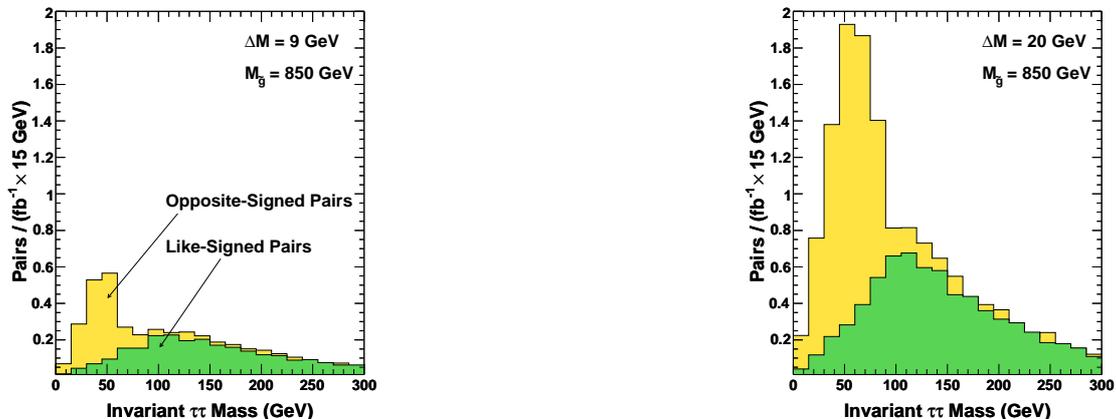


Figure 54: Number of tau pairs as a function of invariant $\tau\tau$ mass. The difference $N_{OS}-N_{LS}$ cancels for mass ≥ 100 GeV eliminating background events (from [42]).

ΔM at the LHC is not significantly worse than at the ILC.

The results on the accuracy of determining DM mass in astrophysics and colliders within the mSUGRA framework is given in figure 57. We see that the cosmological measurement are at present the most accurate one, however, the reader should bear in mind the model-dependence of all these results. We now come to demonstrate this point by repeating the analysis for some class of stringy models.

11.2 A Stringy Model with dilaton sources: making the model dependence of astrophysics constraints on particle physics models explicit

As an illustration of the strong dependence of the constraints on particle physics models, such as supersymmetry, on the underlying theoretical model of cosmology, we discuss in this subsection a particular string-inspired model and show how the constraints from astrophysical measurements of the (thermal) dark-matter relic abundance, imposed on the minimal supersymmetric standard model, discussed within the framework of standard cosmology in previous sections, are modified in our string model due to the coupling of dark matter with the dilaton field, ϕ (which is a spin-zero (scalar) excitation of the string spectrum). It should be noted that this dilaton-dark matter coupling is only a (speculative) model, and it should by no means be considered as generic. It is, nevertheless, an instructive example of the above-mentioned model dependence of the astro-particle physics constraints. We shall be very sketchy in our discussion, and concentrate on outlining only the main results. For further details the interested reader is referred to the literature [28, 2].

As discussed in [28], the model is a non-equilibrium string inspired model for the Universe, where at late eras it has not yet relaxed completely to its equilibrium state. This non equilibrium may be the result of a cosmically catastrophic event at an early epoch. In the modern version of string cosmology, for instance, where our world is viewed as a *brane* domain wall embedded in a higher-dimensional space time, where only gravitational degrees of freedom propagate, whilst Standard model matter are confined to the brane, such catastrophic events, causing departure from equilibrium, could be the result of the collision of two such *brane worlds*. In such models, the presence of a cosmic-time dependent dilaton $\phi(t)$ coupled to dark matter (which can be provided by supersymmetric partners, such as the neutralino), affects [28] the relevant Boltzmann equation by introducing a Γ source term in (10.43), with

$$\Gamma = \dot{\phi}.$$

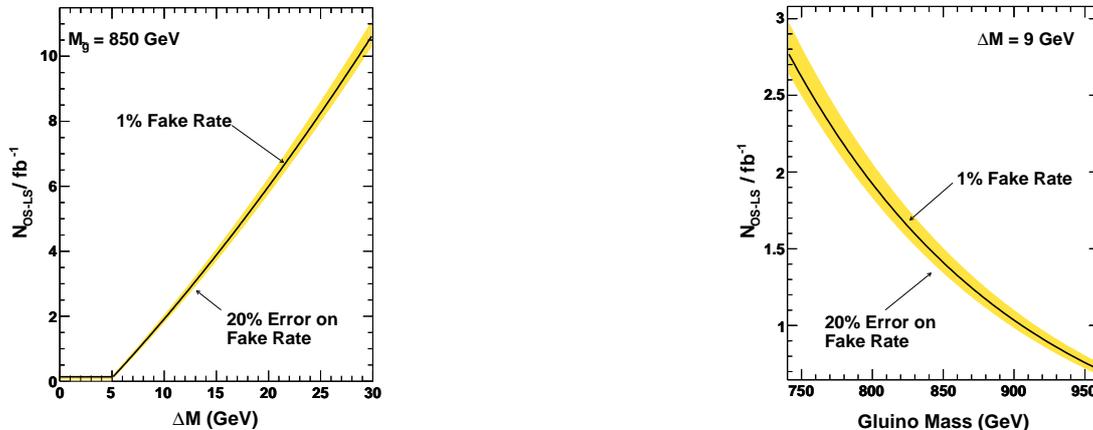


Figure 55: N_{OS-LS} as function of ΔM (left graph) and as a function of $M_{\tilde{g}}$ (right graph). The central black line assumes a 1% fake rate, the shaded area representing the 20% error in the fake rate (from [42]).

Moreover, the corresponding gravitational equations, obtained from variations of the pertinent effective action of the (non-critical/non-equilibrium) string theory at hand, are modified by dilaton terms, in the way explained in [28], which will not be repeated here for brevity. In fact, matter in such systems include ordinary dust-like, with an equation of state $w_d = p_d/\rho_d = 0$, and with conservation equation with a “friction”-type $\dot{\phi}$ -term

$$\dot{\rho}_d + 3H\dot{\rho}_d - \dot{\phi}\rho_d + \dots = 0 .$$

as well as “exotic” dark matter D components, coupled to the dilaton in such a way that they are characterised by an equation of state with exotic form $p_D/\rho_D = w_D = 0.4$. Their conservation equation is more complicated than that for dust matter, as a result of their non conventional couplings to the dilatons. As a result, the total matter, including dust and exotic forms, obey a conservation equation of the form [28]:

$$\begin{aligned} \dot{\tilde{q}}_m + 2Q\dot{Q}e^{2\phi} = & -3\hat{H}(\tilde{q}_m + \tilde{p}_m) + \dot{\phi}(\tilde{q}_m - 3\tilde{p}_m) + \\ & 4(\hat{H} + \dot{\phi})(-\hat{H}^2 + \dot{\phi}^2 + e^\phi Q(\dot{\phi} + \hat{H}) + \tilde{p}_m) \end{aligned}$$

where the $\tilde{\cdot}$ denotes quantities pertaining to total matter, including exotic forms, and $Q(t)$ is a parameter in the (non-equilibrium) string model that quantifies the deviation from equilibrium. This parameter can be computed rigorously within the string theory framework. In equilibrium (critical) strings $Q \rightarrow 0$.

I make here a note of caution: the overwhelming majority of the string literature deals with critical strings, so the reader should be aware that the above discussion is speculative, but may not be un-realistic, given that the string Universe may be described by such non-equilibrium situations after all. At any rate, the point of this discussion, as I stressed above, is to demonstrate the model dependence of the astrophysical constraints on particle physics models, and most of the features of the above model, namely coupling of the dilaton, exotic forms of dark matter *etc.*, might be sufficiently generic to characterise other theories as well. Solving these equations, together with their gravitational counterparts (not exhibited here), consistently one can show [28] that the delicate era of Nucleosynthesis (i.e. the formation of the light elements, characterised by a delicate balance between expansion rate of the Universe and the pertinent nuclear reaction rate of the relevant interactions in the early Universe) is not affected much, and its standard-cosmology predictions continue to hold in this model. In particular, we find that, at temperatures

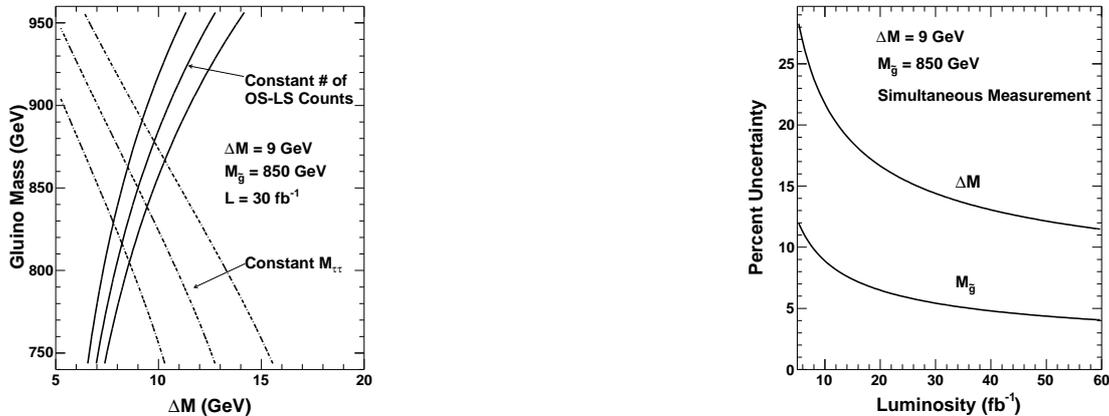


Figure 56: Left: Simultaneous determination of ΔM and $M_{\tilde{g}}$. The three lines plot constant N_{OS-LS} and $M_{\tau\tau}^{\text{peak}}$ (central value and 1σ deviation) in the $M_{\tilde{g}}-\Delta M$ plane for the benchmark point of $\Delta M=9$ GeV and $M_{\tilde{g}}=850$ GeV assuming 30 fb $^{-1}$ luminosity. Right: Uncertainty in the determination of ΔM and $M_{\tilde{g}}$ as a function of luminosity (from [42]).

$T \simeq 1$ MeV, radiation prevails over ordinary matter by almost seven orders of magnitude as demanded by Primordial Nucleosynthesis. It is worth noting that the radiation to matter ratio depends rather sensitively on the value of w_D (equation of state of the dark matter species coupled to dilaton) and it is remarkable that the cosmologically interesting values for w_D , according to the current astrophysical data, coincide with those for which the photon to matter ratio for successful Primordial Nucleosynthesis is in the right ball park, while diluting at the same time the Lightest Supersymmetric Particle (LSP) relics by factors of $\mathcal{O}(10)$. This dilution is the result of time-dependent dilaton source terms in the corresponding Boltzmann equation (10.43), which affect the relic density (10.62),

$$\text{Relic Density Dilution factor} \equiv \left(\frac{\tilde{g}_*}{g_*}\right)^{1/2} \left(1 + \int_{x_0}^{x_f} \frac{\Gamma H^{-1}}{\psi(x)} dx\right).$$

The reader should recall that in the absence of a source, $\Gamma = 0$, $\tilde{g}_* \rightarrow g_*$, so indeed the dilution characterises only situations with a non-trivial Γ .

In fact, the dilution of the the neutralino Dark Matter density in the string model is to such a level [28] that, while it relaxes the severe constraints imposed by conventional cosmology (c.f. discussions in previous sections on minimal SUGRA model), still keeps it in a SUSY parameter space exploitable by LHC. The relevant results are indicated in fig. 58, where comparison with the Standard-Cosmology constraints is given. The reader might also be interested in knowing that it has been shown [28] that for the set of parameters that provide the best fit to all supernovae data until recently [17], the non-critical string model predicts a rather smooth evolution of Dark Energy, for the last ten billion years, thus in accordance with the very recent supernovae data.

In addition to thermal dark matter, in string [11] or other theories there is also the case of models involving non-thermal dark matter, which requires very different techniques to detect, and has not been discussed here. We hope that the reader is by now convinced about the stringent theoretical-model dependence of many of the astrophysical constraints imposed on particle physics models. Such a dependence complicates matters, e.g. as far as “smoking-gun evidence” for supersymmetry at LHC is concerned. In the above-discussed non-critical string model, for instance, the dilution of the dark matter relic abundance by a factor of 10, and almost no dilution for baryons, was just about right in order that neutralino Dark matter continues to be the leading Dark Matter candidate, but clearly this dilution was a consequence of fine tuning of the equation of state of the

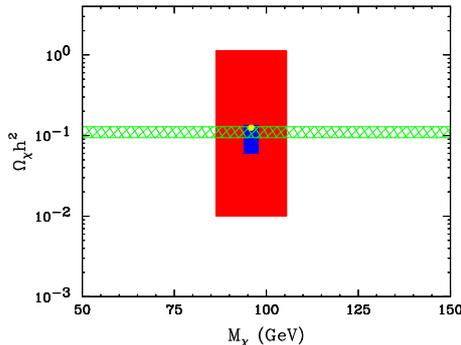


Figure 57: Accuracy of WMAP (horizontal green shaded region), LHC (outer red rectangle) and ILC (inner blue rectangle) in determining M_χ , the mass of the lightest neutralino, and its relic density $\Omega_\chi h^2$. The yellow dot denotes the actual values of M_χ and $\Omega_\chi h^2$ for a sample point in parameter space of mSUGRA: $m_0 = 57$ GeV, $m_{1/2} = 250$ GeV, $A_0 = 0$, $\tan\beta = 10$ and $\text{sign}(\mu) = +1$ (from A. Birkedal *et al.*, arXiv:hep-ph/0507214)

dark matter (even though this seems necessary for nucleosynthesis, nevertheless the value $w = 0.4$ was *not* derived from microscopic considerations). If some of these constraints are relaxed, one may have a further dilution of the amount of thermal dark matter relics in the Universe, which could enlarge the cosmologically allowed parameter space of the model significantly.

For future directions, it would be desirable to explore in more detail SUSY models with *CP violation*, which recently started attracting attention [46], since, due to bounds on the Higgs particle mass, coming from electroweak data (at the LEP (CERN) collider), $m_H > 114$ GeV, we now know that the amount of CP Violation in the Standard Model is not sufficient to generate the observed baryon asymmetry of the Universe [47], and hence SUSY CP violation might play an important rôle in this respect. At this point I should mention that parameters in SUGRA models that can have CP phases are the gaugino and higgsino masses and trilinear sfermion-Higgs couplings. CP phases affect co-annihilation scenarios, and hence the associated particle physics dark matter searches at colliders [46].

Another direction is to constrain SUSY GUTs models (e.g. flipped SU(5)) using astrophysical data [3], after taking, however, proper account of the observed dark energy in the Universe. Personally, I believe that this dark energy is not a cosmological constant, but depends (softly) on cosmic time, due to some quintessence (relaxing to zero (non-equilibrium) field). WMAP data point towards an equation of state of quintessence type, $w = p/\rho \rightarrow -1$ (close to that of a cosmological constant, but not quite -1). Such features may be shared by dilaton quintessence in string theory, as mentioned briefly above. The issue is, however, still wide open and constitutes one of the pressing future directions for theoretical research in this field.

On the experimental side, LHC and future (linear) collider, but also direct [24], dark matter searches could shed light on the outstanding issue of the nature of the Cosmological Dark Sector (especially Dark Matter), but one has to bear in mind that such searches are highly theoretical-model dependent. To such ideas one should also add the models invoking Lorentz violation as alternative to dark matter. Clearly, particle physics can play an important rôle in constraining such alternative models in the future, especially in view of the currently operating or upcoming high-precision terrestrial and extraterrestrial experiments, such as Auger, Planck mission, high-energy neutrino astrophysics experiments *etc.*

Nothing is certain, of course, and very careful interpretations of possible results are essential. Nevertheless, the future looks promising. Certainly particle physics and astrophysics will proceed together and provide a fruitful and complementary experience to each other and exchange interesting sets of ideas for the years to come.

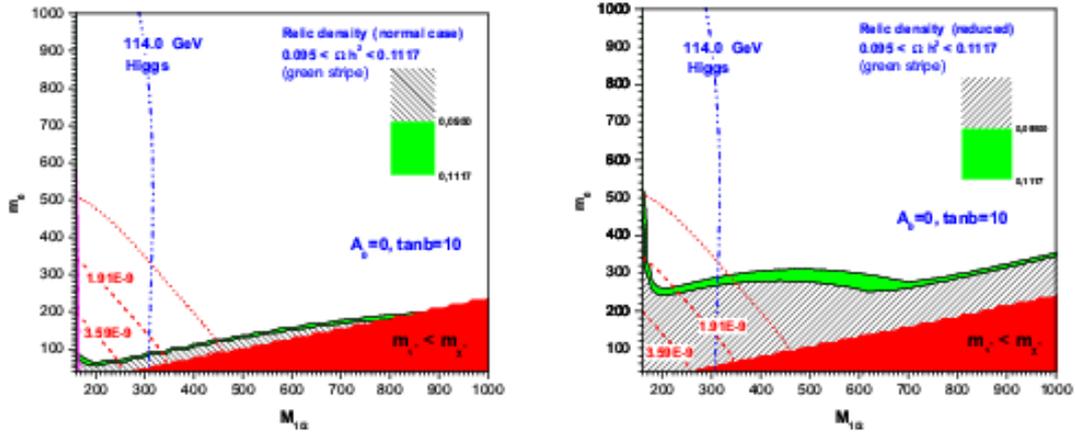


Figure 58: Left: In the thin green (grey) stripe the neutralino relic density is within the WMAP3 limits for values $A_0 = 0$ and $\tan\beta = 10$, according to the source-free $\Gamma=0$ *conventional Cosmology*. The dashed lines (in red) are the 1σ boundaries for the allowed region by the $g - 2$ muon's data as shown in the figure. The dotted lines (in red) delineate the same boundaries at the 2σ 's level. In the hatched region $0.0950 > \Omega_{CDM}h^2$, while in the dark (red) region at the bottom the LSP is a stau. Right: The same as in left panel, but according to the non-critical-string calculation, in which the relic density is reduced in the presence of dilaton sources $\Gamma = \dot{\phi} \neq 0$.

12 Epilogue

With the above analysis we have thus arrived at the end of our discussion of General Relativity and Cosmology. In this course we have only grazed the surface of a huge subject. Our intention was to provide the advanced undergraduate or first-year graduate Physics student with an elementary knowledge of this fascinating subject, which we hope was sufficient to motivate interested students to continue studies at a postgraduate Ph.D. and/or post-doctoral levels.

There are many issues in General Relativity that we did not cover, including spherical solutions for Stars, rotating (spinning) celestial objects, a precise study of the structure of Black Holes, formation of the latter by collapsing stars, etc. All these are topics that can be covered in an M.Sci. or first year graduate Ph.D. specialised course. In addition, there are interesting theoretical models of the early epochs of an expanding Universe, such as the inflationary model, which seem to provide the most satisfactory explanations to date on the observed large-scale homogeneity. Such topics, including the important one on cosmological perturbations, have been covered either very sketchily or not at all during the course, not only because of lack of time, but also because they require knowledge of field theory that undergraduate students do not have. All such topics can be the subject of a specialised graduate programme. Moreover, the astro-particle physics sections on dark matter and how one can use astrophysical measurements to constrain interesting particle physics models, such as supersymmetry, were only superficially discussed, because again a more detailed discussion requires specialised knowledge, which can be acquired at advanced years of a graduate programme, or even at a post doctoral level.

General Relativity and Cosmology are clearly subjects at the frontier of knowledge, and by their very nature are difficult to comprehend, since they deal with topics that pertain to the structure of spacetime itself. However, the hope is that by studying the material covered in this course the student must have realized that general Relativity is not more difficult a subject than, say, Quantum Mechanics or Special Relativity. As we have seen, its language (of tensors) is a bit *peculiar*, appearing difficult at first sight. But the hope is that the physical applications of the theory outlined during the course compensated this apparent difficulty, and showed that knowledge of basic physics helps in grasping the basic ideas and techniques behind Einstein's classical theory of Gravitation.

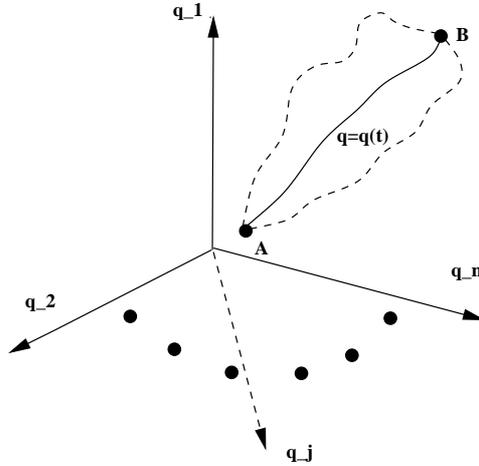


Figure 59: Towards the derivation of Lagrange’s equations. The physical (classical) path (solid line) of a dynamical system (e.g. particle) is selected among the possible paths (dashed lines) as the one that *minimises* the action (Hamilton’s principle) upon infinitesimal variations of the d.o.f.: $\delta q_i(t)$, such that at the end points A and B of the path the d.of. are *fixed*, i.e. $\delta q_i(A) = \delta q_i(B) = 0$.

Acknowledgements

I wish to thank J. Bernabéu and the Department of Theoretical Physics of the University of Valencia for their kind invitation to lecture on their doctorate programme in May 2008 and their hospitality and support during my stay. These significantly extended and revised set of notes, especially on the Modern Cosmology sections, is based on my lectures in this programme.

Appendix A: Lagrange Equations

Lagrange’s equation is an important topic of Mechanics, which finds wide applications as describing the classical dynamics of systems in all areas of Physics, from classical mechanics to field theories, including Gravitation. In this Appendix we give a brief derivation of Lagrange’s dynamical equations, stemming from Hamilton’s principle of least action.

To start with, consider a dynamical Newtonian system which is described by n dynamical degrees of freedom (d.o.f.): $\{q_1, q_2 \dots q_n\}$. Time is an absolute variable in Newtonian Mechanics, to which all observers agree. The path of the system in the parameter (d.o.f.) space is described by solutions of the dynamical equations of motion, yielding $q_i = q_i(t)$, which are derived by means of the *Hamilton’s principle* or principle of *least action*, which can be described as follows:

Consider possible paths of the system, e.g. a particle, in the parameter (d.o.f.) space, as shown in fig. 59 (in the case of a Newtonian particle, the d.o.f. denote the spatial coordinates). The action of the system \mathcal{S} is given by the integral of the Lagrangian function $\mathcal{L}(q_j, \dot{q}_j; t$ over the (universal) time parameter t that parametrises the path (“trajectory”), $\{q_j = q_j(t)\}$:

$$\mathcal{S} \equiv \int_{t_{in}}^{t_f} dt \mathcal{L}(q_j, \dot{q}_j; t) , \quad \dot{q}_j \equiv \frac{dq_j}{dt} . \quad (12.1)$$

Hamilton’s principle, which selects the physical (classical) path among the possible paths of the system (c.f. fig. 59), states that the latter is obtained by minimising the function \mathcal{S} , that is considering its total variation $\delta \mathcal{S}$ upon *arbitrary infinitesimal* variations of the d.o.f., δq_j , $j =$

1, ... n, and setting it to zero:

$$\delta\mathcal{S} = 0 \quad (12.2)$$

Taking into account that the Lagrangian function is a function of both q_j and \dot{q}_j , as well as the fact that the beginning and end points of all paths occur at *fixed* times t_{in} and t_{f} , respectively, we then obtain from (12.1) and (12.2):

$$0 = \delta\mathcal{S} = \sum_{j=1}^n \int_{t_{\text{in}}}^{t_{\text{f}}} dt \left(\delta q_j \frac{\partial \mathcal{L}}{\partial q_j} + \delta \dot{q}_j \frac{\partial \mathcal{L}}{\partial \dot{q}_j} \right) . \quad (12.3)$$

Taking into account that $\delta \dot{q}_i = \frac{d}{dt}(\delta q_i)$, we can partially integrate the second term on the right-hand-side (r.h.s.) of (12.3), to obtain:

$$0 = \delta\mathcal{S} = \sum_{j=1}^n \int_{t_{\text{in}}}^{t_{\text{f}}} dt \delta q_j \left(\frac{\partial \mathcal{L}}{\partial q_j} - \frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}_j} \right) \right) + \sum_{i=1}^n \left(\delta q_i \frac{\partial \mathcal{L}}{\partial \dot{q}_i} \right)_{t=t_{\text{in}}}^{t=t_{\text{f}}} . \quad (12.4)$$

The last term (in square brackets) on the r.h.s. of (12.4) vanishes, on account of the fact that at the end points of the paths the variations δq_j vanish: $\delta q_i(t = t_{\text{in}}) = \delta q_i(t = t_{\text{f}}) = 0$, $i = 1, \dots, n$. Given that one considers *arbitrary* infinitesimal *variations* $\delta q_i(t)$, we then obtain from (12.3) the Lagrange's equations:

$$\frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}_i} \right) - \frac{\partial \mathcal{L}}{\partial q_i} = 0, \quad i = 1, \dots, n . \quad (12.5)$$

The system of n -differential equations (12.5) describes the classical dynamics of the system. As we have seen, they have been derived from Hamilton's principle of least action.

The above construction can be generalised to any relativistic system as well as any field-theory covariant action. For a relativistic system, the coordinates of the particle are now space time coordinates, described by contra-variant four vectors x^μ , and the universal time of Newtonian mechanics is replaced by the proper time τ , as we have discussed in the text, upon which all observers agree. This is the wrist-watch (rest-frame) time of the (massive) observer. To incorporate photons in this picture, where proper time cannot be defined, as there is no rest frame, one generalises the concept of the relativistic-path parameter, by introducing the concept of the affine parameter λ , such that a relativistic path is described by $x^\mu = x^\mu(\lambda)$. We discussed this in the Lectures, when we derived the geodesics of a particle in a background gravitational field. The *affine* parameter is by *definition proportional* to the *proper length* of the relativistic path from the initial point A to the end point B. For a massive particle, $\lambda \propto \tau$.

The relativistic Lagrangian describing a particle, of (rest) mass m , in a background gravitational field, which is "free" in the Einstein sense, that is the particle feels only the influence of Gravity, reads (c.f. Lectures):

$$\mathcal{L}^{\text{free}} = \frac{1}{2} m g_{\alpha\beta}(x^\mu) \frac{dx^\mu}{d\lambda} \frac{dx^\mu}{d\lambda} \quad (12.6)$$

In the general case of a particle with additional interactions, the Lagrangian \mathcal{L} contains also potential terms.

The Lagrange equations for the variables x^μ will now read, in direct analogy with (12.5):

$$\frac{d}{d\lambda} \left(\frac{\partial \mathcal{L}}{\partial x'^\mu} \right) - \frac{\partial \mathcal{L}}{\partial x^\mu} = 0, \quad x'^\mu \equiv \frac{d}{d\lambda} x^\mu, \quad \mu = 0, 1, 2, 3 . \quad (12.7)$$

In the free case (12.6), the corresponding Lagrange equations, as we have discussed in the Lectures, are the geodesics of the particle.

Generalisation to field theories, including Gravitation, are done, upon considering Hamilton's principle for the upon variations of the corresponding field theory actions with respect the relativistic fields $\phi_{\mu_1, \dots, \mu_n}$, which are tensorial in general. A generic covariant field-theory matter

action, consistent with being invariant under general coordinate transformations, as required by the Relativity principle of Einstein, reads:

$$\mathcal{S}^{\text{matter}} = \int d^4x \sqrt{-g} \mathcal{L}^{\text{matter}}(\phi_{\mu_1, \dots, \mu_n}, \phi_{\mu_1, \dots, \mu_n; \nu}; g_{\mu\nu}, g_{\mu\nu, \rho}) \quad (12.8)$$

and depends on both the fields and their (gravitational) covariant derivatives (denoted by ; as usual). Notice that, as discussed in the Lectures, the invariant proper space-time volume element, entering the expression for the action, is $\sqrt{-g}d^4x$, where g is the determinant of the gravitational field (metric tensor). Notice the dependence of the matter action on the metric tensor and its (ordinary) derivatives, as a result of general covariance (contraction of indices, covariant derivatives of matter fields *etc.*)

Ignoring gravitational variations, Hamilton's principle of (12.8) yields the matter field theory Lagrange equations:

$$\partial_\nu \left(\frac{\partial \mathcal{L}^{\text{matter}}}{\partial \phi_{\mu_1, \dots, \mu_n, \nu}} \right) - \frac{\partial \mathcal{L}^{\text{matter}}}{\partial \phi_{\mu_1, \dots, \mu_n}} = 0 \quad (12.9)$$

Notice that in the Lagrange equations for matter fields, one considers the variation with respect to the ordinary derivatives $\phi_{\mu_1, \dots, \mu_n, \nu} \equiv \partial_\nu \phi_{\mu_1, \dots, \mu_n}$ of these fields, although the Lagrangian depends on the gravitational covariant derivatives, which are necessary in the presence of a gravitational field for general covariance reasons.

In a full theory of dynamical gravitation the action consists of two parts. The gravitational part, which contains the kinetic terms for the gravitational field, namely the scalar space-time curvature in the Einstein theory, and the matter part, which includes the matter fields on a gravitational background, as in (12.8). The full action has the generic form

$$\mathcal{S}_{\text{total}} = \int d^4x \sqrt{-g} \mathcal{L}_{\text{Grav}}(g_{\mu\nu}, g_{\mu\nu, \rho}) + \mathcal{S}^{\text{matter}} . \quad (12.10)$$

The gravitational and matter fields are independent field variables and their variations should be considered separately. Variation of the matter action alone with respect to the gravitational field yields the matter stress tensor, $T_{\mu\nu}$, appearing on the right-hand-side of Einstein's equations, as we have seen in the Lectures. Variation of the gravitational part of the total action, with respect to the metric tensor, yields the dynamics of the gravitational field itself. In these Lectures we have discussed explicitly the case of variation with respect to the gravitational field $g_{\mu\nu}$ of the Einstein-Hilbert classical gravitational action coupled to matter, in order to derive Einstein's equations for the dynamics of the gravitational field itself and understand the connection of matter and space-time geometry.

Appendix B: Thermodynamical Equilibrium Formulae

In this Appendix we shall outline (and derive) the most basic formulae characterising Equilibrium Thermodynamics, which we made use of in various parts of the text, especially in Cosmology — the thermal history of our Universe and the calculation of thermal relic densities.

We commence our discussion with the formulae of the number density, n , the energy density, ρ , and pressure, p , of a dilute, weakly interacting gas of particles with g integral degrees of freedom, which is in thermal equilibrium with a heat bath of temperature T . Let $f(\vec{p})$ be the phase-space distribution (or occupancy number) of the particles in the gas. The relevant formulae are [1]:

$$n = \frac{g}{(2\pi)^3} \int f(\vec{p}) d^3p, \quad \rho = \frac{g}{(2\pi)^3} \int E(\vec{p}) f(\vec{p}) d^3p, \quad p = \frac{g}{(2\pi)^3} \int \frac{|\vec{p}|^2}{3E} f(\vec{p}) d^3p . \quad (12.11)$$

where the particles in the gas are characterised by the on-shell dispersion relation $E^2 = |\vec{p}|^2 + m^2$.

For a species in kinetic equilibrium, the occupancy number f is given by the Fermi-Dirac or Bose-Einstein distributions, depending on the spin of the particle:

$$f(\vec{p}) = \left(e^{(E-\mu)/T} \pm 1 \right)^{-1}, \quad (\text{in units of Boltzmann factor } k_B = 1) \quad (12.12)$$

where the $+$ ($-$) sign is for fermions (bosons), and μ is the chemical potential. If the species are in chemical equilibrium, when they interact with other species, e.g. via reactions of the form

$$i + j \rightarrow k + \ell,$$

then the chemical potentials are related via:

$$\mu_i + \mu_j = \mu_k + \mu_\ell.$$

Substituting the equilibrium distributions (12.12) into (12.11), we obtain the following equilibrium-thermodynamics formulae characterising the species of mass m and chemical potential μ , at temperature T :¹⁰

$$\begin{aligned} n &= \frac{g}{2\pi^2} \int_m^\infty \frac{(E^2 - m^2)^{1/2}}{e^{(E-\mu)/T} \pm 1} E dE, \\ \rho &= \frac{g}{2\pi^2} \int_m^\infty \frac{(E^2 - m^2)^{1/2}}{e^{(E-\mu)/T} \pm 1} E^2 dE, \\ p &= \frac{g}{6\pi^3} \int_m^\infty \frac{(E^2 - m^2)^{3/2}}{e^{(E-\mu)/T} \pm 1} dE. \end{aligned} \quad (12.13)$$

In the relativistic limit, which is relevant for the early epochs of our Universe, as we have discussed in the Lectures, one has: $T \gg m$. We can then distinguish two cases. The first, is when there is no degeneracy among the species, so $T \gg \mu$. In such a case, from (12.13) we get:

$$\begin{aligned} n &= \begin{cases} (\zeta(3)/\pi^2)gT^3 & \text{(Bose)} \\ (3/4)(\zeta(3)/\pi^2)gT^3 & \text{(Fermi)} \end{cases}, \\ \rho &= \begin{cases} (\pi^2/30)gT^4 & \text{(Bose)} \\ (7/8)(\pi^2/30)gT^4 & \text{(Fermi)} \end{cases}, \\ p &= \frac{1}{3}\rho. \end{aligned} \quad (12.14)$$

where $\zeta(x)$ is the Riemann ζ -function. We have $\zeta(3) = 1.202\dots$. The second of these relations are the known Stefan-Boltzmann law of thermal radiation. When applied to photons (Bosons) the energy density of photons is expressed in terms of the radiation constant σ as $\rho_{\text{rad}} = \sigma T^4$ (c.f. Lectures). The third of these relations is used to denote the equation of state of a radiation-era dominated Universe.

For the second case of relativistic species with degeneracy, we distinguish the case of fermi degeneracy and that of Bose degeneracy. They must be treated separately. For the fermi degenerate case, we have $\mu \gg T \gg m$, in which case:

$$\begin{aligned} n &= (1/6\pi^2)g\mu^3, \\ \rho &= (1/8\pi^2)g\mu^4, \\ p &= (1/3)\rho = (g/24\pi^2)g\mu^4. \end{aligned} \quad (12.15)$$

¹⁰Above we discussed Thermodynamics in a flat Minkowski space-time. The generalization to a curved Robertson-Walker (RW) (isotropic and homogeneous) space time is straightforward. Since, the infinitesimal line element of such a space time is written as: $ds^2 = -dt^2 + a(t)h_{ij}dx^i dx^j$ (c.f. Lectures), where x^i are ‘‘spatial’’ coordinates and h_{ij} a maximally symmetric spatial metric, the inclusion of curved space-time effects in the above formulae is achieved through the replacement $|\vec{p}| \rightarrow p_i p_j h^{ij} = p^i p^j h_{ij}$. The corresponding dispersion relation of the particle then reads: $p^\mu p^\nu g_{\mu\nu} = -m^2$, from which $E^2 = p^i p^j h_{ij} + m^2$. The spatial part of the homogeneous and isotropic space then decouples when we pass from a three-momentum to an energy integration, and hence the Minkowski space-time discussion carries intact to the RW case. This is understood in what follows.

For a degenerate Bose-Einstein species, $\mu > 0$ denotes a Bose-Einstein condensate, which should be treated separately from the other species. We shall not discuss explicitly this case here. For the relativistic case of fermions or bosons with $\mu < 0$, $T > |\mu|$, we have:

$$\begin{aligned} n &= e^{\mu/T} (g/\pi^2) T^3, \\ \rho &= e^{\mu/T} (3g/\pi^2) T^4, \\ p &= (1/3)\rho = e^{\mu/T} (g/\pi^2) T^4. \end{aligned} \quad (12.16)$$

In the non-relativistic case ($m \gg T$), which is also of interest to us here, especially when we discuss thermal massive dark-matter relics, the relevant formulae are the same for Bosons and Fermions, to leading order, and are given by:

$$\begin{aligned} n &= g \left(\frac{mT}{2\pi} \right)^{3/2} e^{-(m-\mu)/T}, \\ \rho &= mn \\ p &= nT. \end{aligned} \quad (12.17)$$

The second of this relation is expected from the fact that heavy non relativistic species, are (almost) at rest, hence their energy density is equivalent to their rest mass times their number density.

Finally, if the various species i are at thermal equilibrium, but at a temperature $T_i \neq T$, where T is the photon temperature, which is a situation common in early epochs of Cosmology, then the total energy density and pressure during the radiation era can be expressed as sums:

$$\begin{aligned} \rho_{\text{rad}} &= T^4 \sum_{i=\text{all species}} \left(\frac{T_i}{T} \right)^4 \frac{g_i}{2\pi^2} \int_{x_i}^{\infty} \frac{(\xi^2 - x_i^2)^{1/2} \xi^2 d\xi}{e^{\xi - y_i} \pm 1}, \\ p_{\text{rad}} &= T^4 \sum_{i=\text{all species}} \left(\frac{T_i}{T} \right)^4 \frac{g_i}{6\pi^2} \int_{x_i}^{\infty} \frac{(\xi^2 - x_i^2)^{3/2} d\xi}{e^{\xi - y_i} \pm 1}, \end{aligned} \quad (12.18)$$

where g_i are the internal degrees of freedom of the species i (spin, color *etc.*) and we reverted to dimensionless variables $x_i \equiv m_i/T$ and $y_i \equiv \mu_i/T$.

Since the energy density and pressure of non-relativistic species (12.17) are suppressed by exponential factors $e^{-(m-\mu)/T}$, they are not the dominant contributions to $\rho_{\text{rad}}, p_{\text{rad}}$, which are thus dominated by the relativistic species. Upon this observation, the respective expressions simplify significantly, yielding essentially the Stefan-Boltzmann law for radiation:

$$\begin{aligned} \rho_{\text{rad}} &\simeq \frac{\pi^2}{30} g_{\star} T^4, \\ p_{\text{rad}} &= \frac{1}{3} \rho_{\text{rad}} \simeq \frac{\pi^2}{90} g_{\star} T^4, \\ g_{\star} &= \sum_{i=\text{Bosons}} g_i \left(\frac{T_i}{T} \right)^4 + \frac{7}{8} \sum_{i=\text{Fermions}} g_i \left(\frac{T_i}{T} \right)^4, \end{aligned} \quad (12.19)$$

where g_{\star} counts the total number of relativistic (effectively massless, $m_i \ll T$) degrees of freedom and the relative factor of 7/8 in the second sum of the expression (12.19) for g_{\star} is due to the difference in the phase-space distribution function between fermions and bosons (12.12). The g_{\star} is a function of the (photon) temperature T , since the sums run over relativistic species, thus those whose masses $m_i \ll T$. For instance, for temperatures up to MeV, the only relativistic species are the photons and the three light neutrinos, $\nu_e, \nu_{\tau}, \nu_{\mu}$ (assumed they are indeed light), whilst when the temperature exceeds, say, $T = 300$ GeV, all the species within the SU(3)xS(2)xU(1) Standard Model of particle physics behave as relativistic.

In the radiation-dominated era, as we have seen in the Lectures, the scale factor of the Universe, scales with the cosmic time t like $a(t) \sim t^{1/2}$. That era occurs for $t \lesssim 4 \times 10^{10}$ sec, and the total

energy (pressure) of the Universe is very well approximated by $\rho(p) \simeq \rho_{\text{rad}}(p_{\text{rad}})$. From these it follows:

$$H \equiv \frac{\dot{a}}{a} = 1.66 g_{\star}^{1/2} \frac{T^2}{M_{\text{Pl}}} ,$$

$$t = 0.301 g_{\star}^{-1/2} \frac{M_{\text{Pl}}}{T^2} \simeq \left(\frac{T}{\text{MeV}} \right)^{-2} \text{ sec} . \quad (12.20)$$

These relations have been used in the Lectures, especially in the calculation of the thermal dark matter relic density from the Big Bang.

As discussed in the text, the entropy of an Einstein Universe, remains constant, and the corresponding entropy density per (proper) co-moving volume is

$$s = \frac{\rho + p}{T} \quad (12.21)$$

On account of the previous discussion, this entropy density is dominated by the effectively massless (relativistic) species, hence, on account of (12.19), to a very good approximation we have:

$$s = \frac{2\pi^2}{45} g_{\star S} T^3 , \quad g_{\star S} = \sum_{i=\text{Bosons}} g_i \left(\frac{T_i}{T} \right)^3 + \frac{7}{8} \sum_{i=\text{Fermions}} g_i \left(\frac{T_i}{T} \right)^3 . \quad (12.22)$$

These formulae are extensively used in the text. In particular, in the calculation of thermal relics one needs to evaluate quantities like $Y \equiv n/s$ at thermal equilibrium, Y_{eq} for non-relativistic species. According to (12.22) and (12.17), for non-relativistic massive species, this quantity is [1]:

$$Y_{\text{eq}}^{m \ll T} \simeq \frac{45}{2\pi^4} \left(\frac{\pi}{8} \right)^{1/2} \frac{g}{g_{\star S}} x^{3/2} e^{-x} = 0.145 \frac{g}{g_{\star S}} x^{3/2} e^{-x} , \quad x \equiv m/T \gg 3 . \quad (12.23)$$

This quantity is used in solving the associated Boltzmann equation, as we have discussed in detail in the Lectures.

This completes our review of the most important formulae of equilibrium Thermodynamics, used in this course on Modern Cosmology.

References

- [1] E. W. Kolb and M. S. Turner, *The Early Universe* (Frontiers in Physics, Addison-Wesley, 1990).
- [2] N. E. Mavromatos, “*LHC Physics and Cosmology*,” in Proc. of *22nd Lake Louise Winter Institute, Fundamental Interactions*, Lake Louise, Alberta (Canada), 19-24 February 007 (A. Astbury, F. Khanna and R. Moore eds., World Sci. 2008), p. 80-127, arXiv:0708.0134 [hep-ph]
- [3] A. B. Lahanas, N. E. Mavromatos and D. V. Nanopoulos, Int. J. Mod. Phys. D **12**, 1529 (2003), and references therein.
- [4] B. P. Schmidt *et al.*, Astrophys. J. **507** (1998) 46; S. Perlmutter *et al.* [Supernova Cosmology Project Collaboration], Astrophys. J. **517** (1999) 565; A. G. Riess *et al.*, Astrophys. J. **560** (2001) 49.
- [5] S. Perlmutter and B. P. Schmidt, arXiv:astro-ph/0303428; J. L. Tonry *et al.*, arXiv:astro-ph/0305008; P. Astier *et al.*, Astron. Astrophys. **447** (2006) 31. A. G. Riess *et al.*, Astrophys. J. **659** (2007) 98. W. M. Wood-Vasey *et al.*, arXiv:astro-ph/0701041.
- [6] C. L. Bennett *et al.*, arXiv:astro-ph/0302207.
- [7] D. J. Eisenstein *et al.* [SDSS Collaboration], Astrophys. J. **633**, 560 (2005).

- [8] G. F. Smoot *et al.*, *Astrophys. J.* **396**, L1 (1992); C. L. Bennett *et al.*, *Astrophys. J.* **436**, 423 (1994).
- [9] D. N. Spergel *et al.* [WMAP Collaboration], *Astrophys. J. Suppl.* **148**, 175 (2003); **170**, 377 (2007).
- [10] A. H. Chamseddine, R. Arnowitt and P. Nath, *Phys. Rev. Lett.* **49** (1982) 970; R. Barbieri, S. Ferrara and C. A. Savoy, *Phys. Lett. B* **119** (1982) 343; L. J. Hall, J. Lykken and S. Weinberg, *Phys. Rev. D* **27** (1983) 2359; P. Nath, R. Arnowitt and A. H. Chamseddine, *Nucl. Phys. B* **227** (1983) 121.
- [11] P. Binetruy *et al.*, *Eur. Phys. J. C* **47**, 481 (2006); P. Binetruy, M. K. Gaillard and B. D. Nelson, *Nucl. Phys. B* **604**, 32 (2001).
- [12] J. R. Ellis *et al.*, *Int. J. Mod. Phys. A* **21**, 1379 (2006), and references therein. G. A. Diamandis *et al.*, *Phys. Lett. B* **642**, 179 (2006).
- [13] S. Eidelman, talk at ICHEP 2006, Moscow (Russia).
- [14] H. V. Peiris *et al.*, arXiv:astro-ph/0302225; V. Barger, H. S. Lee and D. Marfatia, *Phys. Lett. B* **565**, 33 (2003).
- [15] A. Kosowsky and M. S. Turner, *Phys. Rev. D* **52**, 1739 (1995).
- [16] E. W. Kolb, S. Matarrese and A. Riotto, *New J. Phys.* **8** 322 (2006).
- [17] V.A. Mitsou, “*Constraints on Dissipative Non-Equilibrium Dark Energy Models from Recent Supernova Data*”, in *Proc. of 22nd Lake Louise Winter Institute, Fundamental Interactions*, Lake Louise, Alberta (Canada), 19-24 February 007 (A. Astbury, F. Khanna and R. Moore eds., World Sci. 2008), p. 363-367, arXiv:0708.0113 [astro-ph], and references therein; J. R. Ellis, *et al.*, *Astropart. Phys.* **27**, 185 (2007); N. E. Mavromatos and V. A. Mitsou, arXiv:0707.4671 [astro-ph], *Astropart. Phys.* in press (DOI:10.1016/j.astropartphys.2008.05.002).
- [18] M. Gasperini, *Phys. Rev. D* **64**, 043510 (2001); M. Gasperini, F. Piazza and G. Veneziano, *Phys. Rev. D* **65**, 023508 (2002); R. Bean and J. Magueijo, *Phys. Lett. B* **517**, 177 (2001).
- [19] M. Milgrom, *Astrophys. J.* **270**, 365 (1983).
- [20] J. D. Bekenstein, *Phys. Rev. D* **70**, 083509 (2004) [Erratum-ibid. *D* **71**, 069901 (2005)].
- [21] C. Skordis *et al.*, *Phys. Rev. Lett.* **96**, 011301 (2006).
- [22] S. Dodelson and M. Liguori, *Phys. Rev. Lett.* **97**, 231301 (2006).
- [23] E. Gravanis and N. E. Mavromatos, *Phys. Lett. B* **547**, 117 (2002).
- [24] V. Zacek, “*Dark Matter*,” in *Proc. of 22nd Lake Louise Winter Institute, Fundamental Interactions*, Lake Louise, Alberta (Canada), 19-24 February 007 (A. Astbury, F. Khanna and R. Moore eds., World Sci. 2008), p. 170-206, and references therein.
- [25] F. Zwicky, *Helv. Phys. Acta* **6**, 110 (1933).
- [26] M. Tegmark *et al.* [SDSS Collaboration], *Astrophys. J.* **606**, 702 (2004).
- [27] J. R. Ellis, *et al.*, *Nucl. Phys.* **B238** (1984) 453; H. Goldberg, *Phys. Rev. Lett.* **50** (1983) 1419.
- [28] A. B. Lahanas, N. E. Mavromatos and D. V. Nanopoulos, *PMC Phys. A* **1**, 2 (2007) [arXiv:hep-ph/0608153]; *Phys. Lett. B* **649**, 83 (2007) [arXiv:hep-ph/0612152].

- [29] N. Yoshida, *et al.*, *Astrophys. J.* **591**, L1 (2003).
- [30] J. R. Ellis, J. L. Lopez and D. V. Nanopoulos, *Phys. Lett. B* **247** (1990) 257; J. R. Ellis, *et al.*, *Nucl. Phys. B* **373** (1992) 399. S. Sarkar, arXiv:hep-ph/0005256 and references therein.
- [31] D. J. Chung, *Phys. Rev. D* **67** (2003) 083514.
- [32] J. R. Ellis, T. Falk, K. A. Olive and M. Srednicki, *Astropart. Phys.* **13** (2000) 181 [Erratum-*ibid.* **15**, 413 (2001)].
- [33] A. Birkedal, *AIP Conf. Proc.* **805**, 55 (2006) and references therein.
- [34] U. Chattopadhyay, A. Corsetti and P. Nath, *Phys. Rev. D* **68**, 035005 (2003); K. L. Chan, U. Chattopadhyay and P. Nath, *Phys. Rev. D* **58**, 096004 (1998).
- [35] J. Edsjo and P. Gondolo, *Phys. Rev. D* **56**, 1879 (1997).
- [36] J. L. Feng, K. T. Matchev and T. Moroi, *Phys. Rev. Lett.* **84**, 2322 (2000); *Phys. Rev. D* **61**, 075005 (2000).
- [37] H. Baer and C. Balazs, *JCAP* **0305**, 006 (2003).
- [38] G. W. Bennet *et al.* [BNL-E821 Collaboration], *Phys. Rev. Lett.* **89** (2002) 101804; C. J. Onderwater *et al.* [BNL-E821 Collaboration], *AIP Conf. Proc.* **549**, 917 (2002).
- [39] S. Narison, *Phys. Lett. B* **568**, 231 (2003).
- [40] J. R. Ellis, *et al.*, *Phys. Lett. B* **565**, 176 (2003).
- [41] A. B. Lahanas and D. V. Nanopoulos, *Phys. Lett. B* **568**, 55 (2003).
- [42] H. Baer, *et al.*, *JHEP* **0306**, 054 (2003).
- [43] H. Baer, T. Krupovnickas and X. Tata, *JHEP* **0307**, 020 (2003).
- [44] R. Arnowitt *et al.*, arXiv:hep-ph/0701053 and references therein.
- [45] G. A. Diamandis, *et al.* *Int. J. Mod. Phys. A* **17**, 4567 (2002);
- [46] G. Belanger, *et al.*, *AIP Conf. Proc.* **878**, 46 (2006); *Phys. Rev. D* **73**, 115007 (2006).
- [47] A. Pilaftsis and C. E. M. Wagner, *Nucl. Phys. B* **553**, 3 (1999).